

A guide to language use observation **SURVEY METHODS**

Olatz Altuna

Asier Basurto



EUSKO JAURLARITZA



GOBIERNO VASCO

HEZKUNTZA, HIZKUNTZA POLITIKA
ETA KULTURA SAILA

DEPARTAMENTO DE EDUCACIÓN,
POLÍTICA LINGÜÍSTICA Y CULTURA

A catalogue record of this book is available in the catalogue of the General Library of the Basque Government: <<http://www.bibliotekak.euskadi.net/WebOpac>



First published: January 2013

Number of copies: 850 copies

© Administration of the Autonomous Community of the Basque Country

Ministry for Education, Linguistic Policy and Culture

© Soziolinguistika Klusterra

Publisher by: Eusko Jaurlaritzaren Argitalpen Zerbitzu Nagusia

Servicio Central de Publicaciones del Gobierno Vasco

Donostia kalea, 1 01010 Vitoria-Gasteiz

Authors: Olatz Altuna and Asier Basurto

With assistance from Network to Promote Linguistic Diversity

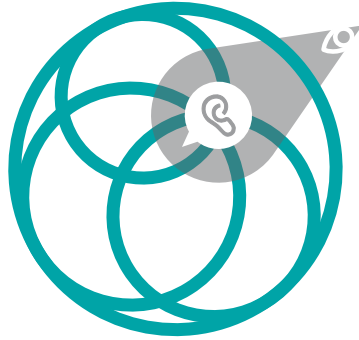
English translation: Alan R. King (BAKUN, S.L.)

Graphics: Zirrimarra (Andoain)

Printing: Ulzama Gráficas, S.L.

ISBN: 978-84-457-3287-8

L.G.: VI-55-2013



A guide to language use observation

SURVEY METHODS

Olatz Altuna
Asier Basurto



HEZKUNTZA, HEZKUNTZA POLITIKA
ETA KULTURA SAHIA
DEPARTAMENTO DE EDUCACIÓN,
POLÍTICA LINGÜÍSTICA Y CULTURA



Eusko Jaurlaritzaren Argitalpen Zerbitzu Nagusia

Servicio Central de Publicaciones del Gobierno Vasco

Vitoria-Gasteiz, 2013

Contents

INTRODUCTION:

Patxi Baztarrika, Vice-Minister for Language Policy 7

INTRODUCTION:

Jannevieske de Vries, NPLD Chair 9

ACKNOWLEDGMENTS:

Olatz Altuna, Asier Basurto, *Soziolinguistika Klusterra* 13

PREFACE: Jacqueline Urla, University of Massachusetts 15

INTRODUCTION 21

| | | |
|-----------|--|-----------|
| 1. | OBSERVING SPOKEN LANGUAGE USE | 27 |
| 1.1. | The subject of study: spoken language use | 28 |
| 1.2. | The method's basic principle: observation | 28 |
| 1.3. | The method's main features | 32 |
| 1.3.1. | Be inconspicuous | 32 |
| 1.3.2. | The unit of observation is the conversation | 33 |
| 1.3.3. | The length of the conversation is not relevant | 33 |
| 1.3.4. | Conversations in a single language | 34 |
| 1.3.5. | Each survey has its own design | 34 |
| 2. | A GUIDE TO STREET USE SURVEYS | 39 |
| 2.1. | Design | 41 |
| 2.1.1. | Variables | 41 |
| 2.1.2. | The place: spots and routes | 45 |

| | | |
|-----------|---|-----------|
| 2.1.3. | Times and duration of survey sessions | 47 |
| 2.1.4. | The number of conversations surveyed: the sample and reliability | 50 |
| 2.1.5. | Trends in language use: continuity from survey to survey | 54 |
| 2.2. | Fieldwork | 55 |
| 2.2.1. | Data records | 55 |
| 2.2.2. | The mobile phone | 60 |
| 2.2.3. | Data collection criteria | 62 |
| 2.3. | Processing the data | 70 |
| 2.4. | Interpreting the results | 74 |
| 3. | OBSERVATIONS OF LANGUAGE USE IN THE BASQUE COUNTRY | 81 |
| 3.1. | Euskal Herria, the land of the Basque language | 81 |
| 3.2. | Background and development of the study | 85 |
| 3.3. | Main conclusions from the study | 86 |
| 4. | OBSERVING LANGUAGE USE AT WORK AND SCHOOL | 99 |
| | REFERENCES | 107 |

INTRODUCTION



Patxi Baztarrika
Vice-Minister for Language Policy

There exists a range of ways to measure the situation and progress of a language, each with different characteristics, which may generally be considered mutually complementary. Typically registries and censuses collect information about language knowledge, albeit somewhat superficially as a rule. Sociolinguistic surveys provide periodically updated data about linguistic competence, generational transmission, language use and attitudes to language campaigns, and are also valuable for tracking trends over time. In the Basque Country a sociolinguistic survey based on a significant sample of the population is carried out throughout the entire Basque-speaking area every five years. Last but not least, there are also methods for directly observing spoken language use, and such methods are the basis of the study presented here.

Observation-based use surveys have improved in power and sophistication over the past twenty years in both theoretical and practical terms. For one thing, mathematical and statistical principles have been worked out in order to ensure the representativity of the observational data. As with other methods, accumulated experience has led to the fine-tuning of various aspects of the process. Direct observation of language use is an activity with little tradition

in the analysis of minority languages elsewhere in Europe and the wider world. Therefore I believe it is a good thing to make known outside our country the method that has been used to analyse the situation of Basque, given that in some other language communities the development of our work has been followed with interest. Here too, as in the case of the processes whereby sociolinguistic surveys have been designed and carried out, there are things worth sharing that can benefit us all.

This book has an eminently practical aim. It tells how to observe and measure oral language use, with explanations of significant methodological concepts. It details every phase of the survey process through examples, in a way that is helpful to anyone wishing to undertake a study of this kind. Given its focus on methods, the purpose of this handbook is to lay out in a clear and comprehensive manner the things that need to be taken into consideration when performing surveys through observation; and that purpose has been fulfilled admirably by Olatz Altuna and Asier Basurto, of Soziolinguistika Klusterra, the authors and coordinators of this excellent book.

I am confident that this manual, which explains and tells how to apply observational methods, will be found invaluable in many other countries, from Wales to Finland and from Ireland to Galicia. Any effective language policy must be based on reliable analyses and profiles of the sociolinguistic situation, given that the most accurate knowledge available about the real situation has to precede any measures to be taken. This can be achieved above all by combining different kinds of study which complement each other, since there are a great many different aspects to be considered in order to achieve success in understanding a linguistic reality which is often hard to perceive accurately. Needless to say, that understanding is needed as a pre-condition for intervention of any kind. Data, and methods for obtaining that data, are key elements of any such endeavour.

INTRODUCTION



Jannewietske de Vries
NPLD Chair

Languages are one of the key traits that define a community. Each language conveys a specific way of viewing the world and a collective imagery, which allows humans to name, organise and classify everything that surrounds us. Languages are tools that we use to communicate and share experiences, as individuals and groups alike.

The best way to ensure the vitality of a language is to use it in all spheres of life and to pass it on to future generations. Languages are capable of adapting to new social requirements, whether through innovation or through enrichment thanks to contact with other languages. All languages are open, flexible, dynamic and incomplete and all languages are constantly enriched by other languages.

The **Guide to Language Observation** represents an extremely valuable tool at the disposal of all language communities to assess the health of a language in its daily use. Governments may encourage, foster and develop language policies and strategies aimed at promoting their languages. However, the best barometer to measure the success of these policies and the vibrancy of the language is the practical and spontaneous use of the language. Only by knowing how widely used is a language will Governments and language planners be able to take measures to further increase the number

of speakers of a language and the contexts in which it needs to be promoted.

When a language is no longer spoken this leads to a major rupture in terms of intergenerational cultural integration and a community's social cohesion. Mechanisms to assess the oral vitality of a language are therefore as necessary as ever to guarantee the health of all languages and to ensure that they continue as a vibrant means of communication.

The Network to Promote Linguistic Diversity (NPLD) works to ensure that Constitutional, Regional and Small-State (CRSS) languages are represented at European level and to raise awareness on the need to promote Europe's languages, regardless of number of speakers or political status. I am deeply convinced that all language communities who are members of NPLD will benefit from this valuable piece of work. The NPLD will continue to support initiatives such as this one, which are extremely beneficial for politicians, policy makers, practitioners and researchers alike across Europe.

ACKNOWLEDGMENTS



Olatz Altuna (*Soziolinguistika Klusterra officer.*
Director, Street Surveys of Euskal Herria)
Asier Basurto (*Soziolinguistika Klusterra officer*)

Before entering into the subject of this publication, we wish to mention the people and organisations who have created, developed and implemented methods for measuring language use, thanks to whom the present book was possible. It is owing to the determination and cooperation of a great many advocates of public observation surveys that the methods presented here could be developed and used for the past twenty-five years.

As a member of the Siadeco research centre, Iñaki Larrañaga was responsible for the creation of this methodology when, in the 1980s, the Committee of Basque Culture (EKB) instigated the observation of language use, or *Kale Neurketa* (Street Survey) as it has come to be better known, all over Euskal Herria.

Subsequently, many people contributed to the improvement of the methodology, among whom we must make special mention of Jose Jabier Iñigo, Kike Amonarriz, Imanol Esnaola, Léonel Joly and — as current members of the Scientific Committee — Iñaki Martínez de Luna, Xabier Isasi, Rosa Ramos and Iñaki Iurrebaso.

We also wish to pay special homage to our late, dear friend Jose Luis Alvarez Enparantza, alias Txillardegí, who contributed a great deal by creating a mathematical model of isotropic language use (Txillardegí, 1994; 2001) based on this method.

Since 2006, Yosu Yurramendi and Eñaut Urrestaratzu have served as statistic consultants, and they are responsible for the statistical model used to calculate samples and margins of error. The initial data were organised by Nestor Urreztaratzu in a structured, accessible way, thanks to which it was possible to track the trends in language use over time.

Assistance was provided by the EMUN cooperative and EL-HUYAR assessors without which it would not have been possible to adapt linguistic observation to the milieu of the workplace and specific spaces.

This is a strategic project for Soziolinguistika Klusterra, in which all of its members have participated in one way or another. To the office staff —Belen Uranga, Denis Elortza, Ibon Usarralde, Iker Salaberria, Isabel Godinez, Larraitz Garmendia, Oihana Lujanbio and Pablo Suberbiola — we express our gratitude.

We would be amiss if we did not also acknowledge the organisations supporting through funding this research and the resulting publication, namely the *Network to Promote Linguistic Diversity* (NPLD), the Office for Language Policy of the Basque Government, the provincial governments of Araba, Bizkaia and Gipuzkoa, and a number of town councils.

Finally, a word of gratitude to the hundreds of surveyors who have collected data over the years since 1989, thereby making an essential contribution to the six editions thanks to their commitment and dedication as volunteers since the beginning of this process. Their generous help sowed the seeds for later, more systematic and, if the term is justified, more “professional” efforts. It is undeniable that without these surveyors’ participation we could not have come all this way, and even more importantly, would not dispose of so much valuable information about the state of the Basque language.

PREFACE



Jacqueline Urla,
University of Massachusetts

It is almost thirty years ago that I first encountered an early prototype of the *Kale Neurketa*, or Street Survey of Basque language use that is the subject of this book. It was in the summer of 1983, during the annual Santa Ixabelak celebrations of Usurbil (Gipuzkoa). This was a year of great hope and enthusiasm for Basque revival: the Law of Euskara (1982) had just been passed and Basque was declared co-official with Castilian in the Basque Autonomous Community. Usurbil youth in charge of organizing the festivities wanted to incorporate a language awareness campaign into the celebrations. Basque was gaining new ground and institutional support after the end of the Franco dictatorship, but how much did it actually circulate in daily life? Conducting ethnographic research in Usurbil at the time, I witnessed how local language advocates worked with the technical assistance of SIADECO, an independent Basque research institute, to collect data on the relative amount of Basque or Spanish being spoken in public conversations. Observers were trained in the method of listening and recording the language of conversations taking place along the main streets of this small town. Data in hand, they then posted the low results of Basque language use in the town square for the duration of the celebrations for all to see. Along-

side the statistics, they hung a beautifully painted mural painted by schoolchildren that called out to residents to make Basque a part of their celebrations.

If I am not mistaken, this was the first, or one of the very first, trial runs of what would become the Kale Neurketa. This popular use of statistical measurement of language use prompted me to delve deeper into the role of statistics as ubiquitous and powerful instruments in defining and debating modern social reality both in the Basque Country and beyond (Urla, 1993; 2012). I found a longer history going back to the early 20th century of Basques looking to statistical surveys as a way of more objectively understanding their social reality, whether it is language, public health, or the economy (Urla, 1989). Statistics were seen early on as necessary instruments of modern, rational government. In the field of Basque language advocacy, we find a recurring passion for enumeration that stems from a hope that quantitative measurement conducted according to sound criteria could pierce through the grip of political ideologies and assumptions that many advocates believed made it difficult to accurately assess the linguistic situation (see SIADECO, 1979). Indeed, within the wider field of lesser-used languages in Europe, Basque language advocacy stands out for its robust and sustained tradition of demolinguistics.

The Street Survey can be seen as part of this tradition. But it is also unique in its attempt to devise a means of direct observation of language use, rather than rely on the more typical method of self report. As Iñaki Larrañaga of SIADECO said to me in 1983, the Street Survey was an attempt to “take a snapshot” of the public language use. Why measure usage? Why is this important and why do it through observation? What does it contribute?

Scholars know that numbers of “speakers” can often be deceptive when it comes to minoritized languages. Histories of stigmatization, prohibition and ideologies of language that dictate accommodation to majority language speakers can often lead to very limited public use of a minority language (Woolard, 1989). As a result, the

measure of competence in a minoritized language may be a feeble indicator of the vitality of a language as a medium of communication. Similarly, the social and political milieu has been known to sway how speakers report their knowledge of the language in question as well as their estimations of usage (Bourhis and Sachdev, 1984). While this is generally true for the reporting of all kinds of behavior, reporting on the use of minoritized languages may suffer more acute distortion due to what I call the phenomenon of “misplaced scale”. We see this in relation to judgments about the presence of all kinds of minorities – women, sexual or racial others, or minority language speakers–. An objectively small percentage of minorities in public spaces will often be remembered as larger than it actually was. Two or three black people, or women, or gays, in a group of 20 white men will be recalled as “integrated”. Similarly, when it comes to language, the use of an emblematic word or an occasional phrase, may be enough for a majority language speaker to recall an interaction or event as having transpired in the minority language (Hill, 2008).

Observational methods, while not immune to error, can step into this gap. Like measures of the linguistic landscape that document written language in public space (Landry and Bourhis, 1997), observation-based measures of spoken language give us data that stands independent of speakers’ perceptions. This can be quite informative, provide a baseline of data to compare over time, and serve as a useful contrast and complement to self report. Linguistic anthropology and the ethnography of communication in particular makes abundant use of observational data collection in the belief that it can reveal patterns and behaviors about which speakers are often unaware or take for granted. Speakers, for example, often do not “hear” certain accents, perceive code switches, or accurately gauge the amount of interrupting that they do. As studies of dialect shift have shown, observational data can also help to identify social change in process before the changes are apparent to social actors or investigators.

Thus we have several compelling reasons for wanting to gather information on language use and to do this observationally. There

were not, however good ways of doing this. This is where the Street Survey breaks new ground. It is a means of collecting quantitative data on language use in public space. The focus on public space is worth underscoring. Certainly this focus is partly conditioned by pragmatic factors: it is after all easier for researchers to have access to public space. But the focus on language use in public space also reflects the long-standing goal of the language revitalization movement to transform Basque into a public language – a language for generalized use that is not restricted to an ethnic group or a limited set of social domains (Woolard, 2008; Gal and Woolard, 2001). Becoming a public language is very much what language normalization has meant in both the Basque and Catalan language movements. Conducted periodically over the span of more than two decades, the Street Survey offers a series of “snapshots” that can be compared to each other. In that way, they can be used as a kind of barometer of the efficacy that language promotion and education policies are having in meeting this specific goal of making Basque a “public” language.

An innovative instrument, the Street Survey is, to my knowledge, the only one of its kind. It was developed by volunteer language advocates and researchers to document the disjuncture between linguistic competency, advocacy and actual language use. It is a tremendous undertaking spanning all seven of the Basque provinces. Its existence and continued refinement as a survey instrument is a testament to the vitality and innovation that comes from the nongovernmental sphere of language advocacy in the Basque country. Its history also speaks to positive developments that have taken place in language advocacy. In its early years, official Basque language policy ignored the survey, casting doubt on its scientific validity, and rarely if ever acknowledging the survey results in governmental reporting on the status of the language. I am happy to say that this is no longer the case and that in general we see a more harmonious and collaborative relationship between governmental and nongovernmental language researchers. This kind of pluralism and

mutual respect bodes well for the future of Basque sociolinguistics and language policy.

With this book, Basques are extending that collaborative spirit outward to other linguistic communities, hoping to make the methodology of the Street Survey more widely known so it can be used and tested in new contexts. I have no doubt that other language advocacy groups will find it useful. At the same time, its limitations need to be kept in mind. First, this survey is meant to obtain broad-scale quantitative data. By the nature of its design, it cannot capture the subtler features of speech, features such as codeswitching or other forms of linguistic alternation that often occur in conversations, particularly in minority or multilingual contexts. It reduces, of necessity, the complexity of linguistic practices. The snapshot is helpful, but not the whole story. Understanding patterns of language use will require supplementing this data with other more detailed observational studies of linguistic practice in the tradition of the ethnography of communication so that we can understand the dynamics of language choice and linguistic interaction.

Second, the meanings of the percentages the survey delivers are not self-explanatory. Over time the survey can show us the evolution in public use of the language. It can give some clues as to contexts or social groups that may correlate with greater or lesser public language use. The presence of children, for example, has been found to be a factor strongly correlated with Basque language use. But explaining the results requires, once again, complementing this survey data with in-depth inquiry into the symbolic values associated with the available languages, norms governing language choice and studies of the evolving linguistic marketplace. It is important not to fall back on intuition to explain the percentages that the survey produces. Basques have come to realize, for example, that it is in fact mathematically unreasonable to expect the percentage of Basque language use would match the percentage of Basque speakers (Martínez de Luna, Isasi and Altuna 2006). One had to take into account the probability that speakers will find themselves in public spaces

with other people who can speak the language. Thus I encourage all who would take up this method to follow the Basque example and not only publicize the results, but also invite a broad range of scholars to comment on them. Combine the survey with other research methods to better understand the results. The publication of each Street Survey is an occasion to invite discussion and analysis, not simply a “report card” on the language movement. The Survey, if it serves its purpose, should stimulate new questions and clues for further inquiry, and in that sense contribute to a deeper understanding of the social life of languages in multilingual contexts.

Jacqueline Urla

Department of Anthropology

University of Massachusetts Amherst



INTRODUCTION

Observations of language use have a long history in the Basque Country. The first surveys were carried out by Siadeco during the 1980s in a wide range of localities, and have been repeated periodically every four or five years, among other reasons because the data thus obtained are very useful for tracking the language's vitality over time. Today it may be stated without hesitation that such data have become essential for a full acquaintance with the situation of the Basque language. Consequently, as the years pass more and more towns have been performing such observations, or street surveys.¹

With twenty-five years' accumulated experience, we now consider this a proven and established method. Outside the Basque Country this research tool is also becoming more widely known, and we have witnessed a rise of interest in it on the part of foreign researchers.

We wish to make known this tool, which has proved so useful for diagnosing the situation of Basque, to a wider audience and in particular to communities with minority languages and researchers who want to quantify spoken language use in order to analyse linguistically diverse situations.

1 In the Basque Country observations of language use have come to be known as *kale-neurketak* which we translate as "street surveys".

The purpose of these observations is to measure a language's use in spoken interactions in public places. The object of study is use and the technique used is observation. In this manual the technique will be explained. This book aims to be a user's guide of sorts.

For the most part we shall talk about observations carried out in public, on the street. However, at the end of the book we will also discuss observations at the workplace and in schools, together with some general comments.

In the preface, the anthropologist Jackie Urla of the University of Massachusetts writes about the need of such studies and the importance of this observation technique from point of view of a foreign specialist who is familiar with the subject.

In the section titled "Observing spoken language use", we will present the values that inspire this research. In our view there is a basic principle: if we wish to ensure the future of a language, it is essential to keep close track of its use as an oral medium, because it is necessary for a language to be used and to be present in our day-to-day conversations in order for it to live. This study will indicate the extent to which this is happening and who is speaking the language. It also addresses more specific questions, such as: How is use distributed by districts? What are the characteristics of those districts, and what percentages of use are found in them? Do young people use the language more than adults, or vice-versa? Does the presence of children in conversations favour the use of the language?

This work presents some methodological principles in order to answer such questions properly.

- The basis of the methodology is observation. Thus we will deal with objective facts, without asking the speaker about anything: researchers limit themselves to noting the language of the conversation and the characteristics of the speaker. In other words, researchers collect plain data without taking the opinion or self-evaluation of the subject into consideration. In the second section we shall examine the particular features of observational studies in comparison with other research techniques.

- What is noted is the quantity of spoken interaction, not its quality. The amount of language use is quantified, independently of any evaluation of the correctness or appropriateness of language use.
- Normally we note the language of informal conversations, i.e. spontaneous spoken usage among people who are mutual friends or acquaintances.
- The unit of study is the conversation, and information about conversations must be recorded inconspicuously. The length of conversations is immaterial: long or short exchanges are accorded equal value in the research results. What is noted is the language in which the conversation takes place, and if the speakers in a given conversation switch between languages the predominant language will be noted, without taking the other language into account.

Having explained these methodological principles, the second section presents a guide for carrying out street surveys in which, before entering into details about how to perform the fieldwork, we discuss the need for a particular design prior to each survey. When fieldwork is described, in addition to criteria for data collection the media employed in data collection (the data record and the mobile phone app) are also presented. These respond to criteria for data collection; thus, their design is important. The guide also includes guidelines on the treatment of data and interpretation of results.

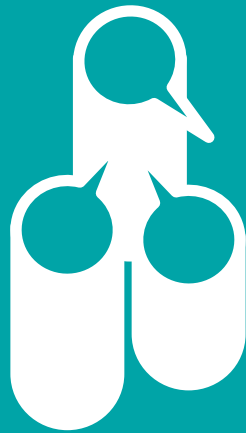
The third section reviews the most significant study that applies this technique in Euskal Herria, namely the Street Use Survey of the Basque Country, looking in particular at the latest survey (of 2011), and so focusing on the situation today. But given its importance, space is also allotted to its development over the years.

From this experience in the open space of the “street”, we switch to a look at more restricted settings in the fourth section, with a description of the nature and particularities of surveys in workplaces and schools while adhering to the same methodological principles.

A guide to language
use observation

SURVEY METHODS





1.

OBSERVING SPOKEN LANGUAGE USE



Speaking a language is a commonplace, everyday activity of human beings. We use language in all sorts of places, at all sorts of times. Language is a distinctly human trait. Use of language is not a regulated activity, nor does it conform to set patterns. No spatial or temporal limits can be imposed on language use. Consequently, measuring language use is a complex business, and to do it properly a sophisticated, refined method is required.

We have created and developed a method of measuring language use in order to address this purpose, and this is the method whose general principles will be presented in this section. Note that it is a research tool that has been designed with this particular purpose in mind. Therefore we believe that, unlike methods developed for general purposes, this research technique has the right characteristics that allow us to gather data about language use in an appropriate fashion.

The reliability of the measurement of spoken language use is ensured by the use of robust criteria and the right kind of tools. To spell this out more clearly: unless we are careful about methodological principles and apply these correctly, measurements of quality cannot be made, and unreliable measurements are of little use. If we really wish to learn about a language's situation and its trends over time, so that we can move in the right direc-

tion in developing effective language planning measures on the basis of this knowledge, well-designed and well-developed methods of measurement are needed.

1.1.

The subject of study: spoken language use

Experience in sociolinguistics has shown that the most meaningful and reliable index of a language's situation is its use. The best way to evaluate the state of "health" of a language is by considering that depending on whether use is higher or lower, over time it will eventually grow or contract. A given language's level of use among the population provides information that is not given by that population's level of language knowledge or by any other index of its language loyalty.

Another important point to underline is that the language use that this research method studies is oral. Oral or spoken language use is the most natural use of language. It is the most direct kind of linguistic interaction between two or more individuals. Spoken language use is also characterized by spontaneous improvisation. The choice of spoken language is often a reflex action rather than a result of reflexion.

1.2.

The method's basic principle: observation

The method presented here differs from other methods in a third respect: in it spoken language use is measured through observation. Surveying language use through observation is a technique for the quantitative study of social activity (see Table 1) which aims to chart the real frequency of acts using each of the languages in practice.

Table 1. Quantitative, qualitative and participatory methods

| | QUANTITATIVE METHODS | QUALITATIVE METHODS | PARTICIPATORY METHODS |
|--|---|--|--|
| SCIENCE AND RESEARCH OBJECTIVES | ▶ To explain, predict and chart social situations | To comprehend the sense and meanings of social behaviours | To participate in the changing of reality |
| OBJECT | ▶ Acts of individuals. Social events | Actors' discourses. Meanings | Interests and needs of collective subjects |
| LANGUAGE | ▶ Statistics | Concepts, metaphors | Socio-historical, strategic |
| SUBJECT | ▶ The sample | Social actors in their socio-structural situations | Self-diagnosis by social actors |
| DESIGN | ▶ Systematic, structured, repeatable | Open, flexible, variable | Open process: action-study-action |
| TECHNIQUES | ▶ Surveys Case records Observation | Discussion group In-depth interview Three-person group Life story | Participatory research Sociological participation |

Source: Altuna et al. (2008) and Juaristi (2003).

Observation is one of the main quantitative techniques for social research, even though questionnaires and other methods are generally used more. This is how Jone Miren Hernandez defines the observation technique in *Soziolinguistika Eskuliburua* (Zarraga et al., 2010:310):

The goal of direct observation is to examine and analyse people's behaviour systematically and in a controlled way, without any kind of manipulation or mediation. Unlike other techniques, in direct observation there is no effect from special stimuli, individuals' behaviour is observed directly.

In comparison with the more often used quantitative techniques, observation has some advantages and some drawbacks:

Table 2. *Advantages and disadvantages of observation*

ADVANTAGES OF OBSERVATION

- In general, fewer influences are felt than in other techniques when observing people. As Juaristi (2003:197) says, the researcher has less effect on subjects' behaviour and expressions, which helps to produce better data.
- If observation is performed in naturally occurring circumstances, reality will be studied as it is without any kind of change, influence or manipulation.
- In the other techniques the researcher is to a great extent at the mercy of the researched subject who decides what information to provide. But in the case of direct observation the researcher decides what is most important and what to observe.

**DISADVANTAGES
OF OBSERVATION**

- In many cases, observation is also the only possible technique, when on account of the situation, the time or the characteristics of the subject (such as when the subject is a child or an aged person) it is the only feasible technique.
- Observation entails immersion in the context of the study, which adds to the technique's value.

- Observation can only be used to study the present, not the past or the future.
- Some events, situations or behaviours cannot be observed directly because of their intimate nature or their hidden character.
- Some phenomena cannot be studied through observation because mutually related actions are taking place simultaneously in different places. In such cases the solution is to share the task of observation between several people.
- Studies based on observation tend to be bulkier than those carried out by other techniques. Thus (the lack of) time may be a limitation for the use of this technique.

Source: Zarraga et al., 2010: 311-312 (the original text has been adapted)

Having looked at the characteristics laid out in Table 2, we are in a position to conclude that direct observation is a technique suited to our purpose, since the advantages outweigh the drawbacks. So this technique is the one chosen for surveying spoken language use.

1.3.

The method's main features

1.3.1.

Be inconspicuous

One essential condition that any observational survey must meet is that the data must be obtained inconspicuously in the area under study. Speakers should not know that they are being observed in order to make it possible to register their real behaviour with respect to language use, which as we have already seen is what determines to a large extent the value of observation-based surveys. This criterion of staying inconspicuous is what makes it possible to observe the speaker's practice by obtaining data about use as if the observer were not present.

It helps to collect data inconspicuously if, without speakers being aware that they are being observed, researchers work in pairs when collecting data. Two further measures taken in the most recent studies, to provide for the use of suitable media for data-collection, are:

1. If data are noted on paper, the pages should be kept as small as possible.
2. We have created a special mobile phone app as an alternative to the use of paper.

When we come to discuss fieldwork we will describe these procedures of data collection on paper or on the phone.

1.3.2.

The unit of observation is the conversation

In this method the basic unit observed is the conversation. We do also note the number of individuals taking part in conversations, but the conversation is the primary unit.

In each observation session a given conversation is only registered once. The conversation is delimited by a change of language or of participants. This means that if the same conversation-group in the same language should be observed more than once, it will be collected and registered as a single conversation. After observing and recording a group's conversation, if the group switches to a different language then a different conversation will be registered in the observation records. This will also happen if a member of the group leaves or a new participant enters into the conversation: it will be treated as a new conversation.

When another participant in the conversation speaks (i.e. there is a change of active participant) this does not mean that another conversation has started, provided the language remains the same.

1.3.3.

The length of the conversation is not relevant

It is immaterial how long conversations last. No data are recorded about the length of conversations. The purpose of the survey is to track the frequency of language *choices*, not to measure the length of time for which each language is spoken.

1.3.4.

Conversations in a single language

The observers record the language of each conversation. To do that it is not necessary to observe all the participants in the conversation; unless there is a reason to think otherwise, it is assumed that the same language used by one member of the group is used by all of them.

The application of this criterion may raise two main types of uncertainty: In the first place, there is the issue of code-switching, where a speaker uses more than one language within a single stretch of spoken discourse. Here we include situations in which utterances in the different language alternate continually with similar frequencies, and also cases where utterances consist of words from different languages. Such conversations are not counted in the survey.

The other type of uncertainty concerns bilingual conversations, where some participants in a single conversation speak one language and others speak another. If we wanted to include such cases in the survey it would be necessary to observe all conversation participants in all conversations.

1.3.5.

Each survey has its own design

The survey is designed in accordance with the characteristics of the area that is to be observed. The design will involve specification of the time and duration of observation sessions, places, the variables to be studied, etc. The way to design each survey and the characteristics that should be taken into account will be considered in the next section. However, it is worth noting here that it is desirable for designs to be adjusted to each particular case. People familiar with the real situation to be analysed should par-

ticipate in this design process. Intimate knowledge of the context may be very useful in order to make the best decisions about appropriate places, times and so on. It is not a good idea to copy over verbatim a design used in one survey to another without bearing in mind the specifics of each place.

On the other hand, if we plan to analyse trends over time in language use in a given town or place, it *is* desirable to repeat exactly the same survey design in order to make sure that the data that are compared have been collected in conditions that are as close as possible to identical.

A guide to language
use observation

SURVEY METHODS





2.

A GUIDE TO STREET USE SURVEYS



Measurement of language use through observation have been used for different purposes and in different settings, although one of those settings, the street, has the longest tradition. Street use surveys are a means of assessing the language situation in a particular town or neighbourhood, and they serve to tell us what the trend is for each language, in terms of whether it is gradually being used more, or less.

Numerous other quantitative indicators might be used to identify a town's sociolinguistic profile, such as census data, enrolment statistics in schools and language schools, questionnaires and so on; but whatever situation is described by such data, if it is manifested anywhere, is manifested in daily language use on the town's streets.

The street or public space is the most neutral space there is. It is a place where individuals of all ages, genders, classes and types cross paths. It is a space belonging to everyone and to no one, a place where real coexistence between people occurs, and that is also true on the linguistic level. It is a setting for citizens' public activities (such as meeting one's friends, playing, or interactions with neighbours); but also a place of transit from one private domain to another, the route connecting the home, the workplace, school, administrative offices, etc. The boundaries of this street-space are

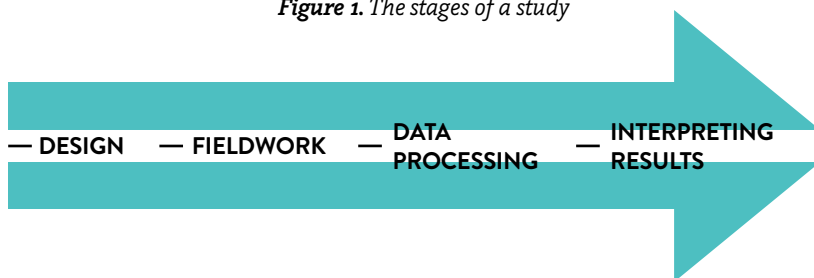
formed by buildings; hence the activities of the street are bounded by those of such buildings, such as commerce, the exchange of information, publicity, and so on.

An increase in the use of a given language on the street is a sign of that language's expansive tendency and an indicator of its vitality. If on the contrary street use diminishes, that may be a sign of contraction of the language away from the public domain towards the private, and thus be an indicator of sorts of a regressive trend.

Measurements of a language's street use may be useful tools for a wide variety of groups and institutions, such as local administrations wishing to design and implement language policies, grassroots language movements wishing to evaluate the achievement of their objectives, or groups and researchers interested in the sociolinguistic aspect of cultural diversity in urban environments.

In general terms, the main stages in carrying out street use surveys are similar to those which must be followed in most forms of research: design, fieldwork, processing the data and interpreting the results (Figure 1).

Figure 1. The stages of a study



Each of these steps will now be considered in turn to provide anyone interested in performing a street survey with some general guidelines.

2.1.

Design

In any research project, the design stage puts together a plan of the actions that must be performed to achieve the goal. Most commonly, the goal of street surveys of language use is to produce a quantitative description of the language use that takes place on the streets of a town or neighbourhood.

In order for the results of such a survey to provide a description of linguistic behaviour in a town or a part of a town, it is important that the information obtained should come from places, times and individuals that are representative of the locality being studied. Thus the value of the results obtained is dependent on the quality of the study's design. To carry out a street use survey it is necessary first to stipulate the variables to be observed, the places where and times when observation will take place, and the sizes of the samples to be produced.

In order for the design of a project to respond to its purpose, it is a good idea for people who are fully familiar with the real situation under study to participate in the design process. Information provided by the town's inhabitants are the best source of knowledge for identifying the most meaningful areas and points in which to make observations. In this way the design will be more satisfactory, its effect on the quality of the study more noticeable.

2.1.1.

Variables

The main variable in a street use survey is the language spoken. Thus the language is the *dependent variable*. All other variables apart from this which may be noted in the collection of information about speakers will be the

so-called *independent variables* which may somehow be linked to the dependent variable.

Suppose, for instance, we have a hypothesis that *people over forty speak Basque more than people under forty*. In this case, in addition to observing in our fieldwork what language people speak in the street, we shall also want to record whether people are older than forty or younger than forty, since that will give us a way to either confirm or refute our hypothesis by means of the study.

When designing the study, therefore, it has to be decided what information (i.e. which variables) are to be recorded in addition to information about language use. In order to decide that, three aspects are important to consider:

1 INTEREST

Unless we hypothesize that there may be some relationship between a given variable that we want to record and language use, there is unlikely to be any point in recording it.

If the purpose is to discover the profile of speakers who speak a language, the variables that it makes most sense to record, and which normally are registered, are major social descriptors such as age, sex and so on.

2 AVAILABILITY

It should be borne in mind that the individuals observed in the street survey will have to be classified according to the variables that are chosen. Therefore the information in such variables must be obtainable through simple observation.

3 FEASIBILITY

This criterion has to do with the amount of data that it is decided to record about each subject. Given that the data must be recorded on a form, there may be practical limits or difficulties associated with the possibility of recording information about a lot of variables at the same time. The way in which information is recorded during data collection should be simple, clear and practicable, so only variables that are really needed should be collected.

In the street use surveys that have been carried out to date in the Basque Country, the following information has been collected:

- **Age:** Speakers are assigned to four age groups. Although these are sharply defined, it is up to the surveyors themselves to determine which group each speaker belongs to during street observation. It is for the researcher to decide whether to differentiate four or more age groups and what age range should be covered by each. Here is one possibility:
 - Children: 0-14 years old
 - Young people: 15-24 years old
 - Adults: 25-64 years old
 - Elderly: 65 or older²

2 This distribution was part of the design of the 1989 Street Survey and has been maintained ever since in order to keep the data compatible.

- **Sex:** It is recorded whether speakers are female or male.
- **Presence of children:** It is also recorded whether children are present in the conversations that are registered and whether a child is speaking when the observation is made. This makes it possible, when analysing the data, to distinguish between conversations with no children present, conversations between children, conversations where adults are talking to children and conversations where children talk to adults.

These are the directly measured variables (i.e. the data that appear explicitly in the data records produced in the course of fieldwork). Other kinds of information may be obtained indirectly, such as the following:

- Is there much difference in language use between neighbourhoods within a given town?
- Do the results from surveys suggest a difference between data of language use on weekdays and weekends, or between mornings and evenings?
- Taking the results for different numbers of interlocutors, is there a difference between conversations between two people and larger groups?

2.1.2.

The place: locations and routes

In order for survey results to reflect what really happens in the streets of a town or city, how we choose the places for the survey is particularly important. For one thing, these places should make it potentially possible to measure the use of speakers pertaining to a full range of profiles; it is not a good idea to limit oneself to places only frequented by a certain kind of people because the survey is supposed to represent people of all kinds. Secondly, defining places and routes clearly is one way to make sure that the results will be statistically meaningful. Surveying busy areas and streets will make it possible to document a large number of conversations. Normally both criteria go hand in hand: areas frequented by many people are often places where people of many kinds meet.

The surveyed places are delimited and organised as routes, according to the characteristics of the town or city being surveyed. Often when a neighbourhood or district of a city is to be surveyed, a route made up of a contiguous or continuous set of streets is defined. Surveyors move from street to street, although it is not necessary to proceed in a specific direction. Other times, rather than specifying a route, geographical spaces are delimited, such as a large square, a playground, the area around a station and so on. These are all areas of special interest where surveyors should move about from place to place even when not changing streets.

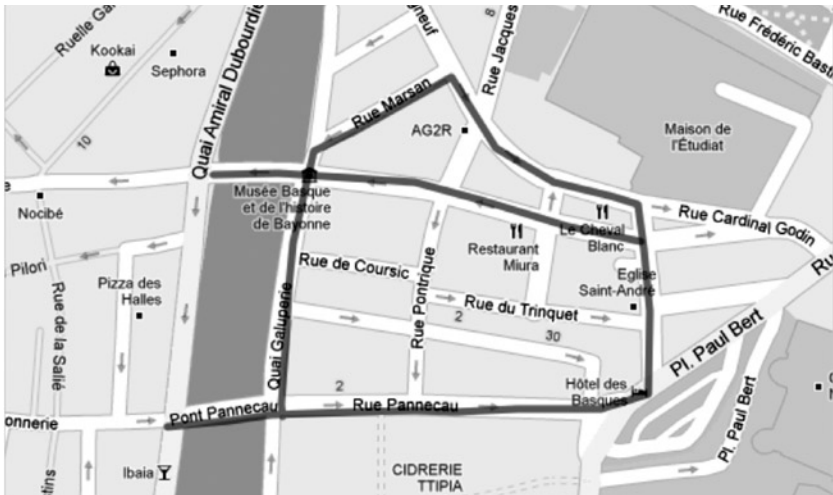
Surveyors will move along the specified routes or within designated areas, at the specified times, collecting data about the conversations that occur around them. When specifying the routes or areas to be surveyed, their extent and other characteristics should be taken very much into account because it is necessary to ensure that the real situation is collected properly. The following phenomena should be taken into consideration:

- Specifying too small a route or area may result in observers becoming too conspicuous. If surveyors spend too long collecting data in the same spot this may attract the subjects' attention. Designating too large a route or area, on the other hand, runs the risk of including points with widely differing characteristics.
- Putting together places that are meeting points for speakers with very different profiles or performing very different activities can distort the results.
- Unless the length of time spent in a part of a given route is kept constant from survey to survey (if, that is, the area of the route is not allotted the same time from one survey to the next), results may vary considerably.

Two aspects must be taken into account when determining the size of the route or area, then: on the one hand, the problems and dangers of having too short a route, and on the other, the distortions that may ensue from having too long a route. It may be necessary to establish one or more routes depending on the area (e.g. the town or city) that one wishes to study.

Here in Figure 2 is a route in Bayonne, for example:

Figure 2. Route in the “Baiona ttipia” district of the city of Bayonne



2.1.3.

Times and duration of survey sessions

Two aspects of timing need to be specified at the design stage. It must be decided how long the total duration of a given fieldwork session is to be. Also, the schedule must be established, e.g. whether observations should be made in the morning, in the afternoon, and so on. These decisions must be made in accordance with the two main criteria involved in specifying the survey area: (1) potentially including speakers of all profile types in the survey, and (2) observing the number of conversations needed to make the results statistically significant.

In order for the survey results to reflect what happens on the streets faithfully, it is desirable for observations to focus around times when many

conversations take place in the street. Thus we need to know the times of the day and the week when people go out in order to specify good survey times. The aim is to perform the survey at the times when most people are out.

However, it is not enough to identify the busiest times of the week and limit observation sessions to those times, because if we did that we would run the risk of leaving some parts of society out of the survey. Not only do we want to identify the busiest times but also to include times when a variety of different activities take place outside: weekdays, holidays, times when people go shopping, when children are out playing, and so on.

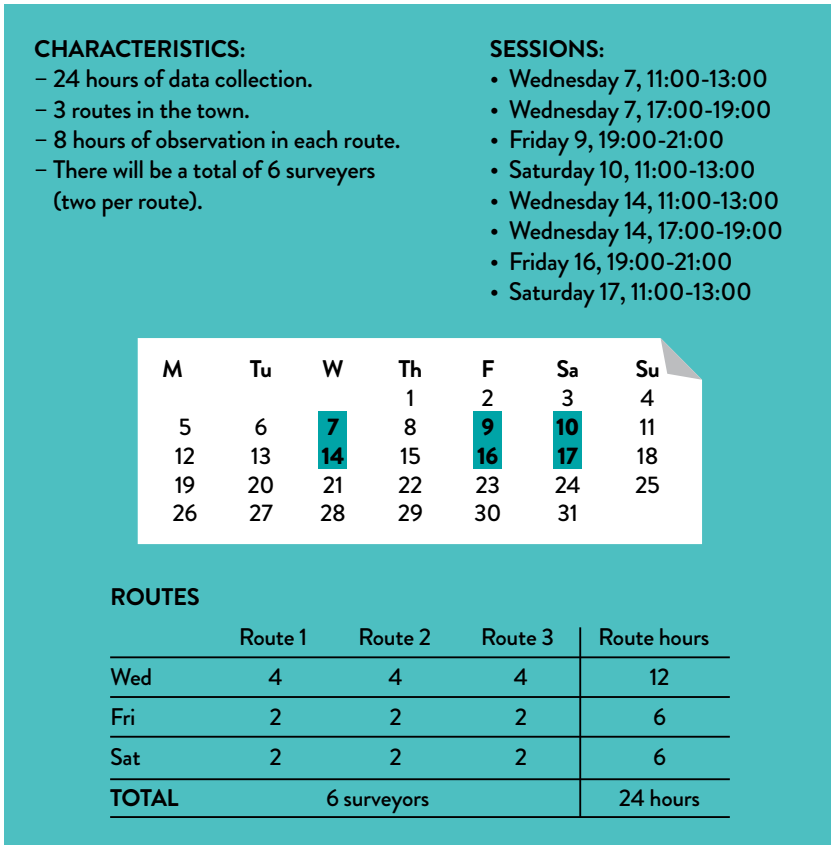
The ultimate purpose is to obtain information about the language use of people with all sorts of profile in conversations observed in the streets: men and women, people in every age group, all social classes and so on.

It is recommended that a survey in any given town should be divided into sessions of at least two hours' duration, because usually if sessions are shorter than this more survey sessions would be necessary to obtain data for a given number of hours and that would involve more work.

Furthermore, it is recommended that there should be at least two survey sessions on each day of the week and time of day. This is a precaution against the danger of data for a town being skewed by the occurrence of an atypical day at a given place and time. Hence we suggest doubling up survey sessions in the design.

To understand better these two criteria (sessions lasting at least two hours and doubling sessions), let us take the example of the session schedule shown in Figure 3:

Figure 3. An example of a session schedule



From the data collected in these sessions, totalling 24 hours altogether, we will calculate the degree of language use in the town. In this way we will obtain a snapshot of the use of the language from place to place. In order to plot the development of language use over the years, we will perform another survey after a period of time (for this purpose there should be a lapse of at least one year), with repetition every aspect of the first survey's design. For the data to be fully comparable it is essential that there should be no changes in the routes, days and times between the surveys.

2.1.4.

The number of conversations surveyed: the sample and reliability

In order to be able to consider that the data of language use obtained through observation are reliable, it is important to develop a design that is adequate in terms of variables, time and space, and to register as many conversations as possible. How many conversations are needed to yield valid results?

In quantitative research techniques, it is habitual to state margins of error and measures of reliability based on the data that is presented. When preparing a design it is necessary to know what relation there is between the levels of error and confidence and the sample from which these will be obtained. Often, for example in surveys covering a whole region, a calculation is made of the sample size necessary to produce a particular level of error and of confidence in a random sample for infinite populations.

In the case of street use surveys through observation, it is not possible to employ the standard formulae used in other techniques to calculate the sample, so a mathematical model for the calculation of the sample has had to be specially developed for observational use surveys. A specific mathematical model was developed under the guidance of Prof. Yosu Yurramendi, the principles of which are set out in a report published in 2009 (Yurramendi & Altuna, 2009).

Despite this book's limited scope, it will be found useful to explain a few notions in order to help understand how the sample size is calculated.

In our mathematical model the surveyed sample and the population it represents are made up of conversations. We record a subset of all the conversations that could take place on the street. The relation between the conversations that could possibly occur and the ones that we record indicates the degree of reliability of the results we obtain.

The number of conversations that might occur in the streets of a town is in principle infinite, but for the purpose of calculating the sample a figure is taken as a reference by assuming certain limits. For instance, the population of the town is adopted as a departure point. Of course the total number of people who may be observed on a town's streets is not equal to the town's total population. It is clearly not usual or normal for all of a town's inhabitants to be in the street. Therefore, even though it would be most exact when calculating the population that is in the street on any given occasion to add the number of people from outside the town who are present and to subtract the number of members of the town's population not present in the town's streets, the town's population figure is taken as a practical reference for the purpose of calculating the sample.

In order to be able to calculate the number of possible conversations, apart from the population size a way must be found to quantify the number of possible conversations between individuals. To this end three values need to be estimated in the construction of this mathematical model.

The first of these is the average quantity of personal relations. In the majority of cases by far, conversations in the street occur between people who are in a close relationship. For instance, typically these could be members of the same family, friends, workmates or neighbours. If we could determine the average size of the each person's network of relations of this type, we would have another piece of information that would be useful for establishing the number of conversations that might take place in the street. Having consulted the social sciences literature on this subject, a standard average value for the size of personal relationship networks was incorporated into the mathematical model for calculating the sample. The main reference number used in our study is called *Dunbar's number*³.

The second value is related to the number of interlocutors in conversations occurring in the street. As we have said, we calculate the sample

3 http://en.wikipedia.org/wiki/Dunbar%27s_number

This number represents a sort of cognitive limit, and suggests that a person is capable of maintaining enduring relationships with about 150 people.

in terms of a number of conversations, since these are the basic unit of language use surveys. But as we have also seen, our point of departure for the calculation of the sample is the town's population figure. To get from a number of people to a number of conversations, we need to know how often people tend to engage in two-person conversations, three-person conversations or conversations in larger groups. A study was carried out of the number of interlocutors that conversations in the street tend to have, yielding a value which was found useful for calculating the sample.

The third value is the factor of language competence. Obviously people who cannot speak a language are not able to use that language. Thus, the number of conversations that can occur in a given language in the street is also limited by people's language competence. People first need to be able to speak a language before they can have any sort of conversation in it. Hence the percentage of the population that knows each of the languages considered in the survey is incorporated into the calculation of the sample.

These, then, are the values that are considered when calculating the size of the sample. According to those values we calculate the amount of confidence or margin of error of a survey.

Table 3 shows calculations with a confidence level of 95% and margins of error of ± 5 and ± 2 as a reference for the sample that needs to be surveyed according to the model.

Table 3. Margins of error and samples according to population size

| TOWN'S POP. | PERC. OF BILINGUALS | MARGIN OF ERROR | LEVEL OF CONFIDENCE | CONVERSATIONS REQUIRED |
|-------------|---------------------|-----------------|---------------------|------------------------|
| 2000 | 10% | ± 2 | 95% | 864 |
| 10000 | 10% | ± 2 | 95% | 864 |
| 25000 | 10% | ± 2 | 95% | 864 |
| 100000 | 10% | ± 2 | 95% | 864 |
| 300000 | 10% | ± 2 | 95% | 864 |

| | | | | |
|--------|-----|----|-----|------|
| 2000 | 25% | ±2 | 95% | 1797 |
| 10000 | 25% | ±2 | 95% | 1800 |
| 25000 | 25% | ±2 | 95% | 1801 |
| 100000 | 25% | ±2 | 95% | 1801 |
| 300000 | 25% | ±2 | 95% | 1801 |
| 2000 | 50% | ±2 | 95% | 2395 |
| 10000 | 50% | ±2 | 95% | 2400 |
| 25000 | 50% | ±2 | 95% | 2401 |
| 100000 | 50% | ±2 | 95% | 2401 |
| 300000 | 50% | ±2 | 95% | 2401 |
| 2000 | 80% | ±2 | 95% | 2208 |
| 10000 | 80% | ±2 | 95% | 2212 |
| 25000 | 80% | ±2 | 95% | 2213 |
| 100000 | 80% | ±2 | 95% | 2213 |
| 300000 | 80% | ±2 | 95% | 2213 |
| 2000 | 10% | ±5 | 95% | 139 |
| 10000 | 10% | ±5 | 95% | 139 |
| 25000 | 10% | ±5 | 95% | 139 |
| 100000 | 10% | ±5 | 95% | 139 |
| 300000 | 10% | ±5 | 95% | 139 |
| 2000 | 25% | ±5 | 95% | 289 |
| 10000 | 25% | ±5 | 95% | 289 |
| 25000 | 25% | ±5 | 95% | 289 |
| 100000 | 25% | ±5 | 95% | 289 |
| 300000 | 25% | ±5 | 95% | 289 |
| 2000 | 50% | ±5 | 95% | 384 |
| 10000 | 50% | ±5 | 95% | 385 |
| 25000 | 50% | ±5 | 95% | 385 |
| 100000 | 50% | ±5 | 95% | 385 |
| 300000 | 50% | ±5 | 95% | 385 |
| 2000 | 80% | ±5 | 95% | 354 |
| 10000 | 80% | ±5 | 95% | 355 |
| 25000 | 80% | ±5 | 95% | 355 |
| 100000 | 80% | ±5 | 95% | 355 |
| 300000 | 80% | ±5 | 95% | 355 |

As already mentioned, a fuller explanation of the mathematical model may be found in Yurramendi eta Altuna (2009). However, an application for calculating the sample without entering into such details, using basic survey data (population of a town and the percentage of bilinguals) is available on the Internet at www.soziolinguistika.org/lagina.

2.1.5.

Trends in language use: continuity from survey to survey

Of the various benefits of use surveys, one stands out in particular: the possibility of studying trends over time. Even a single snapshot provides valuable information. But it is even more useful to be able to track the tendency over time.

By repeating the survey at regular intervals we will obtain a dynamic picture which can reveal trends and tendencies. If our purpose is to exert an influence on language use, we need to obtain and analyse information about the language's changing panorama.

However, we cannot benefit from this option without methodological rigour. Two surveys can only be compared if the same design features are replicated in both of them. This point must be kept in mind by the survey's designers: the design should be such as to ensure that it will be possible to replicate it in the future.

For the data to be directly comparable, the features of the survey should not vary; the same survey locations or routes, days and times need to be maintained from one survey to the next.

Not only the present but also the past and the future should be taken into account. Otherwise there is a danger that mistakes made now will be per-

petuated in the future. The need to maintain the same design may make it impossible to correct such mistakes later on.

Therefore, continuity from one survey to another is of fundamental importance and increases the value of street surveys, but it is essential that this option should not be obstructed by design errors.

2.2.

Fieldwork

In surveys of language use through observation it is the people who carry out the survey who are responsible for the fieldwork. Therefore it is very important to be clear about all possible eventualities that may arise in the course of street observation. The results of data collected by different surveyors cannot be mutually comparable unless these criteria are precisely defined.

The surveyors' job is to observe everything that goes on at the times and in the places specified in the survey design. They must record on data forms or on a phone application the characteristics of the conversations they observe within the time interval and along the route that has been defined. We will now look at all this in more detail.

2.2.1.

Data records

Figures 4, 5 and 6 show the layout of a survey form on language use in the street. Figure 4 has been divided into six sections in order to explain how the data collection is done:

- Information is noted at the top (1) identifying each survey session: the town, route, day, time and the surveyor's name.
- In the main table, each conversation is assigned a line and each is identified by a number (2).
- The language of the conversation is noted by checking a box in the column to the right of the number (3).
- The middle part of the table (4) contains a block where the main characteristics of the speakers taking part in the conversation are noted down. Interlocutors are classified in four main columns according to their ages. Within each column a grey square and a blue square distinguish female and male interlocutors. The number placed in each square indicates the number of participants in the conversation of that age group and sex category.
- Finally, in the rightmost column (5), the surveyor indicates in the boxes whether the children were speaking or not in each conversation while it was being observed.
- The same areas for data collection (2, 3, 4 and 5) appear on the other side of the form. But in the area corresponding to where the identification data is situated on the front side, on the back there is a place to add observations on any circumstances that might have had an influence on the survey (6).

Figure 5. Front of the model data collection form

**HIZKUNTZA
ERABILERAREN
KALE NEURKETA**

SOZIOLINGUISTIKA KLUSTERRA:
T: 943 592 556
klusterra@soziolinguintika.org

HERRIA: _____

IBILBIDEA: _____

EGUNA: _____ ORDUA: _____

NEURTZAILEAK: _____

| | EUSKARA | GAZTEKERA FRANTZESA (ipar EH) | BESTERIK | Haurrak (2–14 urte) | | Gazteak (15–24 urte) | | Helduak (24–64 urte) | | Adinekoak (65–... urte) | | Haurrak hitz egiten | |
|----|--------------------------|----------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|----------------------------|--------------------------|--------------------------|--------------------------|
| | | | | | | | | | | BAI | EZ | | |
| 01 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 02 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 03 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 04 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 05 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 06 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 07 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 08 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 09 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 10 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 11 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 13 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 14 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 15 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 16 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 17 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 18 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 19 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 20 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Emakumezkoa | Gizonezkoa

Figure 6. Back of the model data collection form

OHARRAK: _____

| | EUSKARA | GAZTELERA FRANZESA (lpar:EH) | BESTERIK | Haurrak | Gazteak | Helduak | Adinekoak | Haurrak hitz egiten | |
|----|--------------------------|---------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | | | | (2-14 urte) | (15-24 urte) | (24-64 urte) | (65-... urte) | BAI | EZ |
| 21 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 22 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 23 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 24 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 25 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 26 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 27 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 28 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 29 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 30 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 31 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 32 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 33 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 34 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 35 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 36 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 37 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 38 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 39 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 40 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 41 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 42 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Emakumezkoa | Gizonezkoa

2.2.2.

The mobile phone app

We have developed a mobile phone or smartphone version of the data form presented in the preceding section, with a special software application and interface. Figure 7 shows some screenshots.

The new application is a web application using standard, free technology which permits georeferencing; the data can be stored locally or on a central server, and a data base is automatically created.

The utility of this application may be summed up as follows:

- It can provide a pair of surveyors following the established work method with the information needed to perform a survey: pending or unfinished survey sessions, each session's route, and supplementary information (e.g. data) required for performing the survey.
- The application registers the physical location of all conversations and observations through the phone's GPS.
- The application can work with or without connection to the data network. When connected, it sends data straight to the server. When working off-line, it is capable of storing data locally until it comes back on-line. This feature solves any possible issues with 3G or network coverage.

Figure 7. Mobile phone application screens



2.2.3.

Data collection criteria

Data collection is subject to certain rules or criteria:

- Because the survey is performed through observation, unlike questionnaire-based surveys the survey team must never ask questions of the speakers in conversations either about the language they are speaking or concerning their own characteristics.
- We normally only hear one person in a conversation even though it is between two, three or more people. On the data sheet all the members of the group are registered. For instance, when four people meet, although one of them speaks and the other three are hearers, on the data sheet we will indicate that we have “heard” four people.
- Each conversation is to be covered by two surveyors who are on the street at a stipulated time and place to record information about the conversations they observe on survey forms. It is considered desirable for there to be two surveyors for several reasons.

In the first place, it is useful for two surveyors to work together to deal with various issues and difficulties that could arise. For example, there are often doubts about char-

acteristics of conversation participants such as their age or sex, or about which language they are speaking. It may also happen that while collecting data about one conversation the surveyors hear other conversations going on around them. In such cases it is possible to record the data and observe the other conversations at the same time by working as a team.

Secondly, to avoid becoming conspicuous, which is a feature of fundamental importance in observational surveys, it helps to work as a two-person team, especially as we fill in forms. It would attract attention and arouse people's curiosity if one person were to spend a long time in the street with a pen and paper in hand watching people and writing things down. It is easier for two people who are chatting together to pass unnoticed.

- An attempt should be made to cover as many conversations as possible in the survey. Provided other criteria (such as remaining inconspicuous) can be met, surveyors are to observe as many groups as possible and register the data on their data forms or mobile phones. This is the way to obtain the most reliable results. However, we must realise that in places or on routes where there are a lot of people it is not going to be possible to record *all* the conversations.
- Only data about conversations of people

on the street must be collected. People in shops, bars and restaurants, or the entrances to residences are not to be included.

- Often people in the street are not evenly distributed around the space chosen for observation. In such cases it is best to walk the whole route once or twice during the survey session, but the time spent by the surveyors in different parts of the route need not be evenly distributed. Ideally more time should be spent in areas with more people in order to be able to register more conversations. Except as necessary to avoid leaving the observation area, one should not pass through the same places in the same direction more than once; it is preferable to reverse the direction. It may help surveyors to remain undetected and unnoticed by people in the street when passing along a street more than once if they switch to the other side of the road, or stop for a while somewhere along the route. It may be a good idea to sit on a bench provided this doesn't stop them from observing properly and recording plenty of conversations.
- When people in the street address the surveyors, those conversations should not be recorded. The purpose is for the surveyors to have no influence whatsoever on the conversations that are recorded. This may be considered a safety measure to ensure there is no cheating!

- Establishing the limits between different conversations is important. A change of participants and a change of language are events that mark the beginning and end of discrete conversations for observational purposes. If the same group (with the same members) is observed more than once speaking the same language, this should be registered as a single conversation in the survey. If the same group is observed speaking a *different* language, that will be registered as a distinct conversation. And if a change occurs in the composition of a group under observation (through a member dropping out of the group or the arrival of a new participant), even though they continue speaking the same language, this will be registered as a new conversation.
- Doubtful cases will not be recorded. Surveyors may be uncertain about what they have heard and be unsure what language someone is speaking. This can happen for various reasons, such as difficulty hearing what is said or because both languages were mixed up in a complicated way in the same sentence. In such cases, surveyors are recommended not to register the conversation in spite of the general goal of trying to record as many conversations as possible.

Surveyors are given the most important instructions in writing, together with the data forms (see Figures 8, 9 10 and 11). A brief audio text with those instructions has also been prepared for mobile phone.

Figure 8. General notes on how to fill in the data form (I)

LANGUAGE USE STREET SURVEY

What should be measured?

Language use among people who are in the street only. Closed-in places (such as bars, shops, sports installations, houses, factories, transportation, premises of any kind, etc.) are therefore excluded from this survey.

Observe, do not ask!

As we walk along the street we must not ask people what language they normally speak in or what language they are speaking in now. **Do not ask, only listen** and record what you hear, without telling people anything. Our purpose is to record the linguistic reality on the street as objectively as possible. The people in the street must not know that a study is being conducted.

Surveyors

Surveyors will work **in pairs**. A single person can also work alone, but it is better to work in pairs. Reasons:

- It's easier. One person can listen to people and tell the other one what to note down.
- It is much easier for two people to disguise what they are doing, and it is very important that nobody should realise that we are conducting a survey. Otherwise people may change their behaviour by keeping quiet, changing language, being annoyed, etc.

The route

Each pair of surveyors has a **set route** (on the back of the notebook there is a map of the route with street names, days and times of observation). The route will consist of the town's busiest streets, normally in the centre of the town (understanding "centre" in a broad sense): shopping areas, areas of bars, playgrounds, promenades and parks, etc. The route must start and end at the same point so that the surveyors can repeat the route as many times as needed during the the session's two hours' duration.

Each pair is to walk along its assigned route. It is fine for them to stop occasionally and sit on a bench to observe, and in fact this will necessary sometimes to disguise what they are doing. But it is not desirable to spend too long sitting down since the point of the exercise is to survey the whole route.



Figure 9. General notes on how to fill in the data form (II)

LANGUAGE USE STREET SURVEY

How to record data

Data should be recorded on special forms (see the example below). **Each conversation will be registered on one line**, and four kinds of information are relevant:

1. **Language:** "Basque", "Spanish/French (northern BC)" and "other" (other languages). In the southern Basque Country we will mark any language other than Basque or Spanish in the "other" column. Similarly, in the northern Basque Country "other" will mean any language except Basque or French.
2. **Age group:** "Children (2-14)", "young adults (15-24)", "adults (25-64)" and "elderly (65 and over)". Surveyors should guess people's approximate age.
3. **Gender:** Note down the sex of interlocutors, "female" or "male".
4. Lastly, we note **whether children speak**, i.e. whether the children of between 2 and 14 years of age are active speakers, "yes" or "no".

Here are **THREE EXAMPLES** to illustrate how to conduct the survey:

- Two children (a girl and a boy) are speaking Basque (**conversation 01**):
 1. Put an "X" in the Basque box.
 2. In the children's column write "1" in the female box and another "1" in the male box.
 3. Put an "X" in the "yes" box in the "Children speaking" column.
- A husband and wife are talking to each other in Spanish (or in French in the northern Basque Country); a 5-year-old daughter and a 16-year-old son are with them but they are silent (**conversation 02**):
 1. Put an "X" in the Spanish box.
 2. In the children's column write "1" in the female box; in the young adults' column write "1" in the male box; and in the adults' column write a "1" in the female box and another "1" in the male box.
 3. Put an "X" in the "no" box in the "Children speaking" column.
- Five young people, two girls and three boys, are speaking Spanish (**conversation 03**):
 1. Put an "X" in the Spanish box.
 2. In the young adults' column write "2" in the female box and "3" in the male box.
 3. Put an "X" in the "no" box in the "Children speaking" column.

Figure 10. Example of how to fill in a data form

**HIZKUNTZA
ERABILERAREN
KALE NEURKETA**

SOZIOLINGUISTIKA KLUSTERRA
T: 943 592 556
klusterra@soziolinguintika.org

HERRIA: Bilbo
IBILBIDEA: Abando 3
EGUNA: Urriak 08 ORDUA: 18:00-20:00
NEURTZAILEAK: Ane Larrañaga
Xabier Artola

| | EUSKARA | GAZTEKERA FRANTZESA (Ipar EH) | BESTERIK | Haurrak (2-14 urte) | | | | Gazteak (15-24 urte) | | | | Helduak (24-64 urte) | | | | Adinekoak (65-...urte) | | | | Haurrak hitz egiten | |
|----|-------------------------------------|-------------------------------------|-------------------------------------|---------------------|----|-----|----|----------------------|----|-----|----|----------------------|----|-----|----|------------------------|---|--|-------------------------------------|-------------------------------------|--|
| | | | | BAL | EZ | BAL | EZ | BAL | EZ | BAL | EZ | BAL | EZ | BAL | EZ | | | | | | |
| 01 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | / | / | | | | | | | | | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 02 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | / | | | | / | / | / | | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 03 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | 2 | 3 | | | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 04 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | | | | | | | | | | | 2 | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 05 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | | | | | | | | | | 2 | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 06 | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | | | | | | | / | | / | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 07 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 2 | | | | | | / | | | | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 08 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | / | | | / | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 09 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | | | | | | / | / | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 10 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | | | | 2 | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 11 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | | | 4 | | | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 12 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | / | | | | | / | | | | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 13 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | | | | | 3 | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 14 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | 2 | | | | | | | | / | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 15 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | | | | 2 | / | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 16 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | / | / | | | | | | | | | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 17 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | | 2 | / | | | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 18 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | / | | | | | | | / | / | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 19 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | 3 | | | | | | / | | | | | | | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | |
| 20 | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | | | | | / | / | | | | | | | | | | <input type="checkbox"/> | <input checked="" type="checkbox"/> | |

Emakumezkoa Gizonezkoa

2. A GUIDE TO STREET USE SURVEY

Figure 11. General notes on how to fill in the data form (III)

LANGUAGE USE STREET SURVEY

Notes:

- This is a point that needs to be explained: when four people walk along talking, **we will normally hear one** of them because the other three will be listening (paying more or less close attention; that is of no concern to us). In this case **we will say on the form that we have heard four people**.
- For as long as the survey session lasts we will listen to **as many conversations as possible**, and fill in as much data as we can.
- **When starting a survey session** we will start a new data form, and fill in the top of the form with information about session including the date and time.
- **Within one survey session and conversation** do not observe the same person more than once. If we hear the same person speaking the same language with the same people, we will only note them once. If however one of the participants changes (because someone else comes along or someone leaves), that will be another event and will therefore be recorded.
- **In bilingual conversations** (when someone speaks Basque and someone else replies in Spanish or vice-versa) we will register the conversation twice, once in Basque and once in Spanish. Then we will join the two conversations with an arrow. See conversations 04 and 05 in the data form above (In the example, two elderly men are speaking, one in Basque and the other in Spanish).
- People who talk to the **surveyors** will **not** be recorded.
- **Do not record doubtful cases:** when we are not sure which language we have heard, skip it.
- **Never ask questions:** do not ask people their age, which language they usually speak, or what language they are speaking at the time. Do not ask anything because we are not conducting a questionnaire; our job is to survey through observation, so we are to note what we hear.
- There is room on the back of the form for **notes**, where observations may be made about unexpected things that happen during the session, surveyors' doubts or other pertinent information.
- **A final note:** Although it is disheartening not to hear much Basque spoken, **always record truthfully**. This survey is a **very important means of learning about the situation of Basque** and we want to know what the reality is. Please observe this strictly!



2.3.

Processing the data

In order to be able to describe and interpret the linguistic reality of a town's streets, the raw data collected in fieldwork has to be subjected to statistical processing.

It is important to distinguish between the collection and processing of data. A very basic kind of data processing is presented here which aims to produce answers to basic questions about language use. However, it is also possible to treat the data collected in more sophisticated ways, and the data are available to anyone who wishes to perform such statistical studies.

To answer the basic questions that are asked about the measurement of street use, it is sufficient to carry out some simple calculations and combinations. The percentage of people who speak in one or another language can be found through simple arithmetic on the information that has been collected. The results can also be presented in accordance with the variables that were incorporated in to the design. For example, we can find:

the proportion of all the observed speakers...
the proportion of observed children...
the proportion of observed young adults...
the proportion of observed adults...
the proportion of observed elderly people...
the proportion of observed males...
the proportion of observed females...
...who speak each language on the town's streets.

The proportion of observed speakers speaking each language...
...on each route.
...in each session.

...in conversations between children.
...in conversations without children.
...in conversations involving children and other age groups.

Let's look at examples of how to calculate results according to different variables.

- (Proportional) use of language X in the town:

$$\frac{\text{Number of persons observed in conversations speaking language X}}{\text{Total number of persons observed in the town}}$$

- (Proportional) use of language X in the town by a given age group:

$$\frac{\text{Number of young people observed in conversations speaking language X}}{\text{Total number of young persons observed in the town}}$$

- (Proportional) use of language X in the town by females:

$$\frac{\text{Number of females observed in conversations speaking language X}}{\text{Total number of females observed in the town}}$$

To process the data it is advisable to use a data base or a spreadsheet. It is up to researchers to decide what computing resources are best suited to their purposes.

While the calculations that must be made to obtain such results from the raw data are simple enough, certain doubts tend to arise at this stage.

Imbalance between subsets of the population in the survey sample

The town's population data are used as a reference point for the calculation of the survey sample. We have also had available to us figures about the subsets of the whole population prior to collecting our data. By "subsets" we mean the groups that arise from a classification of the population according to the variables we have used, such as female, male, young adult and so on. Provided the fieldwork was carried out in accordance with the guidelines laid down in the design stage, we will have registered a quantity of conversations (a sample) that is adequate to provide a measure of reliability. A doubt arises when the subsets of the population have a different proportion in the whole population and in the sample. What should we do if a subset of the sample is over-represented or under-represented in relation to the original population? How should we deal with this group's value when calculating the overall data? Should we weight them differently?

The answer is no. As we have observed, the key to an adequate design is to make sure we are able to include in the survey anyone who may potentially be in the streets through good choices of times and routes. If we design with this criteria in mind the data we obtain will reflect the reality of the streets in the town in question without needing any weighting. If, for instance, the data obtained for a subset are proportionally lower than it ought to be in terms of the town's overall population data, what this may mean is that the people in this subset tend to be in the street less.

Let us discuss this issue of imbalance through an example.

Suppose we have seen in the population data that a town's inhabitants are divided up by age groups as follows: 17% children, 28% young people, 33% adults and 22% elderly.

In our own data, however, speakers divide into age groups as follows: 22% children, 18% young people, 35% adults and 25% elderly.

As will be seen, in the survey data the young people's age group is represented in a smaller proportion than in the town's population data. When we see this we should ask whether a neighbourhood or a time where or when a lot of young people go out has been omitted in the survey design.

This raises the question as to whether we should weight data for language use of each age group according to the relative proportions of each group in the population (i.e. 17-28-33-22)?

The answer is no, because the objective established from the start was to measure real language use in the street. It may be the case that there is a difference between age groups in terms of their presence in the street; for example, children may spend longer in the street than old people. Given this, if we were to resort to weighting, old people's conversations would be attributed greater importance than those of the children in our survey. That would not meet our criterion of measuring accurately the use of Basque in the street⁴.

4 As a by-product we *can* find out what the language's street presence by age groups is.

Interpreting the results

Once the data have been collected and undergone initial processing, it is up to the researcher to subject the results to further processing in order to draw whatever conclusions are possible. This process of analysis and reflection is known as *interpretation of results*.

At this stage, it is very important not to get lost amidst all the numbers. The data obtained are so numerous and varied that it becomes necessary to analyse them in a highly orderly and structured way. Otherwise there is a risk of coming up with results that are imprecise, or worse still, simply inaccurate.

The results cannot be interpreted without keeping methodological principles clearly in mind. Decisions at the design stage are highly relevant to analysis of results.

In the methods used to measure language use and respecting the interpretation of the results, the following points are of special importance.

Not individuals' language but groups' language of conversation

Spoken use of a language is by nature a collective matter. Oral use of a language does not depend on the individual, but on the group. Knowledge of the language, on the other hand, is a matter for the individual.

When giving general results on use we should bear in mind that what we are measuring is not so much the language of individuals as the language of conversation of the group. Where street use is concerned it is conversations that are observed and counted, and those conversations always take place, by definition, between two or more people. Consequently, when

we observe a conversation in a group we take into account the language of the person who is speaking at the time, on the assumption that if this is the language they are speaking, then they also speak it at other times, and assuming also that if it were not this person but another member of the group who were talking, they would speak the same language⁵.

In measurements of street language use, we observe each conversation group for such a short time that it does not permit us to discover the behaviour of every member of the group. Nevertheless, our procedure is founded on a solid methodological basis. The groups of speakers that we observe in the street tend to have a fixed pattern of linguistic behaviour which seldom varies. The people we observe are in most cases mutual friends, relatives or partners. In such cases, depending on their linguistic habits it is normally already decided (perhaps unconsciously) what language the conversation is going to be in.

A given language is chosen among a given group of speakers, and thereafter conversations between those people will in most cases by far take place in that language. It is true that linguistic habits do change, but such a change does not happen “on its own” or easily, and this rigidity in linguistic habits between particular speakers results in their enduring nature.

Percentage by individuals

The main result of the measurement of language use will be expressed as a percentage. This use statistic can be presented in two ways: according to the language of conversation (out of X conversations, Y are in Basque, Spanish, French, Catalan...), or according to the individuals (out of X people whom we have observed, Y were speaking Basque, Spanish, French or Catalan).

5 This is not the same situation as in bilingual conversations, where one person speaks one language and the other people answer in another, or where, in a conversation between three people, two of them speak one language and the third another language.

One important advantage of giving the information in terms of individuals is that the characteristics of the speakers can be analysed, calculating indices of language use according to the variables. For example, if we specifically want to find the use of children in a given place, we need to know whether a speaker is a child or not.

The favoured option and usual practice of the Soziolinguistika Klusterra is to calculate percentages of use on the basis of individuals. There are two main reasons for this. One is that this is how the first surveys were calculated and it is important to maintain continuity in order to be able to study trends in use over time. The other is that, as we have seen, it is necessary for results to be given in terms of individuals in order to obtain use data according to speakers' characteristics (such as age or sex).

However, the use data of surveys in 97 towns carried out in 2011 were calculated two ways, by conversations and by individuals. In most cases the two results are quite similar, within a few decimals of each other. The biggest difference is 3.7 points. In most towns — 64 out of the 97 — there is a difference of less than one point. In 29 towns there is a difference 2 or 3 points. In the remaining four towns it is between 3 and 3.7. A special study in greater detail is needed to conclude that both ways of calculating yield similar results, but it does appear that calculating the percentage of use according to either the individual or the conversation does not make a great deal of difference. We should also bear in mind that if the statistic is calculated by the individual we can obtain data of use according to speakers' characteristics.

Level of use versus level of knowledge

The fact that a given person chooses to use one language or another depends on many factors, a very important one obviously being knowledge of the language. But conclusions drawn from simple comparisons between the amount of knowledge of a language and the proportion of use are often inaccurate. When conclusions from the chief Street Survey of the Basque

Country are presented in section 3.3, the relation between knowledge and use will be explained intelligibly by means of an example.

Town use versus citizens' use

After carrying out a survey of a town, we must take into account the fact that it cannot be assumed that all the people observed necessarily come from that place. All we observe is someone speaking at a given time and place. Thus we cannot be certain that what we record in a survey is the language use of *local* inhabitants. It would be more accurate to say that what we analyse is the use recorded in the town's streets.

A changeable universe

The speakers surveyed when performing a street use survey are not always the same ones each time, and we cannot predict who we will find in the street, and when. Therefore, although the data obtained are meaningful, the variable under study — language use in the street — are slightly changeable. That is one of the reasons why it is particularly desirable to carry out surveys with a certain frequency and trace the general trends. In any case, the method followed provides a guarantee of the validity of the general conclusions drawn from the results of each survey.

A guide to language
use observation

SURVEY METHODS





3.

OBSERVATIONS OF LANGUAGE USE IN THE BASQUE COUNTRY



Now that the principles and guidelines of this method have been presented, let us look at a particular case of its application by turning to the 2011 survey of language use in the Basque Country, which we shall now review.

We will begin with a brief overview of the Basque Country in order to provide some context for the study.

3.1.

Euskal Herria, the land of the Basque language

Let us quote from articles about Euskal Herria and the Basque language in the Basque-language edition of *Wikipedia*, the free, community-edited encyclopedia⁶:

⁶ In the English Wikipedia, see http://en.wikipedia.org/wiki/Euskal_Herria and http://en.wikipedia.org/wiki/Basque_language. The content and wording of Wikipedia articles vary over time and from language to language: here we offer in English translation excerpts from the Basque version of the encyclopedia in its current form: see http://eu.wikipedia.org/wiki/Euskal_Herria and <http://eu.wikipedia.org/wiki/Euskara>.



Euskal Herria is a European country. Historically it is the land of Basque speakers and the Basque language, spanning the border between France and Spain in the western Pyrenees and extending along the coast of the Bay of Biscay. It comprises the territories or provinces of Araba, Bizkaia, Gipuzkoa, Lapurdi, Low Navarre, High Navarre and Zuberoa. Euskal Herria has a total area of 20,950.3 sq km, and its population in 2006 was 3,005,670.


Basque, Castilian, French and Gascon are spoken in Euskal Herria.

Castilian and French predominate, the former in the south and the latter in the north. However, approximately one-third of its inhabitants speak Basque. Gascon is spoken in a few areas around the Greater Bayonne region (BAB) and the north of Low Navarre, but its presence is weak.

The official status of the Basque language varies among the three parts of the territory: in the Basque Autonomous Community (BAC), it is a co-official language in the three provinces; in Navarre, it is only co-official in the so-called “Basque-speaking zone”, semi-official in the “mixed zone” and has no official status in the “non-Basque-speaking zone”; in the northern Basque Country it has no official status.

The Basque language

*Basque or Euskara (possibly from Old Basque *enu(t)si (‘say’) + -(k)ara (‘way’)) is the language of Euskal Herria. It is an ergative language*



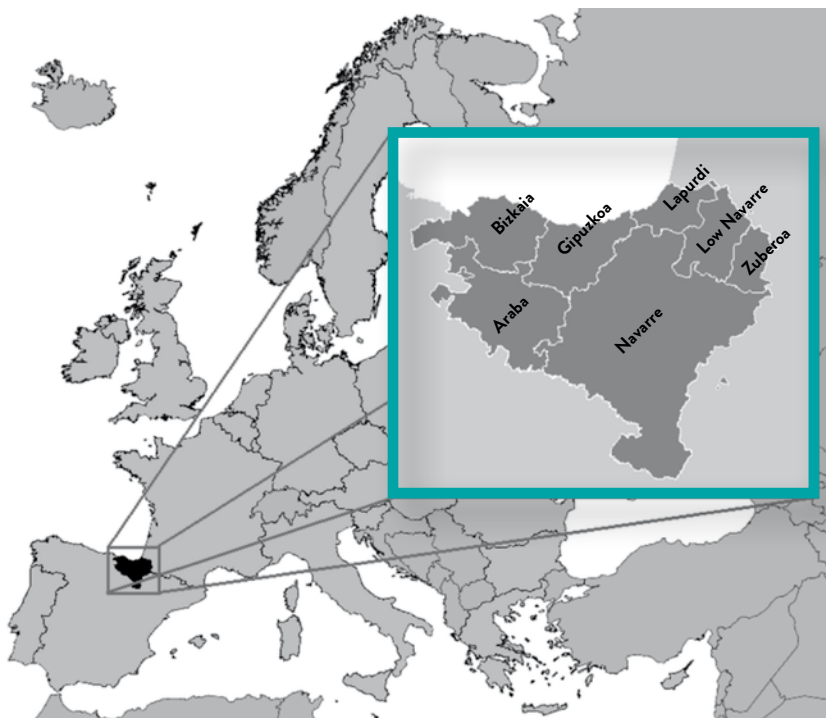
and an isolate. The term for speakers of Basque is euskaldun. Today Basque has become a minority language in Euskal Herria itself, where the newer Castilian and French languages have come to predominate.

Euskara's most ancient roots (from before the Roman Empire) have shown up throughout Aquitaine and along both sides of the Pyrenees from the Bay of Biscay to Andorre, and in the Middle Ages there is documentation of its having been spoken at least as far south as Rioja and the northeast of the province of Burgos. It has been subjected to vigorous oppression, gradually losing more and more ground. In High Navarre this process was very notable over recent centuries. In the late nineteenth and the early twentieth centuries there was a degree of revitalization due to the influence of intellectuals and politicians such as Arturo Kanpion and Sabin Arana, and this was closely linked to the growth of Basque nationalism. After suffering renewed persecution in the era of the Franco regime (1936-1939), the language began a comeback from the mid-twentieth century onwards helped by the establishment of a written standard.

Starting in the 1980s the Basque language has achieved acceptance by official institutions, despite the territory's considerable fragmentation. The Gernika Statute designated it as the Basque Autonomous Community's own language and gave it an official status together with Castilian in the provinces of Araba, Bizkaia and Gipuzkoa (but not in Trebiñu and Villaverde Turtzioz). In High Navarre, in accordance with the Law of

the Basque Language, it is only co-official in the northwest of the province (the Basque-speaking zone). In Pamplona and other important towns in the central region (the mixed zone), it meets less official acceptance and faces greater obstacles. In the south of Navarre (the non-Basque-speaking zone), Basque has no recognition. In the northern Basque Country, the Basque language is not recognised by the public institutions with authority, the only officially recognised language being French.

Figure 12. Location of Euskal Herria



3.2.

Background and development of the study

The most recent survey was performed in 2011. This was the sixth survey since the first one was carried out in 1989. Briefly, these were the characteristics of the survey:

General objective:

To obtain data of language use in the Basque Country and its regions.

Fieldwork:

Ninety-seven towns were surveyed across Euskal Herria. The localities surveyed were chosen according to various criteria including the need to coincide with places surveyed on previous occasions and the need to cover all the sociolinguistic zones of all provinces.

The survey was conducted in the autumn months of September and October, in the centres of towns, for six hours in three separate sessions.

Table 4. Street Survey of Language Use in the Basque Country. Technical specifications

TECHNICAL SPECIFICATIONS

UNIVERSE: Conversations that can be heard in the streets of the Basque Country.

SAMPLE: 154,277 observed conversations among 363,616 individuals.

CALCULATION OF MARGIN OF ERROR (attributable to completely random samplings): $\pm 0.4\%$ for the whole sample of conversations, 95.0% reliability, $p=13.3\%$ (percentage of conversations in Basque out of all conversations observed).

Main conclusions from the study

We will pick out seven conclusions from the results of the observations of language use in Euskal Herria:

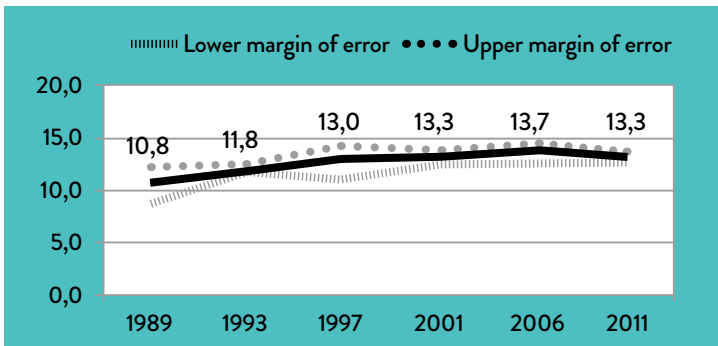
Conclusion 1: The percentage of use of Basque in the latest street survey of 2011 was 13.3%.

The use of Basque on the street has not varied from ten years ago, and seems to have stagnated.

Over the twenty-two years since the first Street Survey of the Basque Country was carried out, the use of Basque has increased by 2.5 percentage points on average for the whole of Euskal Herria. In the first survey of 1989 the recorded use was 10.8%.

But if we take the period of the last ten years, use seems to have stagnated or fallen slightly. Use of Basque in 2001 was the same as in 2011: 13.3%.

Figure 13. Street use of Basque in Euskal Herria. % Basque



Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 1989-2011

Conclusion 2: The data of use are better than statistically expected.

The data of Basque language use are better than is to be expected statistically considering the number of Basque speakers. On the whole Basque speakers are loyal to their language; if not, according to the statistical odds, they would speak even less Basque than they do. This claim is based on J.L Alvarez Enparantzak (Txillardegi)'s theory of isotropic use (Alvarez, 2001).

Txillardegi started out from the proportion of Basque speakers in a given place to make calculations about the use of Basque. Precisely what this predicts is how much Basque would be used if all the Basque speakers in the place are loyal to Basque (i.e. if they speak Basque to other Basque speakers every chance they get) and if relations between all the people in the place are randomly distributed (the isotropic situation). This is called "expected use" (or "isotropic use"), meaning how much Basque would be used provided the theoretical conditions were met. It is calculated by this formula:

$$P_B = m_B (w_2 e_x^2 + w_3 e_x^3 + w_4 e_x^4)$$

where:

P_B = level of Basque use

m_B = loyalty to Basque

e_x = proportion of bilinguals

w_2 = weight of the pair

w_3 = weight of the group of three

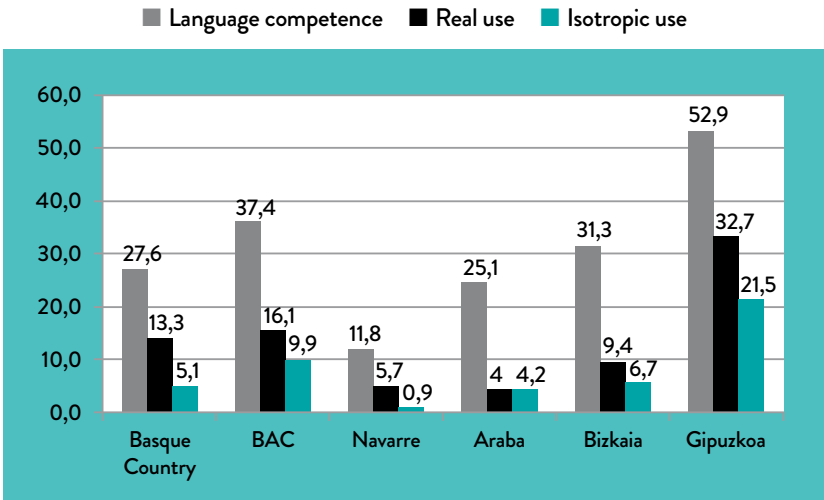
w_4 = weight of the group of four

Let us now apply this formula to Euskal Herria:⁷

$$P_B = 1,0 \cdot (0,60 \cdot 0,269^2 + 0,22 \cdot 0,269^3 + 0,18 \cdot 0,269^4) = 0,049 = \% 4,9$$

According to this calculation, the “expected use” of Basque in the Basque Country is 4.9%. Thus given random interactions between Euskal Herria’s inhabitants and assuming that Basques will speak to Basques in Basque, we can expect Basque to be spoken 4.9% of the time. However, given the very varied situations with regard to the proportion of bilinguals in Euskal Herria (i.e. their language competence), it is more realistic to limit the area and carry out studies of different regions. Figure 14 shows a comparison between real use in the street as observed in each province and the expected use.


Figure 14. Language competence, real use and expected use. Provinces of southern Euskal Herria. % in Basque



7 We have taken the proportion of bilinguals in Euskal Herria (ex) from the Sociolinguistic Survey of the Basque Government (ISL, 2011). However, it should be borne in mind that this figure refers to inhabitants aged sixteen or over. The real figure for the proportion of bilinguals is higher than 27%.

It is important to realise that this is a theoretical point of reference; relation networks are not random, of course. Everyone does not interact with everyone else: we interact more with the people around us and people of our own age. Needless to say, in large cities we do not interact at all with most of the people there. We also know that Basque speakers' language loyalty is not as high as 100% and that conversations between people who know Basque are not always in Basque.

Given all these things, the indirect implication of the fact that real use is higher than “expected use” is that Basques in the Basque Country live in fairly compact groups and the relation network that connects them is quite vigorous and dense. Thus, since in their daily lives Basque speakers have fewer interactions with non-Basque-speakers, the use of Basque is higher than that statistically predicted. This also seems to say something about Basque speakers' language loyalty.



Conclusion 3: Linguistic diversity in our streets is increasing.

In the last five-year period, use of languages other than Spanish and French⁸ has increased significantly, from 2.6% to 3.7%. This point is only studied in the two most recent surveys; precious use of other languages was not registered.

⁸ Spanish is included in the “other language” group in the north, as is French in the south.

Conclusion 4: Four zone types can be differentiated in terms of sociolinguistic zones.⁹

The main findings in each zone are as follows:

Zone 1: Most of the population of Euskal Herria lives in the non-Basque-speaking zone in which fewer than 25% of people can speak Basque. Around 70% of the entire population live in such areas, where use of Basque is about 3% according to the latest survey. The same use level was recorded in 1993, but in 2006 it was 4%.

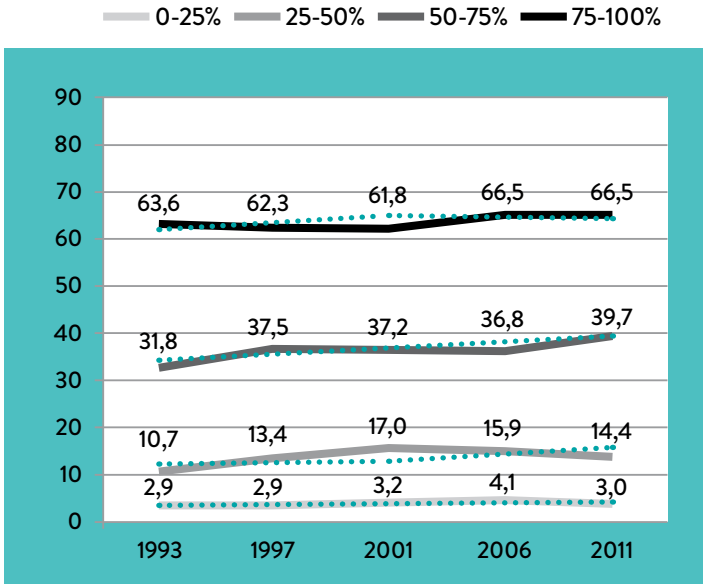
Zone 2: In the zone where between 25% and 50% of the population can speak Basque the level of Basque language use is around 14-15%. This has risen almost four percentage points over eighteen years.

Zone 3: In the areas where between 50% and 75% of the people can speak Basque, use of Basque is around 40%. There has been a very significant rise in use here since 1993, of almost eight percentage points.

Zone 4: Finally, in zones where most people (between 75% and 100%) can speak Basque, about 66% of the people observed spoke Basque. This has risen by three points since 1993.

⁹ Two notes here. We have data on zones by language knowledge since 1993, so we can analyse the trend over the past eighteen years. Unfortunately we have no language knowledge data for the northern Basque Country so we must limit our conclusions to the south here.

Figure 15. Use of Basque by language zone. % in Basque



Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 1989-2011

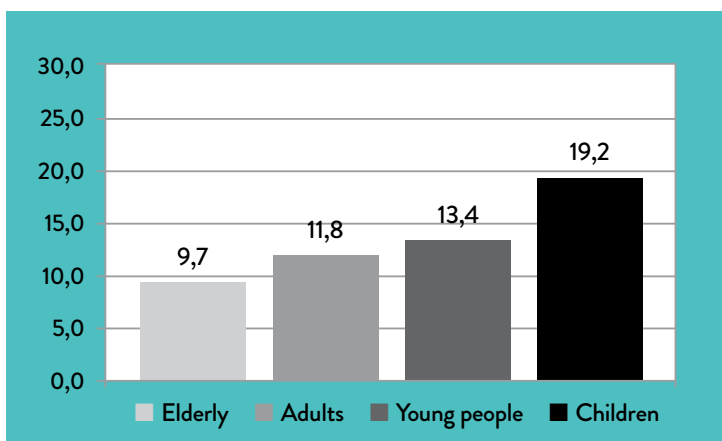
Conclusion 5: Younger people use more Basque

Four age groups were distinguished: children (2 to 14 years old), young people (15 to 24), adults (25 to 64), and elderly (65 and more).

Two interesting things are worth pointing out about the analysis by age groups:

If we order the age groups from those who speak most Basque to those who speak least, children are first (19.2%) followed by young people (13.4%), adults (11.7%) and, in last place, the elderly (9.7%). This suggests that the younger people are the more Basque they use. Although this pattern holds in most parts of the country, in Bizkaia there is a departure from it since elderly people use more Basque than the adult group.

Figure 16. Use of Basque by age group. Basque Country, 2011

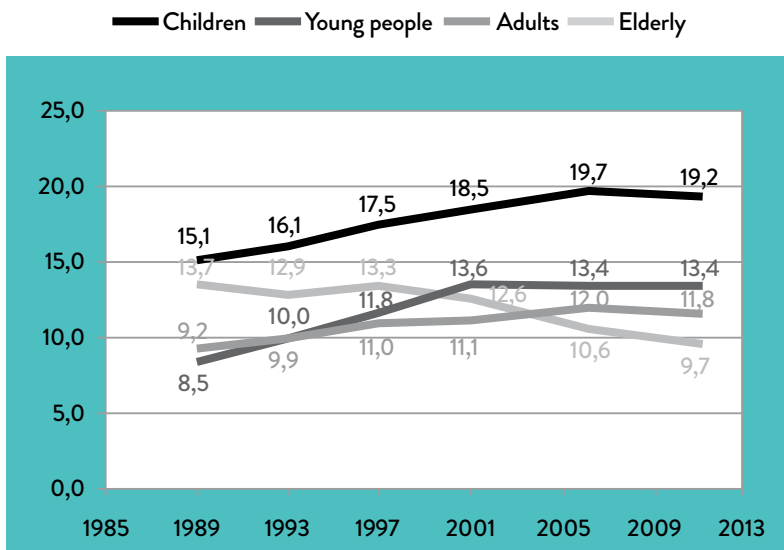


Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 2011

When the first street survey was performed in 1989, the highest level of use of Basque was registered in the childrens' group but the next group in order was that of the elderly, followed by adults, with young people in last place.

When we break down the trend over twenty-two years, we find that in general the use of Basque has risen in children, young people and adults but not in the elderly.

Figure 17. Trend by age groups, Basque Country, 1989-2011

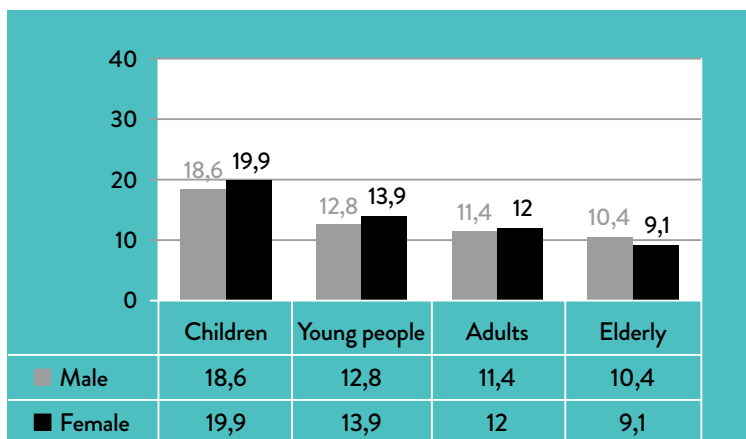


Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 1989-2011

Conclusion 6: Females speak more Basque than males in all age groups except the elderly.

Females speak Basque more than males in all the age groups, except in that of elderly people. This pattern is found in all parts of the country except the northern Basque Country and Navarre.

Figure 18. Use of Basque by sex and age group. Basque Country, 2011

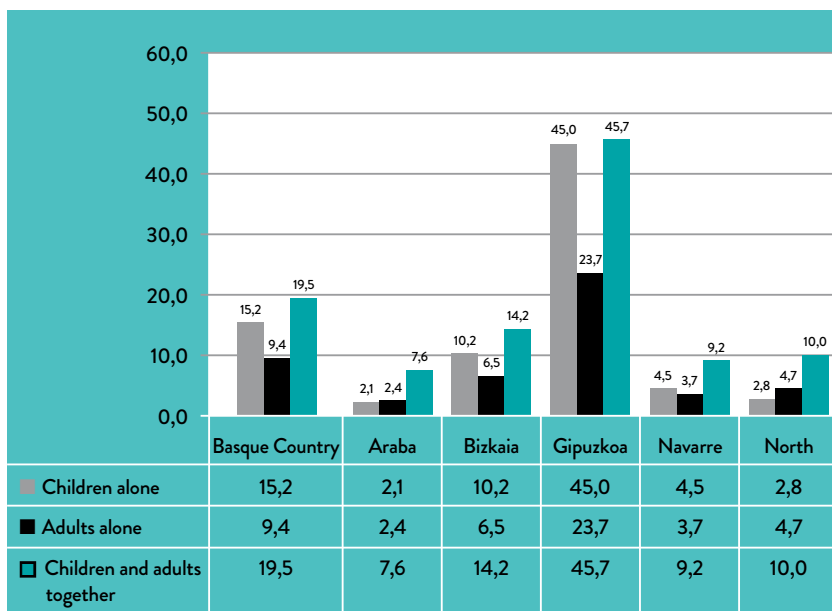


Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 2011

Conclusion 7: The presence of children is a very significant factor affecting use of Basque in the street.

The highest figures for use of Basque are recorded when children and grownups¹⁰ are together. Uses of Basque decreases considerably when children are talking in the absence of grownups or when grownups are talking in the absence of children. The exception is Gipuzkoa, which is where the most Basque is spoken.

Figure 19. Use of Basque depending on the presence of children. By province, 2011



Source: Soziolinguistika Klusterra, Street survey of language use in the Basque Country, 2011

10 “Grownups” encompasses young, adult and elderly people.

A guide to language
use observation

SURVEY METHODS





4.

OBSERVING LANGUAGE USE AT WORK AND SCHOOL

So far we have discussed street surveys. The street is an open space where you can run into anyone. In this section we will think about more restricted places, because interest has grown in recent years in the observation of language use in “closed” spaces.

Figure 20. Survey types by the kind of space



In our case, the most common are surveys in workplaces and in schools. Some language normalization assessment services in workplaces¹¹ have been monitoring and evaluating spoken language use in companies for years. They carry out annual studies to analyse the effect of intervention aiming to promote the language.

¹¹ Cf. the well-known programmes at companies such as Elhuyar, AEK-Ahize, Emun and Artetz.

The method described above is by and large applicable in closed spaces as well. Apart from the specifics, we can maintain the same methodological principles. It is just the specific details which need to be adapted or adjusted to the particular conditions. These will now be discussed.

Information about the subjects

One specific feature of closed domains is that we possess more information about the individual speaker. For instance, a surveyor who belongs to the organisation may already know where a subject comes from or how well a subject knows each language. In street surveys we do not know the speaker personally and the information recorded about the variables that are examined is obtained purely through observation. Thus the information about each individual that can be registered together with the language use data will be more complete and more reliable in surveys of restricted domains.

A big challenge: remaining unnoticed

One of the hardest things when carrying out fieldwork in closed or restricted spaces is keeping subjects from realising they are being observed. Sometimes the physical space is small and the same groups of individuals are found there every day. Some sort of measures is necessary in order to be able to do something unusual (data collection) unnoticed. One thing that can be done is to adapt the data collection mechanism to the place; in an office, for example, the data can be recorded on a computer without needing to move around, as if the surveyor were working on something. Another interesting place for carrying out a survey in workplaces is the room where the coffee machine is; here there is a more relaxed atmosphere and staff tend to converse a lot. In such spaces we may use the mobile phone app to record the data.

Another way to avoid being overly conspicuous during fieldwork is for the surveyor to be a member of the group. It is not a good idea for a person who is not usually there to be collecting data, because people will realise that something funny is going on. It is therefore important for the surveyor to be a member of the staff (or in the case of schools, a pupil, a teacher or a member of the school staff).

Another advantage of the surveyor being somebody who belongs there is that they already know the subjects. Given that it is not allowed, in this method, to ask the subject any questions, this permits us to analyse the effect of different factors on spoken language use by noting down, together with the language of conversations, other kinds of information such as language competence, or the language model of the school.



Different observation spaces

It is very important that the data recorded in closed places should be representative of all settings within the domain. For oral activities, the most significant places are chosen. Thus when choosing places we should think about which are the natural settings for spoken communication to take place. In certain cases, some settings may be more formal and others more informal.

If so, the survey of each workplace and school may have different observation points. Let us look at some examples.

At a place of work some particularly interesting places for surveying might be, for instance: offices, the coffee room, dining rooms or the entrance area.

In schools, good places might include playgrounds, corridors, dining rooms or entrances.

Each survey area will have its own language use profile, and if the people

found at each point are representative of all the speakers in the domain, the sum of all these settings will give the domain's overall language use.



Variables and the observation data sheet

Basically the data sheet used in closed domains is the same as that used in street surveys. A few small adaptations are made for the specific domain.

For example, on the data sheet for workplaces, there are two dependent variables, in addition to the language of a conversation: speaker's age and position in the company.

Figure 21. Example of a data collection sheet for observation of language use in workplaces

SURVEY OF LANGUAGE USE

SOZIOINGUISTIKHA KLUSTERRA
T: 943 592 556
klusterra@soziolinguistika.org

WORKPLACE: _____

LOCATION: _____

DATE: _____ TIME: _____

SURVEYOR: _____

| | Language | | | Director | Management | Office staff | Technician |
|----|--------------------------|--------------------------|--------------------------|----------|------------|--------------|------------|
| | A | B | OTHERS | | | | |
| 01 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 02 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 03 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 04 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 05 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 06 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 07 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 08 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 09 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 10 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 11 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 12 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 13 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 14 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 15 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 16 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 17 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 18 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 19 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |
| 20 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | | |

Under 40 years old
 Older 40 years old

NOTES: _____

On the above form we recorded the speaker's age and position in the organisation. We drew a line at age forty, distinguishing between speakers who are younger and older than this. For position in the company we classified jobs into four types: director, management, office staff and technician. There are obviously other possible ways to classify or specify positions in an organisation, and the data collection sheet may be adapted to the nature of the company or the purposes of the researcher.

Instead of or in addition to these variables, others may be incorporated, such as language competence, sex or place of residence. Once again, when we choose a variable for study we bear in mind that the surveyor must record data without asking the speaker any questions, so the surveyor must be able to find out about such variables with regard to speakers; thus, the surveyor will need to know where the employee is from, or that language or languages the employee is able to speak.



Sample size: how many conversations should be observed?

In 2.1.4 we looked at the mathematical model used as the basis for calculating sample sizes in order to ensure adequate levels of reliability and acceptable margins of error.

That model lets us calculate the size of an adequate sample in either open or closed domains.

Table 5 shows how many conversations need to be recorded in a workplace in order to achieve 95% reliability and a margin of error of ± 5 , depending on the number of members of staff and the proportion of Basque speakers.

Table 5. Number of conversations that need to be surveyed in a workplace

| EMPLOYEES | % BASQUE SPEAKERS | RELIABILITY | ERROR | CONVERSATIONS |
|-----------|-------------------|-------------|-------|---------------|
| 50 | 90 | 95 | ±5 | 223 |
| 50 | 50 | 95 | ±5 | 349 |
| 50 | 25 | 95 | ±5 | 267 |
| 100 | 90 | 95 | ±5 | 233 |
| 100 | 50 | 95 | ±5 | 375 |
| 100 | 25 | 95 | ±5 | 283 |
| 250 | 90 | 95 | ±5 | 236 |
| 250 | 50 | 95 | ±5 | 383 |
| 250 | 25 | 95 | ±5 | 288 |

It is also desirable to carry out more than one survey session in each setting, and if possible to take several days to do this, in order to be sure that all situations are covered properly.

REFERENCES



- Altuna, Olatz; Berrio-Otxoa, Kontxesi; Jausoro, Nekane; Martinez, Zesar. 2008 (unpublished). "Ikerketa soziolinguistikorako metodologiak eta teknikak". In *HIZNET, Hizkuntza Plangintza Ikastaroa*. 2008-2009. Asmoz foundation, Eusko Ikaskuntza.
- Alvarez Enparantza, J. L., Txillardegí, Isasi, X. 1994. *Soziolinguistika matematikoa*. Bilbao: Udako Euskal Unibertsitatea.
- Alvarez Enparantza, J. L., Txillardegí, 2001. *Hacia una Socio-Lingüística Matemática*. Donostia: Euskal Soziolinguistika Institutua Sortzen.
- Bourhis, Richard Y.; Sachdev, Itesh. 1984. "Vitality Perceptions and Language Attitudes: Some Canadian Data". *Journal of Language and Social Psychology*, vol. 3., nº 2: pp. 97-126.
- Eusko Jaurlaritz 2008. *IV. Inkesta Soziolinguistikoa*. Vitoria-Gasteiz.
- Gal, Susan; A. Woolard, Kathryn. 2001. *Constructing Languages and Publics: Authority and Representation*. In *Languages and Publics: The Making of Authority*. Manchester (UK) eta Northampton (Massachusetts): St. Jerome Publishing. pp. 1-12.
- H. Hill, Jane. 2008. *The Everyday Language of White Racism*. Malden (Massachusetts): Wiley-Blackwell.
- Juaristi, Patxi. 2003. *Gizarte ikerketarako teknikak*. Bilbao: UPV-EHU

- Martínez de Luna, Iñaki; Isasi, Xabier; Altuna, Olatz. 2006. Use of the Basque Language, Key to Language Normalization. In *The Case of Basque: Past, Present and Future*. Azurmendi, M. J.; Martínez de Luna, Iñaki. Andoain: Soziolinguistika Klusterra. pp. 67-88.
- Martínez de Luna, Iñaki; Isasi, Xabier; Altuna, Olatz. 2006. Use of the Basque Language, Key to Language Normalization. In *The Case of Basque: Past, Present and Future*. Azurmendi, M. J.; Martínez de Luna,
- Rodrigue, Landry; Bourhis, Richard Y. 1997. "Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study". *Journal of Language and Social Psychology*, vol. 16, n° 1: pp. 23-49.
- SIADeco. 1979. *Conflicto Lingüístico en Euskadi*. Bilbao: Euskaltzaindia.
- Soziolinguistika Klusterra 1991. *Euskararen Erabilpenaren neurketa* BAT Sociolinguistics journal, 3-4. Andoain.
- Soziolinguistika Klusterra 2002. *Euskararen Kale Erabilpena Euskal Herrian* BAT Sociolinguistics journal, 28. Andoain.
- Soziolinguistika Klusterra 2007. Kale Neurketaren V. Neurketa, 2006. Emaitzak, azterketak, gogoetak. BAT Sociolinguistics journal, 64. Andoain.
- Urla, Jacqueline. 1989. Reinventing Basque Society: Cultural Difference and the Quest for Modernity, 1918-1936. In *Essays in Basque Social Anthropology and History*. Douglass, William A. Basque Studies Program Occasional Papers Series, n° 4. Reno: University of Nevada Press. pp. 149-176.
- Urla, Jacqueline. 1993. "Cultural Politics in an Age of Statistics: Numbers, Nations and the Making of Basque Identity". *American Ethnologist*, vol. 20, n° 4: pp. 818-843.
- Urla, Jacqueline. 2012. *Reclaiming Basque: Language, Nation and Cultural Activism*. Reno (Nevada): University of Nevada Press.
- Woolard, Kathryn Ann. 1989. *Doubletalk: Bilingualism and the Politics of Ethnicity in Catalonia*. Stanford: Stanford University Press.
- Woolard, Kathryn Ann. 2008. "Language and Identity Choice in Ca-

talonia: the Interplay of Contrasting Ideologies of Linguistic Authority”. In Lengua, nación e identidad. La regulación del plurilingüismo en España y América Latina, pp. 303-324. Madrid-Frankfurt am Main: Iberoamericana-Vervuert Verlag.

Yurramendi, Y. & Altuna, O. (2009) Zuzeneko behaketaz hizkuntza-erabilera neurtzeko metodologiaren eredu matematikoa. Laginketa eta estimazioa. Soziolinguistika Klusterra. <http://www.soziolinguistika.org/files/TXOSTENA.pdf>

Zarraga, Arkaitz; Coyos, Jean-Baptiste; Hernandez, Jone M.; Joly, Lionel; Larrea, Imanol; Martinez, Loren V.; Uranga, Belen; Bilbao, Paul. 2010. Soziolinguistika Eskuliburua. Eusko Jaurlaritzak. Soziolinguistika Klusterra