

# Linguistic Variation in the Basque language and Education-II

## Euskararen bariazioa eta bariazioaren irakaskuntza-II



Editoreak:  
Aitor Iglesias  
Asier Romero  
Ariane Ensunza



*CIP. Biblioteca Universitaria*

**Linguistic** variation in the basque language and education - II [Recurso electrónico] =  
Euskararen bariazioa eta bariazioaren irakaskuntza -II /editoreak, Aitor Iglesias, Asier Romero,  
Ariane Ensunza. – Datos. - Bilbao : Universidad del País Vasco / Euskal Herriko  
Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, [2016]

1 recurso en línea : PDF (140 p.)

Modo de acceso: World Wide Web.

Textos en inglés y euskara.

ISBN: 978-84-9082-465-8.

1. Euskara (Lengua) - Dialectos. 2. Lenguaje y lenguas - Variación. 3. Prosodia (Lingüística).  
I. Iglesias, Aitor, coed. II. Romero, Asier, coed. III. Ensunza, Ariane, coed. IV. Título:  
Euskararen bariazioa eta bariazioaren irakaskuntza -II.

809.169-087(0.034)

© Euskal Herriko Unibertsitateko Argitalpen Zerbitzua  
Servicio Editorial de la Universidad del País Vasco  
ISBN: 978-84-9082-465-8 (pdf)



## AURKIBIDEA

Iglesias, A.; Romero, A.; Ensunza, A.:	
Sarrera.....	3-4
Elvira-García, W.:	
Transcribing intonation with ETI_TOBI.....	4-15
Aurrekoetxea, G.; Iglesias, A.; Santander, G.; Usobiaga, I.:	
Diatech tresna informatikoaren bertsio berria.....	16-24
Videgain, X.; Aurrekoetxea, G.:	
Euskararen bariazio geo-morfologiaren azterketa.....	25-36
Sainz-Maza, L.; Rodríguez, I.:	
Subject pronoun expression in Basque: description and pedagogical implications.....	37-47
Etxebarria, A.; Romero, A.; Gaminde, I.; Garay, U.:	
Bariazio prosodikoaren eragina testu irakurrien ulermenean.....	48-55
Roseano, P.:	
Dos décadas de dialectometría entonativa.....	56-80
Asensio, N.; Barrios, H.; Lázaro, A.; Sáez, I.:	
Euskaldun berrien eta zaharren ulermena jarrerren inguruan.....	81-86
Ensunza, A.:	
Hizkuntza-aldaketaren inguruko jarrera eta prestigioaz.....	87-94
Romero, A.; Etxebarria, A.; De Pablo, I.; Sanz, A.:	
Keinuen eta bokalizazioen arteko harremanak komunikazio-funtzio goiztiarretan: ELAN softwarea erabiliz.....	95-103
Etxebarria, A.; Gaminde, I.; Olalde, A.; Gaminde, U.:	
Hizkuntza aldakortasuna Larrabetzuko aditz morfologian.....	104-119
Gaminde, I.; Romero, A.; Etxebarria, A.; Eguskiza, N.:	
Bizkaiko aditz laguntzaileen bilakaeraren azterketaz.....	120-141

## SARRERA

### Aitor Iglesias, Asier Romero eta Ariane Ensunza EUDIA taldea (UPV/EHU)

Ikerketaren helburuen artean errealitatea hobetzeko ezagutza handitzea dago. Ikerketaren bidez lortzen den ezagutza jakitera emateko, sozializatzeko eta partekatzeke, eztabaida guneak sortzea da bideetako bat, agian eraginkorrenetakoa, baina ez bakarra. Ikerketaren ekoizpena argitaratzea beti ez da nahikoa izaten. Emaitzak zein metodologia eztabaidatzeko beharra ere izaten da maiz. Eta ikertzailea prest azaldu behar da emaitzak zein metodologia lankide zein urrunagoko ikertzaileekin eztabaidatzeko, jakinik ez dela bide bakar bat arlo honetan ere, kritikak entzun eta aztertzeke, berrikusketak eta eguneratzeak egiteke, teoria etengabeko zalantza zientifikoaren gunean kokatuz, zeren ongi asko dakigu teoria artifizio bat baino ez dela, entelekia bat, errealitatea birsortzeke eta interpretatzeko. Hori dela eta ikertzailea jarrera dogmatikoetatik urrundu beharrean dago, nahiz berea dela egokiena uste osoa izan; ondokoena entzuten asmatu behar du, proposamen berrien iradokizunak interpretatzen ere ikasi behar du, berea umotuz joan dadin.

Ideiak, teoriak, metodologiak eztabaidatzea beharrezko ez eze gustuko ere badugu, zorionez. Lema hori gogoan, ekin zion EUDIA ikerketa-taldeak duela urte batzuk urtero molde eta tamaina desberdinetako jardunaldiak eratzeari. Beti helburu bakarra gogoan: euskararen bariazioari buruzko ezagutza handitzea, era bateko zein besteko ekarpenak egiteko guneak sortuz. Ekarpene guztiak dira baliagarriak eta denetatik egoten da zer ikasi. Ekarpene eta bere ondoren sortzen den eztabaida dira ezagutza handitzeko edo eguneratzeko tresna baliagarrienak. EUDIA *workshop* edo jardunaldi sorta horren adibide argia da.

Bide horretan ikerketa-talde honek jada bost jardunaldi edo *workshop* eratu ditu Gasteizen, Bilbon eta Leioan, 2008az geroztik (2009an, 2011n, 2012an, 2015 eta 2016ko hau). Horietan zehar nagusiki euskararen bariazioaz eztabaidatu da, nahiz izan diren inguruko beste gai batzuetako ekarpenak ere. Ikertzailea beti agertu behar da prest bere ikusmoldeak jendearen aurrean plazaratu eta gainerakoekin eztabaidatzeko. Jarrera irmo, egozentriko, orojakiletik urrundu behar du, izurritik bezalaxe. Ikusmolde dogmatikoak ez datoz bat zientziarekin. Ikertzailea dogmatismotik urrundu behar da, ez da inolaz ere bere irizpidea izan behar. Ez naiz ziurrenik lehena ideia honen gainean idazten duena (nahiz ez ditudan eskura gai honetaz idatzi diren lanak), eta ziurrenik ere, eta zoritxarrez, ez naiz azkena izango.

Apaltasuna izan behar da ikertzailearen ezaugarri indartsuenetako bat. Apaltasuna bere ideiak eta bere teoriak azaltzen, apaltasuna bere teoriak ez direla iraungitze datarik gabeak onartzen. Teoriak sortzen diren bezalaxe iraungi daitezkeela, beste teoria hobeak, osatuagoak sortzen direnean edo delako teoria horren balioa ezereztu dela onartzean.

Teoriak frogatu behar dira eta frogagarriak behar dira, alegia, hipotesi bat ez dela bihurtzen teoria, ezelango frogarik gabe.

Euskararen bariaziora itzuliz, jardunaldi hauek gai honen inguruan ikertzen dihardugunon eguneratze ahalegin bat gehiago dira, beste gune bat non gure lorpenak

eta ideiak eztabaidagai jartzen ditugun eta komunitate zientifikoaren eskuetan uzten ditugun. 2016ko jardunaldian bi ardatzetan bildu dira gure ekarpenak: euskararen bariazioa eta bariazioa eskolan edo bariazioaren irakaskuntza. Lehen ardatzean zortzi ekarpen egin dira, bertako eta nazioarteko ikertzaileen eskutik (Gaminde, Etxebarria, Iglesias, Romero, Aurrekoetxea, Videgain eta beste hainbat ikertzailearen ekarpenak izan dira eztabaidagai), nazioartekoaren artean Sainz-Maza, Rodríguez, Roseano, Elvira-García... izan ditugu euren ekarria aurkezten. Guztira hamaika txosten aurkeztu dira egun bateko baina bete-beteko jardunaldian. Eskerrak hemendik parte hartzaile guztiei.

Ekarpenak orotara bost multzoan sailka genitzake:

- a) Hizkuntza bariazioaren inguruan ikusmolde desberdinak erabiliz, hala nola teoria desberdinen ikusmoldeak (Elvira-García “Transcribing intonation with ETI\_TOBI”, Videgain eta Aurrekoetxearen “Euskararen bariazio geomorfologiaren azterketa”, tresneriaren aldetik (Aurrekoetxea, Iglesias, Santander eta Usobiagaren “Diatech tresna dialektrometrikokoaren bertsio berria”), euskalkien konbertzentzia (Aurrekoetxea, Fernandez, Ormaetxea eta Rubioren “Euskalkien konbergentzia”), bariazio soziolinguistikoa (Ensuzaren “Hizkuntza-aldaketaren inplikazio soziolinguistikoak euskarari: jarrerak eta prestigioa” eta azkenik bariazio morfologikoaren gaineko lanak: Etxebarria, Gaminde, Olalde eta Gaminderen “Hizkuntza aldakortasuna Larrabetzuko aditz morfologikoan” eta Gaminde, Romero, Etxebarria eta Eguskizaren “Bizkaiko aditz laguntzaileen bilakaeraren azterketaz”).
- b) Intonazioaren bariazioa bereziki landu izan da ekarri zenbaitetan, hala nola Etxebarria, Romero, Gaminde eta Garayren “Bariazio prosodikoaren eragina testu irakurrien ulermenean” eta Roseanoren “¿Cuánto nos parecemos? Métodos cuantitativos para el estudio de la variación entonativa en áreas románicas”.
- c) Code-switching edo kode-aldaketa landu dute Asensio, Barrios, Lázaro eta Sáezek bere “Euskaldun berrien eta zaharren ulermena jarreraren inguruan” ekarriarekin.
- d) Eta azkenik, prosodia eta haurren keinua ere landu da ekarpen batekin: Romero, Etxebarria, De Pablo eta Saenzen “Keinuen eta bokalizazioen arteko harremanak komunikazio-funtzio goiztiarretan: ELAN softwarea erabiliz”.

# TRANSCRIBING INTONATION WITH ETI\_TOBI

**Wendy Elvira-García**  
**Universitat de Barcelona**  
[wendyelvira@ub.edu](mailto:wendyelvira@ub.edu)

## Abstract

This paper provides guidance and instructions to use an automatic intonation transcriber for Catalan and Spanish, Eti\_ToBI. Transcribing intonation is, in most cases, the first step in its study and nowadays ToBI systems are probably the most popular transcribing conventions. However, the transcription of intonation is a time-consuming task and this fact has triggered the interest in achieving automatic transcribers. Eti\_ToBI is the first ToBI transcriber for Catalan and the first in Spanish thought for doing research with Praat. Eti\_ToBI is a rule-based transcriber based only on linguistic knowledge that performs as good as predictive systems. This paper aims at showing the program requirements and functioning.

**Keywords:** Speech recognition, automatic transcription of intonation, Sp\_ToBI, Cat\_ToBI

## 1. Introduction

Transcribing intonation is a time consuming task that every phonetician must face before analyzing his/her results. There is evidence that this task can be automatized and yet most phonologists keep performing intonational transcriptions manually. The main reason for it is that, in the path towards speech recognition, rule-based systems and, thus, linguists are left behind more and more since the systems engineers design are meant to be used by the general user as a final product and not designed to perform research. Thus, those systems cannot help phonologists and phoneticians in their daily work.

However, in the last decade, the arrival of Praat (Boersma & Weenink, 2015) has opened a new horizon for phoneticians. Praat gives to researchers the opportunity to design their own tools by using its programming language. Praat scripting language functions are based on Praat GUI commands, what makes the language very intuitive to phoneticians.

A proof of its success is that almost any Phonetics Laboratory develops its own scripts, which allow performing phonetic analysis quicker. Scripts have become so popular that researchers are considering new methods for hosting scripts in a centralized way (CPrAN Team, 2016).

In this context, recently a new Praat script has been created that performs an intonational transcription for Spanish and Catalan, Eti\_ToBI (Elvira-García, Roseano, Fernández-Planas, & Martínez-Celdrán, 2015). The script labels intonational movements according to Sp\_ToBI (Estebas-Vilaplana & Prieto, 2008; Hualde & Prieto, 2015) and Cat\_ToBI (Prieto, Aguilar, Mascaró, Torres Tamarit, & Vanrell, 2009; Prieto & Cabré, 2013) notation systems. In this paper, the instructions for using Eti\_ToBI as a transcriber for Spanish and Catalan prosody are detailed.

The paper is organized as follows: following section deals with general aspects of the program: subsection 0 focus on its motivation and 0 on its novelty. Section 3 deals



with the actual use of Eti\_ToBI explaining how to prepare the materials and how to use the program. Section 4 summarizes the reported assessment of the tool and, finally, section 5 presents the conclusions.

## 2. General aspects of Eti\_ToBI

This section presents Eti\_ToBI giving special emphasis to the motivation and novelty of the program.

### 2.1. Motivation of the program

In the general framework of speech recognition, recognition of prosody and specially intonation is used to segment the speech chain in words or sentences and to disambiguate words. Those systems are thought as a final product and they do not provide the transcription used to get the final result. In that framework phoneticians and phonologists have been forced to create their own tools in order to achieve a transcription of intonation.

Transcribing intonation has become a matter of a great concert among phoneticians resulting in a number of publications dedicated to the subject (Ananthakrishnan & Narayanan, 2008; Avesani, 1995; Campbell, 1996; Lee, Kim, & Lee, 2002; Mertens, 2004; Zhi, Hirst, & Bertinetto, 2010). Out of these, as in general speech recognition systems, the vast majority use prediction systems that rely in statistics and only few use linguistic knowledge.

An issue in transcribing intonation is that there are many different theoretical frameworks that can be used (such as IPO, Aix-en-Provence, analysis by configurations, by levels or the AM model) and each system requires a different tool. However, AM (Pierrehumbert, 1980) and its ToBI notation conventions (Beckman & Elam, 1997) are nowadays the most popular systems for transcribing intonation as it proves the growing number of languages that have their own ToBI (Jun, 2005, 2014). Despite solving the problem of having multiple systems, using ToBI-based conventions has a capital drawback when trying to transcribe intonation automatically. The phonological model is high language-dependent meaning that a different tool is needed for each language (and, arguably, for each dialect).

Such tool does not exist for Catalan and exist with limitations for Spanish (David Escudero, Aguilar-Cuevas, González-Ferreras, Gutiérrez-González, & Cardeñoso-Payo, 2014). The previous Sp\_ToBI transcriber uses fuzzy logic in order to make its predictions, which show up on the screen as a probability in percentage that the researcher needs to confirm. Besides, the software is not integrated in any know tool but the researcher needs to learn to use a new one.

Therefore, the main motivations for creating a new transcription tool were: on one hand, that there were not intonational transcribers for Catalan and, on the other hand, that the existing tools are not specifically thought for helping phoneticians in their daily research.

### 2.2. Novelty

In addition of being the first transcription tool for Catalan, the tool presented in (Elvira-García et al., 2015) presents to main innovations. 1) the method for transcribing the intonation is rule-based showing that we have acquired the sufficient linguistic knowledge in order to transcribe intonation relying in those rules. 2) It is a user-friendly (in this case phonetician-friendly) tool that explodes previous phonetician knowledge by using Praat.

In the world of machine learning and big data, the role of the linguist in the design of speech recognition is usually underestimated relegating him/her to the tedious task of transcribing corpus to use as a training set or assess the machine results. In this context, advances in language technologies are due to the increase of the training corpus available and research done by linguists is not being implemented in computational tools. Therefore, when new tools report that the speech recognition rate improves, the improvement is not due to new discoveries in linguistics not even that there has been any improvement in the statistical techniques, but the improvement is due to the fact that on-line linguistic data is growing every day.

Linguistic research is usually performed in a theoretical way, without taking into account how the results obtained could be used in speech technologies. In fact, linguistic-based systems had not been implemented into prosodic recognition systems since the pioneer works from the 80' (Lea, 1980). Yet, advances in linguistics since that decade are notorious. Furthermore, intonation is a field that has grown especially in the last 20 years due precisely to technical improvement in signal theories and the generalized adoption of digital recordings.

In this context, the usage of the systematic knowledge obtained in the discipline is a priority for Eti\_ToBI. Applying linguistic knowledge, on one hand, proves linguistic theories can explain language phenomena and, on the other hand, gives more information about how language works and where the theory assumptions fail. Using Eti\_ToBI not only can speed up phonetician works in order to let them focus more in theory, but it can also highlight fails in the theoretical framework they use.

Existing software to transcribe or to perform tasks like syllabification are written in several programming languages and have very different interfaces each one of them showing different levels of integration with other tools used by phoneticians. This means that a regular phonetician that uses these tools should have installed in his/her computer a number of interpreters (e.g. Java, Perl, Python, Matlab...), in order to get these independent programs to work. However, the worldwide standard software for acoustic analysis for phoneticians is Praat. And Praat has its own scripting language that can be exploded to create tools that have total integration with regular acoustic analysis. Therefore, the most convenient programming language for a phonetician is Praat scripting language.

Eti\_ToBI has been designed to make phoneticians research easier and faster. That is why the tool has been created as a Praat script. This way, it does not force the researcher to install other software or programming language in his/her computer, which is usually a time-consuming and not transparent task for researchers that are not very familiar with computer sciences.

Furthermore, since it is written in Praat scripting language, *Eti\_ToBI* shows complete integration with Praat TextGrids making possible further analysis with the transcription and using other scripts with the data. Moreover, using *Eti\_ToBI* will not require any specific formation besides knowing how Praat scripts work.

### 3. Using *Eti\_ToBI*

This section covers *Eti\_ToBI* usage. It guides the user preparing materials and using the tool interface. As has been previously said *Eti\_ToBI* is a Praat script that writes the *Sp\_ToBI* or *Cat\_ToBI* label for each TBU (tone-bearing unit, i.e. stressed syllables and boundaries) that finds in a sentence previously segmented in a TextGrid. In Figure 1 the general workflow of *Eti\_ToBI* can be observed.

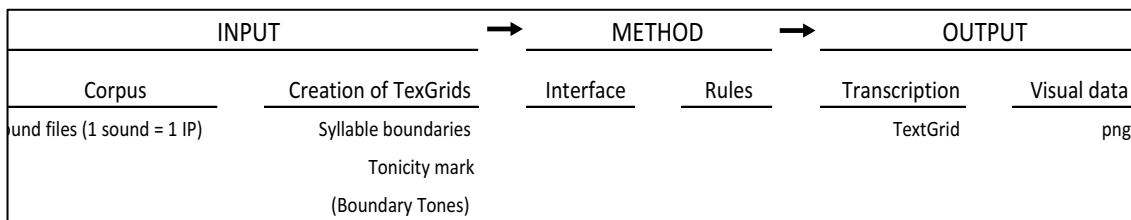


Figure 1. General schema of *Eti\_ToBI* workflow

This general schema can be divided into a) actions that require user interaction: preparing the input and using the interface, and b) those actions that do not require user interaction: applying the rules and creating the output.

In following subsections, the schema is explained according to the previous classification. Section 0 deals with user interaction paying special attention to the TextGrids creation. Section 0 explains the general philosophy behind implementation rules in *Eti\_ToBI*. Section 0 shows the outputs of the program.

#### 3.1. User interaction

User interaction in *Eti\_ToBI* is reduced to preparing the materials (the input) (section 0) and selecting options in the interface (section 0).

##### 3.1.1. Preparing materials

*Eti\_ToBI* cannot work with raw sounds, in order to perform its analysis it needs that data must be annotated previously. Specifically, it needs two pieces of information, syllable segmentation and information about syllable tonicity<sup>1</sup>.

This means that, in order to run, the script needs a folder with a set of .wavs and their TextGrids (with the same name) attached. The TextGrid must have at least an interval tier with the syllable segmentation of the sentence (it does not need the transcription) and a character marking where the stressed syllables are. The tonicity mark can be any character or symbol but is must be only one character. If Praat

<sup>1</sup>This operations can be also performed with other tools like AuToBI or SPASS (Bigi, 2015; Rosenberg, 2010).

trigraphs have been used (see Praat manual) in order to transcribe the sentences, the tier would be converted to Unicode before the analysis begins.

In addition, the user can mark the break indices (BI) in the sentences (that is not compulsory, though). If the BIs are marked, the tool will use the level 3 breaks in order to assign a intermediate phrase boundary tone (L- or H-) to the break.

Once the .wav and the annotated TextGrid are ready, you can launch the script.

### 3.1.2. The interface: Choosing your ToBI

Once the materials are prepared, *Eti\_ToBI* works as any other script. Open it with Praat and click "Run" and "Run" again. Once you run the script, a first form will appear (see Figure 2). You can set several parameters in this form. The form asks for the folder where the .wavs and TextGrids are kept. The second field asks for the symbol or character that has been used in the corpus as a mark of tonicity. If you have used the AFI mark of tonicity (primary stress mark) written in Praat trigraphs (i.e. \1 instead of ' ), you can write the trigraph in the field and the tier will be converted into Unicode characters in order to perform the analysis.

The third field states whether you need a new tier in order to write the transcription or you have already prepared a blank tier in your TextGrid. If a new tier is needed, fill the following field stating the position where the tier will be created. The same applies to the next box: click it if you have written the Break Indices in the the TextGrid. If the BIs are marked in the TextGrid, specify in which tier number they are.

The next two boxes specify the threshold in order to a) consider a pitch movement and b) consider a pitch movement extra high. By default, those values are set to 1.5 semitones and 6 semitones. The methodological reasons for that can be consulted in Elvira-García et al. (in press).

Next fields are used to set the type of transcription. If more than one is active, each transcription is written in a different tier. The types of transcriptions are:

- 1) Surface transcription: a narrow phonetic transcription that would give as a result tritonal pitch accents not included in current ToBI systems. This transcription would label a rise-falling movement as L+H\*+L or a rising movement in the stressed syllable that continues in the poststressed syllable as L+H\*+H.
- 2) Deep transcription: a broad phonetic transcription. In this type of transcription only bitonal pitch accents are included. The previous labels will be converted to L+H\* and L+<H\*.
- 3) Conventionalized transcription: This tier is used to translate into the current ToBI systems some nuclear configuration that would result of the phonological analysis (e.g. L\*+H L% will be converted in L\* HL%). You can see an exemplar sentence with all transcriptions in Figure 6.

The script also gives the opportunity to correct the sentences while the script is performing the transcription by clicking correction. This can be more useful than letting the script perform all the analysis and later on correcting it by opening the TextGrid

again, because when the script is running it prints differences in semitones between syllables and targets and these data can help to decide which label is more appropriate.

The script also gives the chance of creating a figure with the transcription that will be saved in png at 300dpi.

Lastly, the script permits beginning the transcription in a determinate sound of the way list in the folder. This is useful when the researcher begins the transcription and does not finish in one session. The following session can begin, for example, from file 50.

Figure 2. Form of Eti\_ToBI.

In the second form, some of the conventional labels that will appear in the transcription can be customized. For instance, you can decide whether the label L\*LH% should be included in the transcription, or replaced by L\*H%.

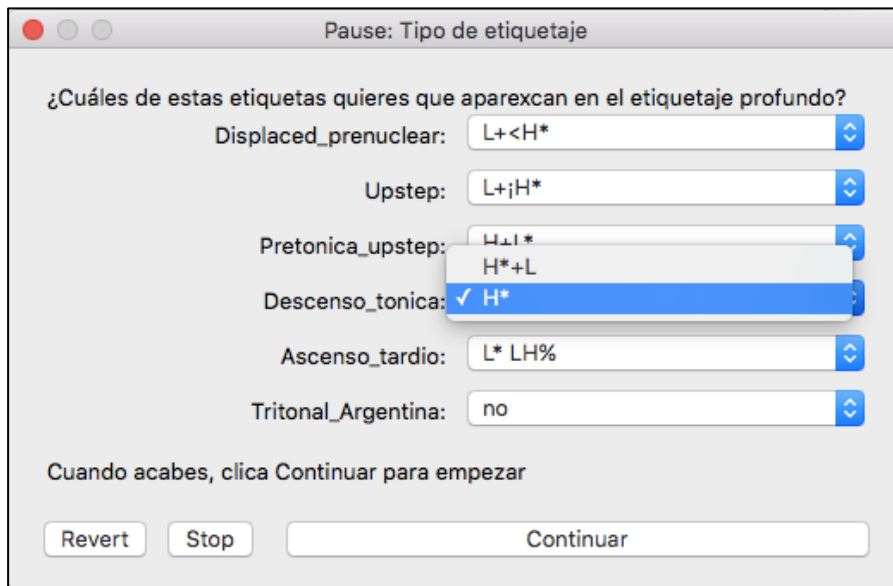


Figure 3. Window showing different options to obtain a customized transcription.

After filling the forms, the automatic analysis will begin. The time lapse until it finishes will depend on the number of sounds that the program has to analyse and mostly whether the figures are being created. The average time for transcribing 100 sentences and creating its pictures in a computer with a 2,4 GHz processor is 10 minutes.

### 3.2. The rules behind *Eti\_ToBI*

As has been previously said, *Eti\_ToBI* is a rule-based approach. This means that after extracting the pitch contour, it applies a series of if-then statements in order to get the transcription. It uses exactly 60 sets of rules and 20 subrules. A more detailed explanation of the rules and the labelling possibilities can be found in Elvira-García et al. (in press).

All rules in the script follow the same schema, they label pitch movements greater than 1,5 semitones (Pamies, Fernández Planas, Martínez Celdrán, Ortega-Escandell, & Amorós Cespedes, 2002). Semitones differences are computing following the formula  $(12/\log_{10}(2)) * \log_{10}(f_{02}/f_{01})$ . However, the set of rules that apply are different depending on the position of the accent that has to be labelled (i.e. pre-nuclear pitch accent, nuclear pitch accent or boundary tone) and its lexical type (oxytone or other). Following subsections explain how the pitch classifications are performed.

#### 3.2.1. Basics of pitch accent classification

According to AM model, pitch accents are anchored to prominent positions (i.e. stressed syllables and phrase boundaries) and that is the main reason why the script needs a tonicity mark in order to work properly.

This means that, in order to find a pitch movement, the script has to look for changes in pitch in a 3-syllable window, what is to say that it has to look for differences between prestressed, stressed and poststressed syllable. And this is the first step it takes. If there is not any change in pitch level between prestressed and stressed syllable and

stressed and poststressed syllable, the label is L\*. If there is a rise from the prestressed to the stressed syllable, it labels L+H\*. If there is a rise between the stressed and the poststressed syllable, it labels L\*+H, etc. But this is not enough in order to transcribe all pitch movements, so the difference between the start and end points of the stressed syllable and the nearest peak and valley are also computed.

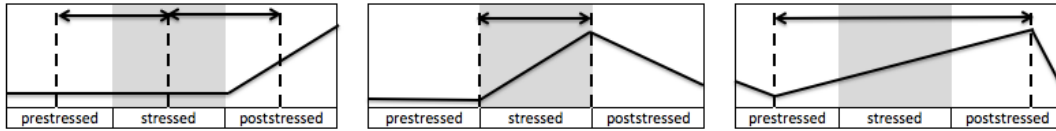


Figure 4. Time points where the script looks for differences (extracted from Elvira-García et al., in press)

Prenuclear and nuclear pitch accents have different rules since the Sp\_ToBI and Cat\_ToBI inventory is different in prenuclear and nuclear positions. In nuclear position, the script gets pitch values in 3 points within the prestressed syllable, 3 points within the stressed syllable and 3 points within poststressed syllable in order to detect complex boundary tones. However, despite differences in label inventory, the *if-then* philosophy behind the rules is the same.

### 3.2.2. Basics of boundary tone classification

The dynamic of boundary tones classification is very similar to the one described for pitch accents. Once the nuclear pitch accent has been detected, the tool assigns a tone level to the last stressed syllable, either low or high. Then, the program computes differences in semitones detected in the boundary tone. The script splits the time range of the poststressed syllable in 6 parts where it gets pitch differences. In addition, it also gets pitch values of the maximum and minimum pitch in the poststressed time range and the maximum pitch value in the last stressed syllable.

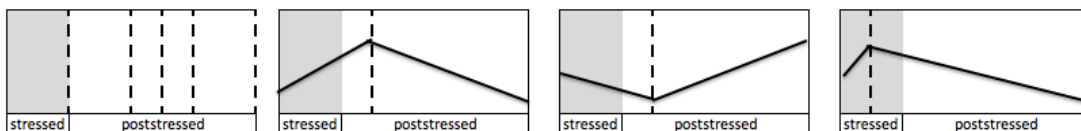


Figure 5. Time points where the script gets pitch values (extracted from Elvira-García, 2015)

The philosophy of the rules using this pitch points will result in an *if-then* condition as the following: If the last stressed syllable is low and within the poststressed syllable there is a rise greater than 1.5 semitones, set the label of the boundary tone to H%.

### 3.3. Outputs

Eti\_ToBI has two types of outputs: written and visual.

The written output consists in the TextGrid of the sound with new tiers for each type of transcription selected in the interface. Since Eti\_ToBI works for every sound in a folder, once the script goes through all the files, every TextGrid used in the analysis will be modified. Having the transcription in TextGrids makes possible to use the new transcription with a new script either to extract its labels or extract any other data that is transcription dependent (e.g. extract pitch excursion where the label equals L+H\*).

Eti\_ToBI overwrites the existing TextGrid meaning that it adds new tiers to your TextGrid. Therefore, if you run the script twice you will find a TextGrid with a double transcription and up to eight tiers. That is why is advisable saving a copy of the prepared TextGrids (with the syllable segmentation and BI) in other folder. If you have not copied your TextGrids before running the script, in the event of getting an unwanted transcription, you will be able to remove unwanted tiers using other scripts (e.g. remove tiers v.1 (Elvira-García, 2015)).

The visual output consists of a picture in the format selected in the interface containing the waveform, spectrogram, pitch contour and TextGrid related to the sound and, thus, its ToBI transcription Figure 6.

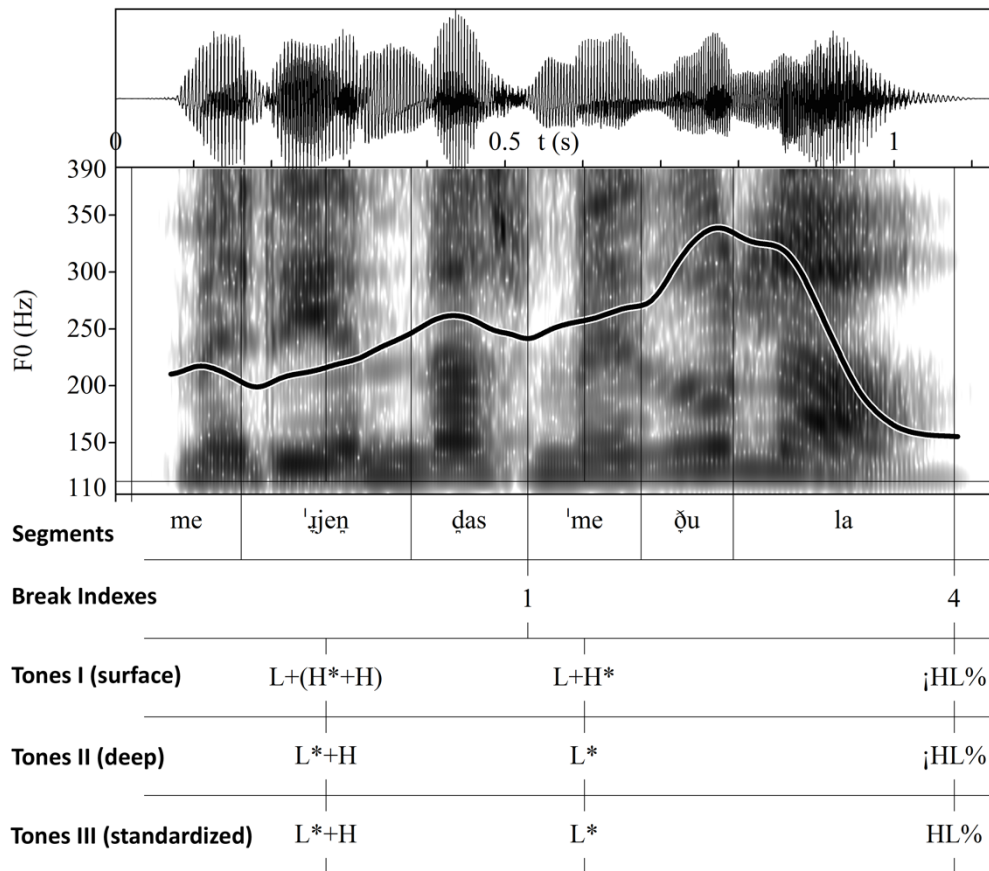


Figure 6. Visual output of Eti\_ToBI showing all three types of transcription and labelled for didactic purposes, extracted from Elvira-García et al, in press). The sentence corresponds to a y/n question produced by a female speaker from Cantabria.

By default, the spectrogram shows frequencies between 0Hz and 5000Hz and the dynamic range is set to 50dB. The pitch contour range is calculated automatically using De Looze (2010) and Hirst's (2011) assumptions about range in non-emotional speech.



#### 4. Assessment

The results in this section have been previously reported in Elvira-García et al., in press. The tool was evaluated by comparing its transcription with a transcription given by human labelers. Different experiments were conducted for Catalan and Spanish but in both cases the metric used in order to get accuracy results were kappa statistics (Cohen, 1960; Fleiss, 1971). Kappa's values vary from -1 (total disagreement) to 1 (total agreement) being 0 chance level. In general, kappa values are considered as moderate agreement (.60-.70), good agreement (.70-.80) and very good agreement (>.80).

The materials used in Catalan consisted of 100 sentences pronounced by 20 speakers of central Catalan and 9 nuclear configurations. Four different human transcribers labeled the corpus. For Spanish, a unique human transcriber labelled the corpus, which consisted of 1186 sentences produced by four speakers from Madrid, Barcelona, Cantabria and Seville.

Catalan				Spanish			
Type	% agreement	Kappa	Assesment	Type	% agreement	Kappa	Assesment
PPA	94.94%	0.907	very good	-	-	-	-
NPA	88.11%	0.831	very good	NPA	83%	0.728	good
BT	81.28%	0.756	good	BT	92%	0.879	very good

Table 1. Summary of the reported assessment in (Elvira-García et al. 2015)

Elvira-García et al. (in press) report that:

“The mean level of agreement (in percentage) is similar to the one performed for multiclass pitch accent and boundary tone labelling by human transcribers between them”(D. Escudero, Aguilar, Vanrell, & Prieto, 2012; Jun, Lee, Kim, & Lee, 2010; Pitrelli, Beckman, & Hirschberg, 1994; Syrdal & McGory, 2000).

Therefore, EtI\_ToBI has proved to be a useful and accurate tool and its results as good as results obtained by human transcribers.

It has to be bared in mind that both corpora were recorded in laboratory conditions. Therefore, the performance of the tool can be worse when used with other data. The voice recordings quality in laboratory speech is the best possible given that there is no noise in the signal and, therefore, Praat pitch detection algorithm can perform adequately. Moreover, both corpora were designed with mostly voiced sounds, which creates a continuous pitch contour.

#### 5. Conclusions

EtI\_ToBI is a Praat script with a user-friendly interface that requires little effort for a trained phonetician. In order to use it, the researcher only needs having Praat installed and a semi-annotated corpus. The corpus must have a TextGrid attached for each sound and in the TextGrid, a syllable segmentation and a mark for each stressed syllable in the sentence.

The tool includes an interface that provides flexibility to perform different types of transcription. Moreover, it permits customizing the transcription in order to meet

researcher's criteria. Eti\_ToBI has proved being a reliable tool that equals previous inter-rater agreement using several ToBI notation systems.

### Acknowledgement

The European Regional Development Fund (ERDF) and the Spanish Government supported this work through the project FFI2015-64859-P, namely, AMPER-based technologies. Paolo Roseano, Ana Ma. Fernández Planas and Eugenio Martínez Celdrán were involved in Eti\_ToBI rules design and were transcribers in the Catalan inter-reliability task.

### 6. References

- Ananthakrishnan, S., & Narayanan, S. S. (2008). Automatic prosodic event detection using acoustic, lexical, and syntactic evidence. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1), 216–228. <http://doi.org/10.1109/TASL.2007.907570>
- Avesani, C. (1995). ToBI: un sistema di trascrizione per l'intonazione italiana. In *Atti delle 5e Giornate di Studio del Gruppo di Fonetica Sperimentale* (pp. 85–98). Povo (TN), Italy.
- Beckman, M., & Elam, G. A. (1997). [Guidelines for ToBI Labelling](#). [On-line resource]
- Bigi, B. (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. In *the Phonetician* (111th–112/20 ed., pp. 54–69). International Society of Phonetic Sciences.
- Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. Retrieved from <http://www.praat.org/>
- Campbell, N. (1996). Autolabelling Japanese ToBI. In *ICSLP 96. Fourth International Congress on Conference on Language Processing Proceedings* (Vol. 4, pp. 2399–2402). Philadelphia: IEEE.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- CPrAN Team. (2016). CPrAN.
- De Looze, C. (2010). *Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais*. Aix-en-Provence. [PhD dissertation]
- Elvira-García, W. (2015). Remove tiers v.1. Retrieved from <http://stel.uab.edu/labfon/en/praat-scripts> [Praat script]
- Elvira-García, W., Roseano, P., Fernández-Planas, A. M., & Martínez-Celdrán, E. (in press, on-line first). A tool for automatic transcription of intonation: Eti\_ToBI a ToBI transcriber for Spanish and Catalan. *Language Resources and Evaluation*, 1–26. <http://doi.org/10.1007/s10579-015-9320-9>
- Escudero, D., Aguilar, L., Vanrell, M. del M., & Prieto, P. (2012). Analysis of inter-transcriber consistency in the Cat\_ToBI prosodic labeling system. *Speech Communication*, 54(4), 566–582.
- Escudero, D., Aguilar-Cuevas, L., González-Ferreras, C., Gutiérrez-González, Y., & Cardeñoso-Payo, V. (2014). On the Use of a Fuzzy Classifier to Speed Up the Sp ToBI Labeling of the Glissando Spanish Corpus. In N. C. (Conference Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA).
- Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica del español: una revisión del Sp-ToBI. *Estudios de fonética experimental*, 17, 264–283.

- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <http://doi.org/10.1037/h0031619>
- Hirst, D. (2011). The analysis by synthesis of speech melody: from data to models. *Journal of Speech Sciences*, 1(1), 55–83.
- Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonational Variation in Romance* (pp. 350–391). Oxford: Oxford University Press.
- Jun, S.-A. (2005). *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Jun, S.-A. (2014). *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Jun, S.-A., Lee, S., Kim, K., & Lee, Y. (2010). Labeler agreement in transcribing korean intonation with K-ToBI. In *Interspeech '10* (pp. 211–214).
- Lea, W. (1980). Prosodic aids to speech recognition. In W. Lea (Ed.), *Trends in Speech Recognition* (pp. 166–205). Englewood: Prentice-Hall.
- Lee, J., Kim, B., & Lee, G. (2002). Automatic corpus-based tone and break-index prediction using K-ToBI representation. *ACM Transactions on Asian Language Information Processing (TALIP)*, 1(3), 207–224. <http://doi.org/10.1145/772755.772757>
- Mertens, P. (2004). The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004, 23-26 March*. Nara (Japan).
- Pamies, A., Fernández Planas, A. M., Martínez Celdrán, E., Ortega-Escandell, A., & Amorós Cespedes, M. C. (2002). Umbrales tonales en español peninsular. In *Actas del II Congreso de Fonética Experimental* (Vol. Sevilla, pp. 272–278).
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. MIT, Cambridge, Massachusetts. [PhD dissertation]
- Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the tobi framework. *ICSLP*.
- Prieto, P., Aguilar, L., Mascaró, I., Torres Tamarit, F. J., & Vanrell, M. del M. (2009). L'etiquetatge prosòdic Cat\_ToBI. *Estudios de Fonética Experimental*, XVIII, 287–309.
- Prieto, P., & Cabré, T. (Eds.). (2013). *L'entonació dels dialectes catalans*. Rubí: Publicacions de l'Abadia de Montserrat.
- Rosenberg, A. (2010). AutoBI - a tool for automatic ToBI annotation. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association* (pp. 146–149). Mihama, Japan.
- Syrdal, A. K., & McGory, J. T. (2000). Inter-transcriber reliability of toBI prosodic labeling. In *INTERSPEECH* (pp. 235–238).
- Zhi, N., Hirst, D., & Bertinetto, P. M. (2010). Automatic analysis of the intonation of a tone language. Applying the Momel algorithm to spontaneous Standard Chinese (Beijing). *INTERSPEECH*.

## DIATECH TRESNA INFORMATIKOAREN BERTSIO BERRIA<sup>2</sup>

Gotzon Aurrekoetxea, Aitor Iglesias, Gotzon Santander eta Iker Usobiaga  
Euskal Herriko Unibertsitatea

[gotzonaurre@gmail.com](mailto:gotzonaurre@gmail.com), [aitor.iglesias@ehu.eus](mailto:aitor.iglesias@ehu.eus), [sunnyworld84@hotmail.com](mailto:sunnyworld84@hotmail.com),  
[ikeruso@gmail.com](mailto:ikeruso@gmail.com)

### Laburpena

Diatech hizkeren mugak zehatzeko tresna informatikoaren bertsio berria aurkezten da ondoko lerroetan. Bertsio berri honek tresna guztiz eraberritzen du, datuak inportatze sistema beretik hasiz, datuen maparatzea eta datuen analisi estatistiko indartsuagoetarako helduz. Datuen inportazioa erraztu egin da sistema berrian datutegi sinpleagoak sortuz. Diatech-en sortzen den proiektu bakoitzak datutegi bakar bat zein bat baino gehiago erabil ditzake. Mapak sortzean ezaugarri bakoitzaren koloreak aldatzeko aukera izango du erabiltzaileak bertsio berrian. Analisi estatistikoetan analisi multidimentsionalak ugaritu eta azkartu egin ditu (cluster hierarkikoez gain fuzzy clusterra gehituz, adibidez). Artikulu honek bertsio berriaren ezaugarriak jendarteratu nahi ditu.

**Gako-hitzak:** geolinguistika, dialektometria, softwarea, euskara

### Abstract

The aim of the contribution is to show the new version of the Diatech tool for dialectometry. This version includes a new format of the database, new tools to create conceptual maps and adds new statistics. The programming has been conceived again, beginning just from the data import tools. This import has been simplified, but allowing users to join more than one dataset in the statistical analyses. Taking advantage of the interactive maps users could change colours of the legend and take information of the selected locality. The new version implements more agglomerative multidimensional analyses, such as the fuzzy cluster and PCA analysis.

**Key words:** geolinguistic, dialectometry, software, Basque

## 1. Sarrera

Artikulu honetan *Diatech* web-aplikazioaren bertsio berria aurkezten da. *Diatech* Euskal Herriko Unibertsitateko EUDIA ikerketa-taldeak 2012. urtean sortutako (Ensunza, Unamuno eta Aurrekoetxea, 2013) tresna informatikoa da. Web-aplikazio hau sortzeko talde multidisziplinari bat egon da lanean: informatikari, estatistiko eta hizkuntzalariak Gotzon Aurrekoetxearen zuzendaritzapean (Aurrekoetxea et alii., 2015). *Diatechek* dialektoen eta hizkeren arteko desberdintasunak estatistikoki neurtu eta sailkapen automatikoa egiten ditu. Tresna hau on-line dago edonork eta edonondik erabiltzeko moduan datu propioekin analisiak egin ahal izateko. Azken urteotan *Diatechek* izan duen arrakastaren ondorioz (Aurrekoetxea 2016, Aurrekoetxea & Videgain, 2009; Iglesias, 2014; Ensunza, 2015...) EUDIA ikerketa Taldeak behar beharrezkoa ikusi du honen garapena bultzatzea.

Lerro hauetan aurkezten dena ez da *Diatechen* garapen bat, bertsio guztiz berritua baino; izan ere *Diatech v3* bertsioa hutsetik hasiz garatutako ekoizpena da eta. Programa berrian taldeak metatu duen esperientzia oso baliagarria izan da. *Diatechv3*-n aurreko bertsioeko ahuluneak konpondu egin dira, gaitasun estatistiko berriak sartu dira eta

---

<sup>2</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

interfazean baliabide gehiago jarri dira erabiltzailearen esku. Beraz, hobekuntzak ez ezik, berrikuntzak ere badauzka.

Hurrengo puntuetan azalduko dira informatikaren alorrean *Diatechen* bertsio berriak dituen ezaugarri teknikoak eta inplementatu diren analisi estatistiko berriak

## 2. Ezaugarri teknikoak

*Diatech* eguneratu egin da hainbat berrikuntza inplementatuz. Hala *Django* 1.8 bertsioa erabiltzen da orain, eta horrekin batera *Celery* ataza kudeatzailea, *RabbitMQ* barne mezularitza sistema, *Rosetta* itzulpen kudeatzailea, *Geopy* geolokalizaziorako plugina eta abar eguneratu dira.

Bide batez, eguneraketa sakonak aplikatu dira interfazean ere, bertsio berrian *Bootstrapframeworka* erabiltzen delarik. *Framework* honen bitartez interfaze garbiago eta intuitiboagoa diseinatu da. Honekin batera, web-diseinu moldagarria du orain aplikazioak eta interfazea sistema desberdinetara (ordenagailu, tablet, mugikorra...) egokitzen da.

Bertsio berriak mapak *SVG* formatuan gordetzen ditu eta erabiltzaileek bideratzen dituzten estatistika emaitzak testu bidez gordetzen dira. Emaitza horiek *Javascript* bitartez prozesatzen dira *SVG* mapa koloreztatuz. Aurreko bertsioan hau guztia hainbat irudien bidez egiten zenez askoz leku eta denbora gehiago behar zen eragiketa guztiak egiteko. Horrez gain, *SVG* formatuak erabiltzeak interaktibitatea ematen dio mapari eta bisualki erakargarriago egiten du. Berrikuntzekin batera, eta *Javascript*-ek funtzioek garrantzi gehiago dutenez, *JqueryJavascript* liburutegia erabili da. Hobekuntza hauetaz gain, *R* pakete estatistiko berriak ere gehitu dira.

Berrikuntza hauek guztiak erabiltzaile arruntak, batez ere, interfazean, datuen inportazioan, datuen kudeaketan eta emaitzen interaktibitatean igarriko ditu.

### 2.1. Interfaze berria

Esan bezala, *Diatechen* bertsio berrian *Bootstrapframeworka* erabili da interfaze argiago eta intuitiboagoa lortzeko (1 eta 2 irudiak).

Proiektuak

## Proiektuak

Gorde proiektu berria

Proiektu izena  Deskribapena

Proiektu irekia

Proiektuak

1

<input type="checkbox"/>	Zki	Proiektua	Jabea	Azken aldatze data	Proiektu irekia	Baimena	Eragiketak	Ezabatu
<input type="checkbox"/>	1	bourciez	gotzon	2016-05-24 14:01:18	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Aukerak	<input type="button" value="Ezabatu"/>

Proiektuak 1

- Bilatu erantzunak
- Estatistikak
- Erantzunen mapak
- Atalen indizea
- Kudeatu herriak
- Kudeatu proiektua

1. irudia: Diatech-en interfaze berria

Proiektua: Bourciez - Datu-base kudeaketa: Herriak

Herriak

Herria:  Latitueda:

Longitueda:

Mapa

Maparen Irudia

Mugak

Muga izena:

Aukeraturako muga:

Aukeratu orria: 1 Herriak: 138

<input type="checkbox"/>	Zki	Herria	Longitueda	Latitueda	Ezabatu	Kudeatu
<input type="checkbox"/>	1	Ahatsa-Altzieta-Bazkazane / Ahaxe-Alciette-Bascassan	-1,164269	43,148752	<input type="button" value="Ezabatu"/>	<input type="button" value="Kudeatu"/>
<input type="checkbox"/>	2	Ahetze / Ahetze	-1,5679129	43,405472	<input type="button" value="Ezabatu"/>	<input type="button" value="Kudeatu"/>

2. irudia: interfaze berria

Interfazean egindako aldaketen ondorioz, eta bertsio berriaren aplikazioak duen web-diseinu moldagarriari esker, interfazea egokitu daiteke sistemara; beraz, gaur egun hain ohikoak diren ordenagailu, tablet edota mugikorretan ere modu erosoan lan egin ahal izango da Diatech-en.



gehiagorekin) batera analisi bakarra egin. Beraz, bertsio berria komeni da eskuartearen dugun proiektua datu-base ezberdinetan igotzea. Hobeto ulertzearren adibide batekin azalduko dugu. Demagun proiektu batean herri bakoitzean genero (emakume eta gizon) eta adin desberdinetako (gazte, heldu eta zahar) lekukoen datuak bildu direla.

	Gazte	Heldu	Zahar
Gizon	1. csv	2. csv	3. csv
Emakume	4. csv	5. csv	6. csv

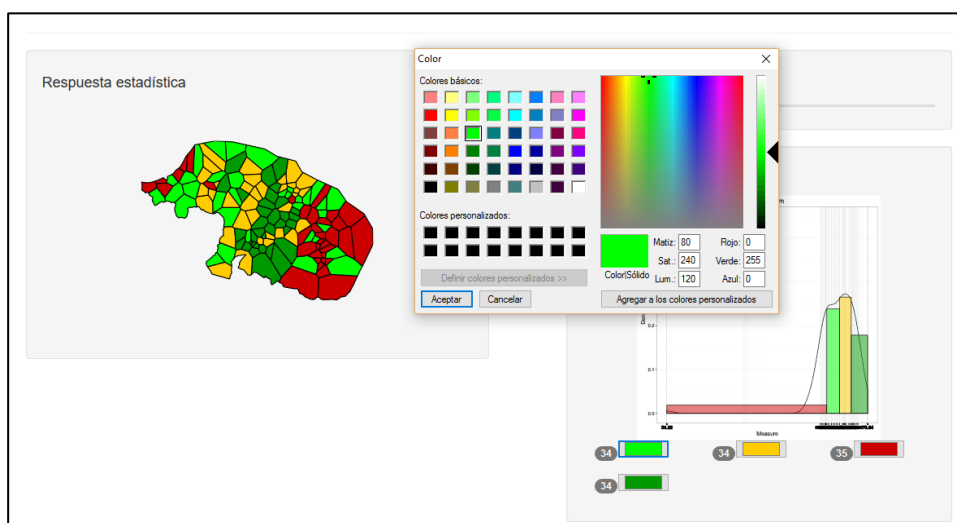
1. taula

*Diatech*-en bertsio berrira igo nahi izanez gero, datu hauek sei *csv* fitxategitan igo behar dira: genero eta adin bakoitzeko datuak bereizi behar dira igotzeko (ikus 1. taula). Horrela azterketa ezberdinak egiteko aukera desberdinak eskaintzen ditu tresnak: sei datu-baseak elkartzen badira analisi orokor bat egin daiteke. Baina era honetako egitura erabiliz, analisi orokorraz gain, analisi partzialak egiteko aukera ere izango da: gizonezkoen datuak bakarrik aztertu nahi badira, aldiz, 1. 2. eta 3. *csv* fitxategiak bakarrik erabiliko dira. Gazteen datuen analisia egin nahi bada 1. eta 4. fitxategiak elkartuz egingo litzateke. Eta horrela nahi den konbinazioaren azterketa egin daiteke.

### 3.3. Emaizen interaktibitatea

*Diatech*-en bertsio berriarekin sortzen diren mapak interaktiboak dira. Hori lortzeko sortzen dituen irudi eta mapak *SVG* formatoa erabiliz lortzen da. Horrenbestez, erabiltzaileek modu interaktiboan erabili ahal izango dituzte. Goebler VDM (<http://ald.sbg.ac.at/dm/germ/VDM/>) programaren antzera, mapan herri batean sakatzean herri horri dagokion informazioa eskainiko du tresnak.

Honetaz gain, aurreko bertsioan ez bezala, mapen koloreak aldatzeko aukera ere eskaintzen da (4. irudia). Baliabide honen bidez hizkuntzalari ikertzaileak analisiaren arabera sortu diren talde bakoitzari programak ematen dion kolorea alda lezake, bere irizpideak erabiliz; horretarako legendan edo historiograman aldatu nahi den gelaxkan sakatuz gero, koloreen paleta agertuko da eta nahi den kolorea aukeratzeko aukera zabalduko zaio ikertzaileari.



4. irudia: koloreak aldatzeko paleta



#### 4. Analisi estatistiko berriak

*Diatechen* aurreko bertsioan bost analisi geolinguistiko egin zitekeen: sinopsi mapak (hizkuntza similaritateen distribuzioa, trantsizio-eremuen analisia, eremu kontserbatzaileen analisia, geografia eta hizkuntza bariazioaren arteko korrelazioak...), izpien mapak, isoglosa mapak, cluster hierarkikoaren analisia eta MDS *multidimensional scaling* analisia. Bertsio berrian lau analisi gehiago inplementatu dira: Fuzzy klusterra, zentroidearen analisia, bariazio gehien eragiten duten ezaugarrien analisia eta Osagai Nagusien Analisia-ONA edo PCA deritzona.

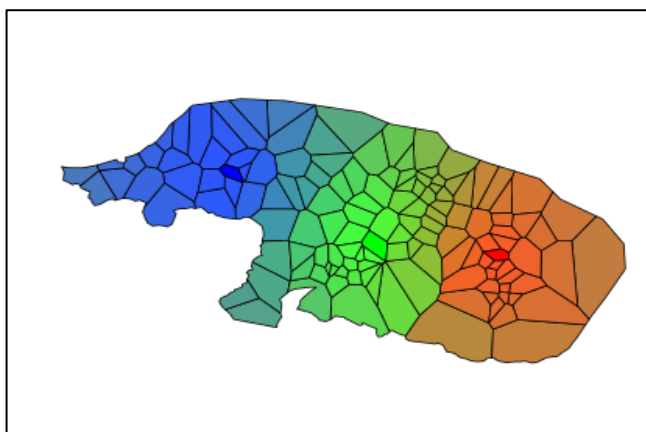
##### 4.1. Fuzzy klusterra

Fuzzy klusterra taldekatze algoritmo mota bat da non aldagai bakoitzak taldeekiko pertenezia maila lausoa duen.

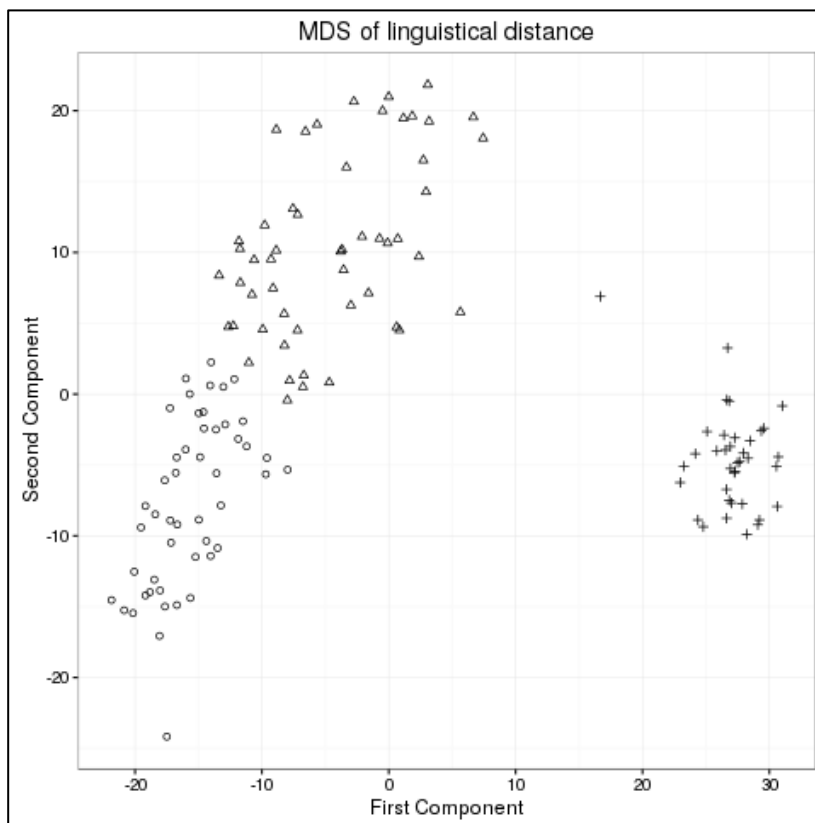
Fuzzy klusterrean, aldagai (herri) bakoitza K kluster desberdinetan sakabanatzen da, taldekatze diskretu bat ezarri ordez (kluster hierarkikoetan gertatzen den bezala). Taldekatzea hizkuntza distantzia izari baten gainean eratzen da (Levenshtein distantzia, adibidez). Aldagai bakoitzak kluster zehatz batekiko duen pertenezia kluster arteko distantzia minimizatuz ebazten da.

Kluster arteko distantzia minimizatzeak bi aurpegiko araua jarraitzen du. Alde batetik kluster berean pertenezia handia duten aldagaiek elkarren artean distantzia txikia izatea, eta bestetik, gehienbat kluster desberdinetan sailkaturiko aldagaiek distantzia handia izatea elkarren artean.

Kluster bakoitzari kolore bat ezarri eta, ondoren, aldagai bakoitzak kluster desberdinetan sakabanatua dagoen arabera kolore propio bat izango du. Adibidez, herri zehatz bat % 80an bada lehenengo klusterrekoa (gorria) eta % 20an bigarrenekoa (berdea), dagokion kolorea % 80 gorria eta % 20 urdin erabiliz sortuko da (5. irudia). Horrez gain, MDS analisi bat egingo da distribuzio geografikoaz gain bi dimentsiotan datuak erakusteko (6. irudia).



5. irudia: Fuzzy klusterra Bourciez proiektuan



6. irudia: MDS analisia

#### 4.2. Zentroidea

*Diatechen* analisi zentroidea inplementatzeko kanpoko lankidetzari esker lortu da<sup>3</sup>. Analisi honen arabera eremu dialektal baten edo kluster baten herrien artean hizkuntzari dagokion erdigunea zehazten da; hots, zein herrik biltzen dituen multzo horri dagozkion ezaugarri gehien edo, beste modu batean, zein herrik duen hizkuntza distantziarik txikiena besteekiko. Ezaugarri hori biltzen duen herria izango da eremu horren ordezkariak eredarriena.

Fuzzy klusterretan kluster bakoitzak aldagai (herri) bat ordezkari bezala erabaki dezake. Kluster K zehatz baten aldagai ordezkaria erabakitzeko, aldagai bakoitzak besteekiko duen distantzia (levenshtein, RIV...) ponderatu totalak kalkulatu dira. Ponderazioa aldagai bakoitzak K klusterrean duen pisuaren arabera sortzen da, zenbat eta pisu gehiago izan orduan eta garrantzia gehiago. Distantzia ponderatu hau txikiena duen herriak du klusterraren ordezkariak.

#### 4.3. Bariazio gehien eragiten duten ezaugarriak

Analisi estatistiko honen bitartez jakin daiteke zein den bariazio gehien sortzen duen ezaugarria, ezaugarri guztiek ez baitute bariazio berdina eragiten. Analisi honekin, azken finean, zein ezaugarri motak sortzen duen bariazio handiena jakitea lortu nahi

<sup>3</sup> Herri zentroideen eta bariazioaren eragileen ideia Esteve Clua (Universitat Pompeu Fabra) eta Miquel Salicrú (Universitat de Barcelona) ikertzaileei zor zaie. Haien burututako ideia *Diatech*-en taldeak garatu du.

da (fonetika, morfologia, sintaxia, lexikoa..) edota hizkuntza alor bakoitzean zein ezaugarri konkretu den.

#### 4.4. Osagai Nagusien Analisia

Diatechen implementatu den azken berrikuntza Osagai Nagusien Analisia (ONA), edo PCA (Principal Components Analysis) da. ONA aldagai anitzeko estatistikan erabiltzen den analisia da. ONA elkarrekiko korrelazioa duten aldagai multzo bat korrelaziorik gabeko osagai izeneko aldagai kopuru txikiago batez laburbiltzen duen teknika da. Bariantzaren bitartez neurtzen den hasierako aldagaien aldakortasuna, osagaietan ahalik eta gehien murrizteko helburua du. Osagai bakoitza aldagai guztien konbinazio lineal bat izango da eta batez ere elkarrekiko korrelazio nabarmena duten aldagaiekin izango da loturik. Analisi hau oso egokia da datu kopuru handiak erabiltzen direnean agerian ez dauden aldagaiak aurkitzen dituelako eta koerlazonatuta dauden aldagaiak aldagai independente berrietan bihurtzen dituelako.

### 5. Ondorioak

Aurrekoetxea et alii. (2015) adierazi legez, lan zientifikoa egin nahi bada, *Diatech* edo honen antzeko tresnarik gabe ezinezkoa da hizkeren arteko mugez jardutea.

*Diatech*-i buruz, esan behar da indarrean dagoen proiektua dela eta jadanik planifikatuta daudela analisi berriak inplementatzea. Berez, analisi geolinguistikoak egiteko garatutako tresna izanik, etorkizun hurbil batean analisi soziolinguistikoak egiteko aukera aurreikusten da. Etorkizuneko erronkarako prestatu da *Diatech*; orain faktore soziolinguistikoen araberrako analisiak egin ahal izateko egitura sortu da eta honek egingai bihurtuko du herri barneko bariazioa aztertu ahal izatea, faktore sozial desberdinak kontuan izanik.

Dialektologiaren aldetik etorkizun hurbileko diren erronka nagusien artean bariazio gehien sortzen duten ezaugarri linguistikoen patroia ateratzea kokatzen da. Emaitzetan aztertu beharko da ea hizkera guztietan patroia berdina ala emaitzak zer erakusten duten. Ezin bazter utzi hizkuntza ezberdinetan lortutako emaitzak konparagarri egitea, hizkuntza guztietarako balio duen azterketa egitea lortu behar da, hizkuntza ezberdinen emaitzak konparatu ahal izateko.

### 5. Bibliografia

- Aurrekoetxea, G. & Videgain, X. (2014). Outils por la géolinguistique automatisée. In Fabio Tosques, Fabio (Eds.), *20 Jahre digitale Sprachgeographie - Tagungsband* (53-66 orr.). Berlin: Humboldt-Universität zu Berlin, Institut für Romanistik.
- Aurrekoetxea, G. & X. Videgain (2009). Le projet Bourciez: Traitement géolinguistique d'un corpus dialectal de 1895. *Dialectologia* 2, 81-111.
- Aurrekoetxea, G., Iglesias, A., Santander, G. & Usobiaga, I. (2015). Diatech muga dialektalak zehazteko tresna. In G. Aurrekoetxea, A. Romero & A. Etxebarria (Eds.), *Linguistic variation in the basque education I* (136-148 orr.). Bilbao: Euskal Herriko Unibertsitatea.
- Aurrekoetxea, G., Fernandez-Aguirre, K., Rubio, J; Ruiz, B & Sanchez, J. (2013). DiaTech: A new tool for dialectology. *Literary and Linguistic Computing*, 28(1), 23-30.

- Ensunza, A. (2015). *Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan*, UPV/EHU.
- Ensunza, A., Unamuno, L. & Aurrekoetxea, G. (2013). Hizkuntzaren barietateak teknologia: EUDIA ikerketa-taldearen ekarria. In A. Etxebarria, U. Garay, A. Romero & I. Gaminde (Eds.), *Ondareaz eta hezkuntzaz I. Jardunaldiak* (99-16 orr.). Bilbao: UPV/EHU.
- Iglesias, A. (2014). *Igorreko hizkeraren azterketa dialektologikoa*. Bilbao: UPV/EHU argitalpen zerbitzua.

### **Webguneak**

- <http://getbootstrap.com>
- <http://www.celeryproject.org>
- <https://jquery.com>
- <https://pypi.python.org/pypi/geopy>
- <https://www.djangoproject.com>
- <https://www.rabbitmq.com>

# EUSKARAREN BARIAZIO GEO-MORFOLOGIAREN AZTERKETA<sup>4</sup>

Xarles Videgain eta Gotzon Aurrekoetxea

Université de Pau et des Pays de l'Adour (UPPA) eta Euskal Herriko

Unibertsitatea

[charles.videgain@univ-pau.fr](mailto:charles.videgain@univ-pau.fr), [gotzon.aurrekoetxea@ehu.eus](mailto:gotzon.aurrekoetxea@ehu.eus)

## Laburpena

Euskararen bariazio geolinguistikoari buruzko ikerketak fruitu ugari eta aberatsak eman ditu dialektologiaz diharduen literaturaren hasieratik beretik. Ikerketa hauek euskararen bariazio dialektalaren ezagutza ikaragarri hazi dute; hala ere, datu homogeneoen gabezia eta metodologia desberdinak erabiltzea eragozpen handia izan dira orain artean. Egun, *Euskararen Herri Hizkeren Atlasaren* (EHHA) lehen 7 liburukiek argia ikusi dutenetik, ikertzaileek hizkuntza datu kopuru erraldoi bat dute eskuartean, inoiz ez pentsatu bezalakoa, eta, gainera, prozedura sofistikuagoak erabiltzeko aukera dute datuon artean gertatzen den bariazioa ikertzeko. Ildo horretatik, ekarpen honetan EHHAREN 5. liburukiko izen deklinabideari dagozkion datuak erabiliz, ondoko prozedura erabiliko da: lehenik deklinabidearen datuen hizkuntza analisia egingo da, zeinen bidez datu fonetikoak erabili ordez fonologikoak erabiliko diren; horretarako, deklinabide atzizkien inbentarioa eta kasu bakoitzean erabiltzen diren arau fonologikoak aztertuko dira, fonologia sortzailea erabiliz. Datuen analisi kuantitatibo dagokionez, berrikitan sortu den *Diatech* tresna erabiliko da; tresna hau aukeratzeko arrazoietan nagusiena erantzun anitzak tratatzeko duen gaitasuna da. Izan ere, *Diatech* tresnak dialektometriari erabiltzen diren gainerakoak baino jaiotasun handiagoa erakutsi du herri berean galdera berarentzat erantzun bakarraren ordez erantzun gehiago agertzen direnean herrien arteko konparaketan modu zehatzagoan kontuan hartzeko. Eta kartografiaren bezainbestean, mapa kontzeptualez gain, mapa sintetikoak erakutsiko dira, atzizkien eta arau fonologikoen eraketa geografiko desberdinak erakutsiz eta, azkenik, hizkeren sailkapen hierarkikoa proposatuz.

**Hitz-gakoak:** bariazio geolinguistikoa, morfologia, dialektometria, cluster, euskara

**Key words:** linguistic regional variation, morphology, dialectometry, clustering, basque

## 1. Sarrera

Euskal dialektologiaren euskalkien sailkapenak garrantzi handia izan du Louis-Lucien Bonaparteren ondotik. Euskalkien sailkapen eta banaketa egiteko orain artean isoglosen metodoa erabili izan da; hots, datu gutxi batzuekin, eta ezaugarri aukeratuak erabiliz, ezaugarri gramatikalen eta datu lexikalen mugak zehaztuz euskalkien eremuak ezarri izan dira. Metodo intuitiboa izan dela esan daiteke.

Egun, aldiz, isoglosen metodoa gainditzen duten metodo berriak garatu ditu dialektologiak, metodo kuantitatiboez baliatuz, analisi multidimentsionalak egiteko gaitu da, teknologia berriak erabiliz. Ekarpen honetan ere bide hau jorratu da, euskalkien sailkapen zientifikoa eraiki nahian eta estatistika sailkatzaileak erabiliz. Lan hau, aitzitik, ez da lehena bide honetan, aurrekariak izan baititu Aurrekoetxea (1992) eta Aurrekoetxea & Videgain (2009), besteak beste.

Ekarpena ondoko ataletan banandurik aurkezten da: bigarren atalean erabili den metodologia azalduko da, hirugarrenean datuen analisia aurkeztuko da, laugarrenean

<sup>4</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

emaitzak ikusaraziko dira, bosgarrenean eztabaida deitu atala aurkezten da, emaitzei buruzko hausnarketak plazaratuz eta seigarrenean ondorioak aipatuko dira.

## 2. Metodologia

Atal honetan datuen jatorria, erantzun anitzen tratamendua eta datuen hizkuntza-analisia zertan mamitzen den aurkezten dira.

### 2.1. Datuen jatorria eta datu-basea

Erabili diren datuak Euskararen Herri Hizkeren Atlas (aurrerantzean EHHA) egitasmoko 5. liburukiko argitalpenetik hartuak izan dira (Euskaltzaindia 2013). Liburuki honetan izen morfologiari buruzko datuak aurkezten dira: deklinabide kasuak eta determinatzaileak, nagusiki. Ikerketa honetarako izen deklinabideari dagozkion datuak aukeratu dira. Galdera hauetatik 156 galdera hautatu dira. Esan behar da galdera batzuk baztertu egin direla, hutsune asko izan dituztenean. Irizpide gisa erantzunetan % 5 herritan edo gehiagotan hutsuneak izan dituzten galderak baztertzea erabaki da. Guztira 22.620 datu biltzen ditu sortu den datu-baseak.

Datu-basearen ezaugarriak zerrendatu behar balira hauek liriateke nagusienak:

a) Erantzunetatik lexema baztertu egin da eta atzizkia hartu da kontuan; zehatz izatearren, lexemaren azken bokala eta atzizkia hartu direla argitu behar da (horrela, *astoaren* erantzunaren kasuan *-oaren* baino ez da jaso datu-basean).

b) Erantzunak eta proposamen onartuak (herrikotzat joak) maila berean tratatu dira; uste dugu metodologia batek ahalbidetu dituen hitzak eta formak lekukoak beretzat hartzen dituztenean, nahiz iradokiak izan, herrikotzat jo behar direla.

c) Erantzunetan agertzen diren desberdintasun guztiak izan dira kontuan (*-oag* eta *-oak*, adibidez, biak onartu dira. Erantzunetan desberdintzat joz onartu badira ere, biak berdindu egin dira hizkuntza analisiaren ondoren, bi formak atzizki beraren bi burutzapen fonetikotzat joz.

d) Testuinguruak eragindako aldaketa fonetikoak ez dira kontuan hartu (hauen artean agertzen da *-m* bukaerakoa, adibidez; ez dira kontuan izan eta denak *-n* bihurtu dira).

e) Azentua ere kontuan izan da: azentuatuta gabeko eta azentuatutako atzizkiak desberdintzat hartu dira.

Sortu den datu-basearen egitura honela gauzatzen da ikus (1. irudia). Lehen zutabean item zenbakia, bigarreanean galderari dagokion kodea, hirugarrenean herriari dagokiona, laugarren zutabean erantzuna eta bosgarrenean hizkuntza-analisiaren ondorio den ezaugarria.

2429	17	119	-oangatik	-oerangatik
2440	17	120	-oi esker	-oi esker
2441	17	121	-oerangatik	-oerangatik
2442	17	122	-oerangatik	-oerangatik
2443	17	123	-u(n) amonkatik	-oeren amonngatik
2444	17	124	-oerangatik	-oerangatik
2445	17	125	-oerangatik	-oerangatik
2446	17	126	-oerangatik	-oerangatik
2447	17	127	-oerangatik	-oerangatik
2448	17	128	-u(n) amonkatik, -ua dela medio, -u(n) esker	-oeren amonngatik, -oa dela medio, -o(n) esker
2449	17	129	-o(n) kausaz	-oi dela kausa
2450	17	130	-oerangatik, -oerangatik	-oerangatik
2451	17	131	-oerangatik, -ua dela kausa	-oerangatik, -oa dela kausa
2452	17	132	-oerangatik	-oerangatik
2453	17	133	-gatik	-oerangatik
2454	17	134	-u(n) kausaz	-oi dela kausa
2455	17	135	-ua dela kausa	-oerangatik
2456	17	136	-ugatik	-oerangatik
2457	17	137	-u(n) kausaz	-oi dela kausa
2458	17	138	-ua dela kausa	-oi dela kausa
2459	17	139	-n hasi gatik	-oerangatik
2460	17	140	-lari	-oeren
2461	17	141	-u(n) kausaz, -u(n) kausaz	-oerangatik, -oi dela kausa
2462	17	142	-u(n) gatik	-oerangatik
2463	17	143	-u(n) gatik	-oerangatik
2464	17	144	-ugatik	-oerangatik
2465	17	145	-u(n) gati, ogati	-oerangatik
2466	18	1	-oikatik	-oerangatik
2467	18	2	-oikatik	-oerangatik
2468	18	3	-oikatik	-oerangatik
2469	18	4	-oikatik	-oerangatik
2470	18	5	-oikatik	-oerangatik
2471	18	6	-oikatik	-oerangatik

1. irudia: datu-basearen egitura

Taulak erakusten duen legez, desberdintasun handia dago erantzunen zutabetik analisiaren ondorengo zutabera: lehenengoan EHHako emaitzak agertzen dira (gorago aipatutako zehaztasunekin) eta bigarrenan burutzapen fonetikoak kontuan izan gabe, atzizkiaren forma edo haren ordezeko esamoldea biltzen da. Esan gabe doa, azken forma honen izaera ikertzaileak aukeratua dela, azaleko formaren azpiko egitura gisa eta horregatik beragatik desberdina ere izan daitekeela.

## 2.2. Erantzun ezak eta erantzun anitzak

Inkestagintzak bere arrakasta baldin badu ere, batzuetan huts egiten du, lekukoak zer esan ez dakielarik. Ondorioz, badira galderak zeinetan herri batzuetan ez den erantzuna bildu; izan dadin inkestagileari ahanzteagatik, izan dadin lekukoak une horretan ez duelako erantzunik izan, edo izan dadin bestelako erantzun bat eman duela. Azterketa honetan baztertu egin dira % 5 baino gehiagoko erantzun huts edo eza erakusten duten galderak. Endemikoa izanez, zubereraz bereziki ezagutzen diren sudurkariak ez dira tratatu. Salbuespen bakarra egin da “-ok” kasuari dagokionean; izan ere deklinabide hurbila ere erabili nahi izan dugu eta horregatik sartu da, nahiz erantzun huts anitz dituen, bereziki Iparraldean.

Aldiz, erantzun anitzak, hots, galdera berari eta herri berean erantzun bat baino gehiago jasotzean, denak izan dira kontuan hartuak. Batzuk aldaera fonetikoak dira, hala nola *astogaittiketaastogaitxik*; beste batzuk, ostera, deklinabide kasu desberdinetako atzizkiak, hala nola *iditzet* eta *iditako*. Denak onartu dira, inkestagileak emandako esaldiari erantzuteko lekukoak erabili dituen erantzunak izanik.

## 2.3. Datuen hizkuntza analisia

EHHako erantzunetatik datu-basearen sorrera izan da lehen urratsa: hitzaren azken hizki edo letra gehi atzizkia bildu dira, atzizki hutsak hartu ordez. Zergatia? Batetik deklinabide kasuak lortzeko hitz desberdinak erabili direlako (ezinezko da Euskal Herri guztian hitz bera erabiltzea ia kasu bakarrean ere), bestetik desberdintasun lexikoak izan lezakeen eragina baztertu gura izan delako izen morfologia aztertzean.

Diptongoetan, -i- bigarren osagaia eta atzizkia bildu da (“oi” eta “ai” diptongoez jaso dira hitzak eta). Baina lema sortzean “ai” bezala eratu da diptongoa. Eta horri erantsi zaio atzizkia. Adibidez, datuak biltzean erabili diren hitzak *zelai* eta *odei*, *trumoi* edo *sagarroi* izanik ere, lema *-aia* egin da.

Transliterazio fonetikotik idazkera ortografikora egin da EHHAn ohiuraz hartu dugun bidea jarraikiz. Basea eratzean ez dira datu fonetikoak hartu eta zuzenean ortografikoki idatzi dira argitalpenean fonetikoki idatzita dauden erantzunak.

Errangabe doa deklinabide atzizki desberdinen erabilera agertu dela. Adibidez, *-arengan* kasuaz gain, *-aren baitan* ere bildu da.

Datuen hizkuntza analisia izan da ondoko urratsa: erantzun fonetikoak (azaleko egitura) eta erantzun fonologikoak (sakoneko egitura) bereizten dira (SPE marko teorikoa erabili da hizkuntza analisia egiteko). Datuen hizkuntza analisia egin edo egin gabe aztertzeak hizkuntza distantzia desberdinak lortzen dira (Clua 2010). Hizkeren egituraren oinarritzea proposatzen da (sakoneko egitura baliatuz) eta ez datuetan (azaleko

egituran). Horretarako datuetatik abiatuz, hizkera bakoitzean zein hizkuntza ezaugarri erabiltzen den zehaztu behar da. Hizkeretan zehar erabiltzen diren deklinabide kasuetan antzekotasun handia dago. Lehen taulan agertzen diren erantzunetan dauden desberdintasun fonetiko (ebakerazko) asko hizkuntza analisiaren arabera ezabatu egin dira. Har ditzagun, adibidez, *-oáingatik*, *-óarengátik*, *-oaéngatik*, *-óarengátik...* erantzunak; denak “-oarengatik” atzizkiarekin ordezkatu dira hizkuntza analisisa egitean. Horrela jokatu da kasu guztietan.

Azterketa honetan morfologia baino ez da kontuan izan; ez dira arau morfonologikoak kontuan izan. Hauek aztertzeke dagoen fonologiaren atalean landu beharko lirakeke, beste ikerketa batean, alegia. Erabaki garrantzitsua da, datuen antolaketan ondorioak baititu.

### 3. Datuen analisi geolinguistikoa

Datuak aztertzeke erabilitako programa *Diatech* izan da. Programa honek eskaintzen dituen aukeren artean sailkapen hierarkizatu-automatizatua edo *cluster analysis* erabili da. Hizkeren sailkapena egiteko estatistika konputerizatuak eskaintzen dituen baliabideen artean indartsuena da *cluster analysis*. Beronen bidez azter-eremuko hizkerak hizkuntza ezaugarrien hurbiltasunen arabera hierarkia baten barnean kokatzen ditu, denak multzo bakarrean ezarri arte. Emaizta dendrograma deitzen den irudi batean agertzen du, zeinaren datuak ondoren mapa batera iraultzen diren.

Mapan hizkerak agerrarazteko Voronöi-ren poligonoen bidez adierazita agertzen ditu *Diatech*-ek. Voronöiren poligonazioa oso ezaguna da dialektologiaren hasiera beretik erabili izan delako (Goebl 1976). Mapa analitikoak (edo bestela esanda erantzun-mapak, kontzeptu bakoitzari mapa bat eskainiz) eta mapa sintetikoak (mapa askotako informazioa erabiliz sintesi-mapak eginez) eskaintzen ditu *Diatech*-ek. Eta hauen artean aukera asko eskaintzen ditu: isoglosa-mapak, izpi-mapak, similaritate mapak, sailkapen hierarkikoak, multidimension Scaling mapak...

Ikerketa honetan, hasteko, similaritate mapa pare bat aztertuko da (Santa Graziren eta Goizuetaren similaritate mapak). Mapa hauetan abiapuntua izango den herri batetik gainerako guztietara dagoen hizkuntza desberdintasuna neurtzen da eta desberdintasun hauek multzokatuz erakusten dira mapan, koloreak erabiliz.

Bigarren motatako mapak sailkapen hierarkikoa erakusteko mapak izango dira; *Cluster analysis* ere deitua dena. Prozedura honen bidez hizkera guztiak aldi berean elkarren arteko hizkuntza desberdintasunak neurtzen dituzte eta desberdintasun txikiena dutenengandik hasiz elkarrekin bilduz joaten dira, hizkera guztiak multzo bakarrean bildu arte. Sailkapen hierarkikorako algoritmo eta hizkuntza distantzia desberdinak erabiltzeko aukera du *Diatech*-en erabiltzaileak: algoritmoen artean *Ward*, *Complete linkage* eta *Average linkage* ditu eskura, eta hizkuntza distantzien artean *LD* (*Levenshtein distantzia*) (Heeringa 2014) eta *RIV* (*Relative Index Value*) (Goebl 1976). Lehen datu fonetikoak (erantzunak) erabili nahi direnerako eta bigarrena datu fonologikoak aztertzean dira erabilgarri (Aurrekoetxea, agertzeko).

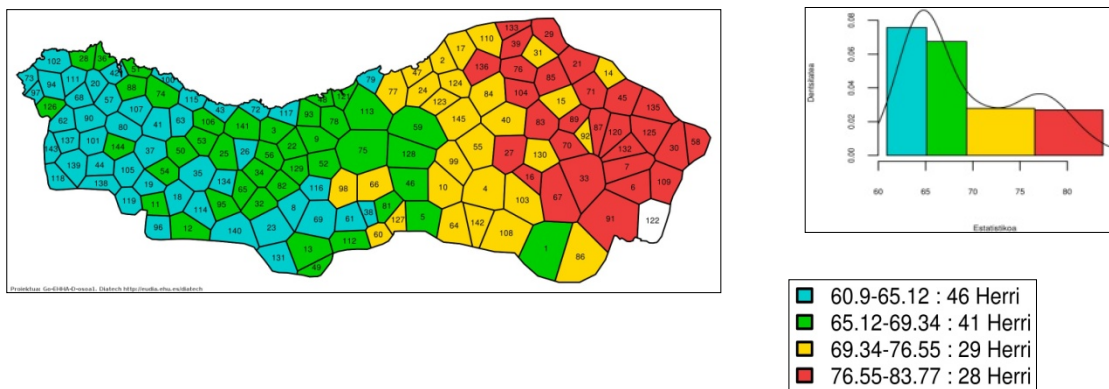
### 4. Emaizak



Azterketa honetako emaitzen artean *Diatech* tresnak eskaintzen dituen guztietarik bakar batzuk erabiliko dira, similaritate mapak (Santa Grazi eta Goizuetakoak) eta *cluster analisia* eskaintzen dituenak, hain zuzen ere. Similaritate mapetan abiapuntuan kokatzen den herri batek gainerakoekin dituen hizkuntza berdintasunak neurtzen dira eta eskala batean irudikatzen dira mapan. Cluster analisian, aldiz, hizkeren sailkapen multidimentsionala egiten da; hau da, hizkera guztiak kontuan izanik, elkarren arteko hizkuntza hurbiltasun edo berdintasunen arabera bilduz, hizkera guztiak hierarkia batean ezartzen dira, hizkuntza ahaidetasunaren arabera.

#### 4.1. Similaritate mapak

Santa Graziri dagokion similaritate mapan lau multzotan banandu dira emaitzak (ikus 1. mapa). Herri hau maparen eskuinaldean agertzen den poligono zuri batek ordezkatzen du. Santa Grazitik hizkuntza hurbiltasun handiena erakusten duten hizkerak Zuberoakoak eta Behe Nafarroako hainbat herri biltzen dira, Bardeze barne (denak gorritz koloreztatuak). Guztira 28 herri edo hizkera dira multzo honetan kokatzen direnak. Bigarren multzoan (horiz koloreztatuak) Iparraldeko gainerako herri guztiak eta Nafarroako herri batzuk agertzen dira. Hirugarren multzoan (orlegiz) nagusiki Gipuzkoako herriak agertzen dira, nahiz Bizkaiko bakan batzuk ere agertzen diren. Eta, azkenik, laugarren multzoan (urdinez) nagusiki euskararen mendebaldean kokatutako herriak agertzen dira, nahiz badiren erdialdean kokatutakoak ere.

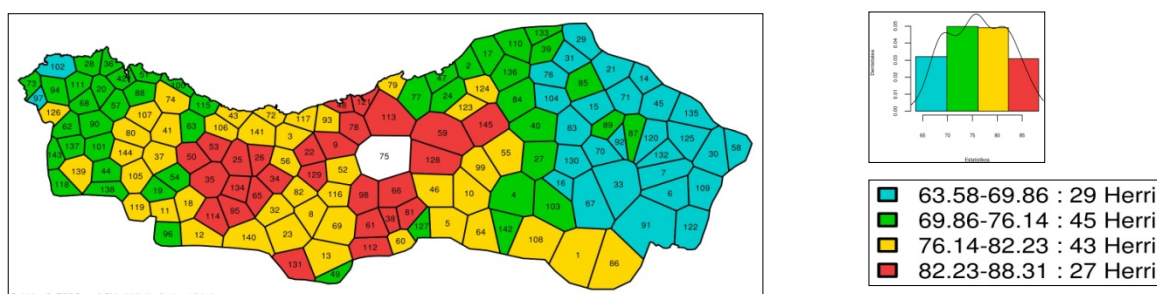


1. mapa: Santa Graziren similaritate mapa (RIV, 4 multzo)

Irakurlea ohartu behar da ezaugarri morfologikoen araberako distantzia neurtzen dela mapa honetan, izen deklinabideari dagozkion datuetan oinarritua. Gainera, datuok hizkuntza analisi bat jasan dute eta lekukoengandik jasotako datuak erabili ordez (azaleko egitura erabili ordez), azaleko datuen azpian dagoen hizkuntza ezaugarriak erabili direla (sakoneko egitura).

#### 4.2. Similaritate mapa: Goizueta

Goizuetari dagokion similaritate mapa eraikitzean Santa Grazikoan erabili diren baliabide berak erabili dira; hots, hizkuntza ezaugarriak erabili dira (eta ez lekukoengandik jasotako erantzun hutsak), RIV hizkuntza distantziak neurtzeko izaria, eta lau multzotan banandutako distantziak. Mapan erakusten den bezala, hizkuntzaz hurbilen diren herri gehienak (gorritz adierazita) geografikoki ere hurbil ditu. Guztira 27 herri dira:



2. mapa: Goizuetaren similaritate mapa (RIV, 4 multzo)

Bigarren multzoan (horiz) kokaturik agertzen diren herriak ere euskararen erdigunean kokatzen dira (43 herri), salbuespen gutxi batzuk gorabehera (Bizkaian eta Nafarroako ekialdean kokatuta). Hirugarren multzoan (orlegiz) mendebaldean herri andana bat eta Lapurdi-Nafarroan kokatzen dira gehienak (45 herri). Azkenik laugarren multzoan (urdinez) Ipar-ekialdean kokatzen dira, mendebaldeko bi izan ezik.

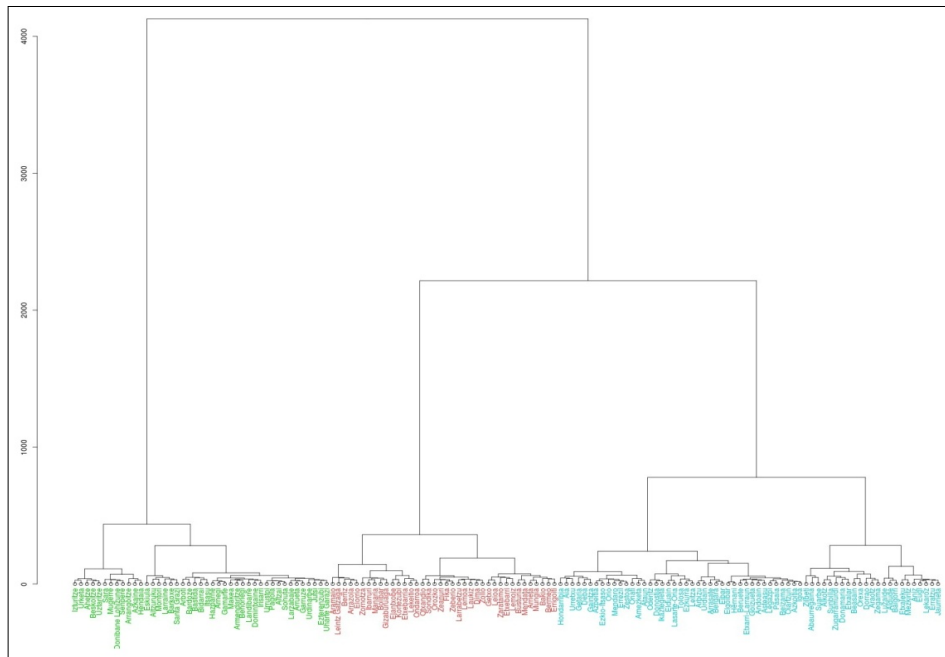
#### 4.3. Sailkapen hierarkikoa

Estatistikak aztertzen diren osagaiak (hemen hizkerak) sailkatzeko prozedura asko ditu. Prozedura hauek bi multzo nagusitan banatzen dira: deterministak eta probabilistikoak.

Sailkapen automatizatu-hierarkikoak erabiltzen dira hizkeren arteko sailkapenak egiteko. Prozedura hau osagaiak hierarkia baten barnean sailkatzeko estatistikak eskaintzen dituen baliabideen artean erabilieta da; gainera, zientzia alor guztietan erabiltzen da, giza-zientzietan, nahiz gainerakoetan. Hizkuntzalaritzan eta zehatzago dialektologian, barioazio geolinguistikoen analisisan aspaldi hasi ziren erabiltzen. Lehenengoetako aplikazioa H. Goebel, dialektologo austriarrari dagokio. Dialektometriari eskainitako ibilbide zientifikoan puntuen arteko analisiari buruz egindako ikerketen ondorioz “dialektometria dendrografikoa” deitu zion hizkeren sailkapen hierarkikoa lortu zuen lehen ikertzailea dugu (Goebel 1991). Hark izendaturiko “dialektometria dendrografikoa”, geroago “sailkapen hierarkikoa”, “sailkapen automatizatu” ere deitua izan da, baina “cluster analisia” terminoa da geroago orokortu dena, batez ere era honetako ikerketak ingelesez ere argitaratzen hasi zirenetik.

Harrezkero dialektologia kuantitatiboak urrats batzuk egin ditu, analisiak ugarituz eta prozedura berriak erabiliz. Hasiera bateko metodo deterministez gain, gaur egun metodo probabilistikoak ere erabiltzen dira, hizkuntza errealitatea modu zehatzago batez irudikatu nahiz. Zertan datza metodo determinista eta probabilistikoen arteko desberdintasuna? Bada, metodo deterministak hizkerak multzo batean edo bestean kokatzen ditu, hizkerak kokatuta dagoen multzoarekiko hizkuntza harremanetan sakondu gabe; hots, zein multzotan kokatzen den bakarrik zehazten du, multzoaren ezaugarriekiko adostasuna zehaztu gabe. Honek balio du hizkuntza muga zehatza eta argia denean, mugako alde bietako hizkeretan ezaugarri desberdin asko direnean eta zalantzarik ez denean hizkera bion artean hizkuntza muga dagoela.

Baina hori ez da beti gertatzen. Gehienetan dialekto muga ez da hain garbi eta argia izaten eta mugako bi aldeetako hizkeretan ezaugarri komun asko agertzen dira. Ondorioz, hizkera horiek ez dituzte eremu bereko gainerako hizkerak dituzten ezaugarri guztiak biltzen. Halakoetan muga dialektal argiaren ordean *continuum* bat egoten da eta *continuum* horretako hizkerak eremu bietako ezaugarriak izaten dituzte, batekoak ugariagoak izan arren. Estatistikan eredu probabilitistikoek kontuan hartzen dute egoera hau eta hizkera batek kokatuta dagoen eremu dialektalean zein ehunekotan hartzen duen parte adierazten du.



Analisi honetan metodo deterministak erabiliko dira, lehen urrats gisa. Zeren eta lehen helburua, izen morfologiari dagozkion ezaugarriekin, euskalkien eremuak identifikatzea den; euskararen kasuan zenbat dialektotan banan daitekeen jakitea oso garrantzitsua delako. Horrela, metodo hauek erabiliz, hizkera bakoitza eremu dialektal batean kokatuko da. Urrats honetan ez da kasu egingo eremu horretan zenbatetan hartzen duen parte; edo bestela esanda, zein mailatan dagoen integratuta eremu batean. Eremu dialektalak identifikatzea da lehen helburua.

Helburua hori izanik, lehen galdera ondokoa da: zenbat eremu dialektal daude deklinabideari dagozkion ezaugarriak kontuan izanik euskaraz? Edo hobe oraindik, nola identifikatu eremu dialektalak? Orain arteko euskalkien banaketan metodo intuitiboak erabili dira eta ikertzaile bakoitzak bere banaketa egin du, euskalkien kopurua ere desberdina delarik. Estatistika sailkatzaileak aldiz, metodo intuitiboak albo batera utziz, datu-base bateko hizkuntza ezaugarri kopuru handiak erabiliz eta hizkeren arteko hizkuntza ezaugarrien desberdintasunetan oinarrituz, sailkapena egiteko aukera ematen du. Analisi honetan *cluster* analisia erabiliko da, gorago esan bezala. Analisi mota honek eskaintzen dituen baliabideen artean dendrograma deitzen den irudia dugu (ikus 2. irud.). Dendrogramak hizkerak, linguistikoki hurbilen direnetatik abiatuz, hizkuntza hurbiltasunen arabera hierarkia baten barnean kokatzen ditu. Cluster analisisian hizkerak biltzeko algoritmo edo desberdintasunak kontatzeko metodo desberdinak ditu eskuragarri ikertzaileak.

Metodo hauen artean hizkuntzalaritzan ohikoena *Ward metodoa* da. Metodo honen oinarrian desberdintasunak kontatzean multzo batean diren hizkeren arteko puntu zentrikoena hartzen du: bi hizkera multzoren arteko distantzia neurtzean multzo bakoitzean diren hizkeren artean puntu zentrikoena hartzen du kontutan. Lortzen den irudia edo dendrogramaren behealdean aztergunean diren hizkera guztiak agertzen dira (hemen 145 hizkera agertzen dira). Hizkera bakoitzetik adar bertikalak sortzen dira eta adar horiek uztartuz doaz, gainerako hizkeren adarrekin hizkuntza hurbiltasunen arabera: zenbat eta beherago uztartu hizkera horien artean hizkuntza hurbiltasun handiagoa izango da (ezaugarri komun gehiago izango dituzte). Dendrograman hizkerak bilduz edo uztartuz joango dira, harik eta hizkera guztiak multzo bakarrean uztartu arte<sup>5</sup>.

#### 4.4. Euskalki kopurua

Behin dendrograma lortuz gero, urrats berri bat eman aurretik, ondo irakurri behar den irudia da, irudi honetatik abiatu behar delako euskalkien kopurua zehazteko prozedura. Jakinik dendrograma honetatik sortzen dela euskalkien mapa, euskalki kopurua egokiena zein den irudiak erakusten du. Irudian ikusten da, lehen mailtan (behe aldean) adar (edo marra) bertikalak oso txikiak direla. Adar txikiak hizkuntza desberdintasuna txikia dela esan nahi du: bi hizkera adar txikien bidez osatuta dagoen multzo batean sartzen badira, honek esan nahi du euren arteko hizkuntza desberdintasuna txikia dela. Adar luzeak, aldiz, hizkuntza desberdintasuna handia adierazten du: adar luzeek biltzen dituzten multzoen artean hizkuntza desberdintasuna handia da.

Dendrogramatik sortzen da euskalkien mapa. Dendrogramako datuak mapara eramateko, (ezkerrean kokatuta dagoen) eskalako maila jakin batean moztu behar da dendrograma: hiru adar mozten badira hiru eremu dialektal marraztuko dira mapan. Mozten diren adar bezainbat eremu sortuko da mapan.

Euskalkiak identifikatu nahi badira eta euskalkien artean dagoen hizkuntza desberdintasun handia izan behar dela uste bada adar luzeetan jarri behar dugu gure arreta. Zer erakusten du dendrograma honek? Bada, adar bertikalen luzerak hizkuntza desberdintasuna adierazten du: zenbat eta adar laburragoa izan hizkuntza desberdintasun txikiagoa izango da biltzen dituen multzoetan, eta zenbat eta luzeagoa izan, hizkuntza desberdintasuna handiagoa izango da. Beraz, estatistikan diren adituei jarraituz, dendrograma moztu behar balitz, mozketa egokiena adar luzeak agertzen diren lekua omen da.

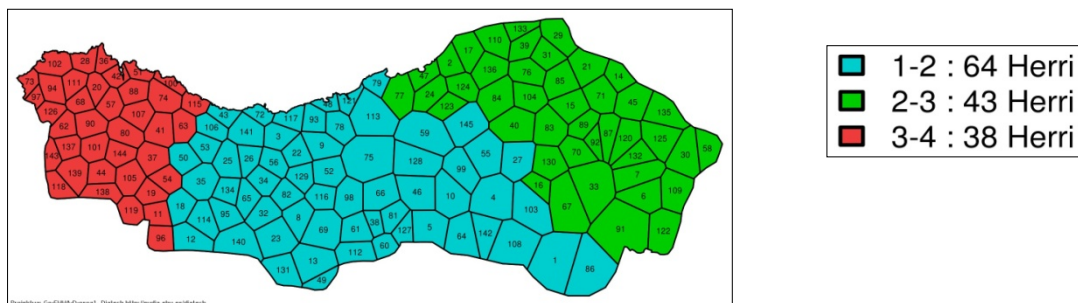
Irudiari ondo begiratzen bazaio, lehen adar luzeak hiru direla ohartuko gara. Adar horiek hizkuntza desberdintasun kopuru handia erakusten dute eta ondorioz pertinentiak dira multzoak bereizteko. Beraz, izen morfologiari buruzko datuen hizkuntza analisi edo modelizazioa egin ondorengo sailkapena egin behar bada hiru multzotan banandu beharko lirateke hizkera guztiak. Eta honenbestez, hiru euskalki daudela ondorioztatu behar da. Hortik beherako irakurketan azpieuskalkiak agertuko lirateke.

#### 4.5. Euskalkien mapa

<sup>5</sup>Hemen dendrograma ez da multzo bakarrean amaitzen, bi adarretan baizik. Adarren luzetasunak eraginik, moztu egin da eta ez da dendrogramaren azken parteak erakusten.

Dendrogramaren mozketak eskalako 1000 mailan egitea erabaki da, honenbestez. Maila horretan hiru adar luze agertzen dira; eta hiru adar moztu direnez, hiru eremu dialektaldun mapa eraikiko da.

Dendrogramaren mozketak maparatzeko kontuan izan behar da mapan marrazten diren eremuak geografikoki koherenteak izan behar direla<sup>6</sup>. Irudikatu den mapan hiru eremuak koherentzia handia erakusten dute.

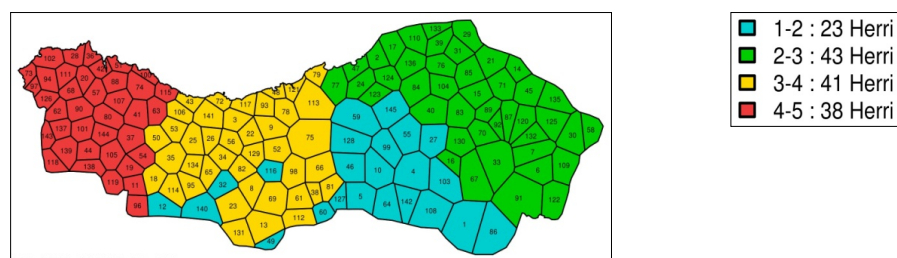


3 mapa: Izen morfologiari buruzko euskalkien mapa (Ward, RIV, 3 multzo)

Hiru eremuok hiru dialekto irudikatzen dituzte: mendebaldekoa, erdialdekoa eta ekialdekoa. Mendebaldekoak, Bizkaiko herri guztiak eta Gipuzkoako mendebaldekoak hartzen ditu (Eibar-54, Leintz-Gatzaga-96) eta Aramaio-11, Arabakoa; guztira 38 herri biltzen dira euskalki honetan. Erdialdeko euskalkiak Gipuzkoa osoa (salbu mendebaldekoan sartu diren herriak) eta Nafarroakoak barne hartzen ditu; guztira 64 herri biltzen dira euskalki honetan. Azkenik, ekialdeko euskalkia dugu, zeinetan Iparraldeko hizkera guztiak sartzen diren; guztira 52 herri.

Hiru eremu dialektal, hiru euskalki oso ongi bereziak eta egituratuta, eta eremu geografikoz ere oso orekatuak direnak. Kontuan izan behar da sailkapen hau izen morfologiari dagokiola; eta zehatzago esanik, izen deklinabideari dagokiona. Zeren ez baitira kontuan izan izenordainak eta determinatzaileak. Halere, uste dugu aski ordezgarria dela eta ziurtasun handikoa erabili den datu-kopurua handia delako eta sailkapena maila fonologikoan sakoneko egitura erabiliz egin delako.

Baina hiru eremu ordez lautan bananduko bagenu euskararen eremua beste honelako mapa agertuko litzaiguke (ikus 4. mapa).



4. mapa: Izen morfologiaren sailkapena 4 multzotan (Ward, RIV)

<sup>6</sup> Beste baterako uzten da eremu dialektal batek koherentzia geografikoa izatea behar-beharrezkoa duen ala ez.

3. mapatik bereizketa bakarra dago: hango kolore urdineko eremua bitan banatzen dela, gainerako biak (gorria eta orlegia) han zeuden bezala gelditzen dira oraingoan ere. Urdin eremua bitan zatitu da mendebaldeko eta ekialdeko hizkerak bilduz. Baina zatiketa ez da guztiz homogenea ikuspuntu geografikotik begiraturaz, ekialdeko eremuari dagozkion hizkerak agertzen baitira mendebaldeko eremuan.

## 5. Eztabaida

Asko eztabaidatu da Euskal Herrian euskalkien inguruan, baina gutxi euskalkien kopuruan. Onartuak izan dira bai L. L. Bonaparteren (1863-1869) aburua 8 euskalki izendatu zituenean, berdintsu gertatu da ondoren R.M. Azkuek 7tara ekarri zituenean (1905-06), eta orobat K. Zuazok 6tan (1998, 2003) eta gero 5etan (2008) ezarri dituenan, ekialdeko nafarra galdutzat eman ondoren.

Hauek guztiak isoglosen metodoa erabiliz eta prozedura intuitiboetan oinarrituz erabaki izan dute euskalkien kopurua, inoiz eztabaidatu gabe zergatik aukeratu duten kopuru hori. Orain prozedura berriak erabiliz eta teknologiak eskaintzen dituen baliabideak eskura izanik, une egokiagoan gaude euskalkien kopuruaz eztabaidatzeko, albo batera utzi gabe euskalki izendatzeko beharrezko diren ezaugarriak (Camino 2004, Aurrekoetxea 2009).

Irizpide eta prozedura zientifikoaren aldekoak garen neurrian ikertzaileak ematen dituen urratsen berri eman behar da; bestetik, hartzen dituen erabakiak argiak izan behar dira eta ondo oinarrituta egon. Bildutako datuen hizkuntza analisi edo modelizazioa egitean azalekoak diren datuak baztertu eta sakoneko egiturako ezaugarriak hartu dira kontuan: datu sistemikoak eta ez fonetikoak. Horrela hizkeren sistemetako ezaugarriak erabiltzen dira konparatzean, hots, hizkeren sistemak konparatzen dira eta ez datu fonetikoak. Bestalde, hizkuntza ezaugarriak datu-basean egituratu ondoren prozedura sailkatzaileak erabiltzea ezinbestekoa da, gaur egun estatistika konputerizatuak eskaintzen dituen baliabideak erabiliz. Tresna hauen bidez hizkeren hizkuntza ezaugarrien araberrako bilketa edo uztarketa jasaten dute, ikertzaileak ezaugarrien hauturik egin gabe. Datu-basean diren ezaugarri guztiak erabiltzen dira sailkapena egitean, bariazioa sortzen dutenak eta sortzen ez dutenak, hizkera guztietan edo gehienetan presente direnak eta hizkera batzuetan baino agertzen ez direnak, denak batera eta denak aldi berean hartzen ditu kontuan prozedura honek.

Ikertzaileak ere bere zeregina du, zalantzarik gabe, prozedura honetan. Haren zeregina da tresna estatistikoek eskaintzen dituzten baliabideak ondo eta argumentatuz hautatzea da. Hemen estatistika sailkatzaileetan erabiltzen den cluster analisia erabiltzea erabaki da, bera delako kopuru handiak izanik erabiltzen den prozedura egokiena. Cluster analisisian Ward algoritmoa erabili da, helburu honetarako modu egokiena delako, hizkuntza desberdintasunak hizkeren taldeko erdigunetik kontatzen dituelako, eta ondorioz, distantzia egokiena eskaintzen duelako. Hizkuntza distantzia gisa, RIV erabiltzea erabaki da; kontuan izanik erabiltzen diren datuek hizkuntza analisia jasan dutela eta datuak izan ordez ezaugarriak direla, RIV distantzia, distantzia nominala edo kategorikoa den neurrian, ezaugarri baten agerrera edo eza hartzen du kontuan. Bestela jokatzen du, adibidez, *LD* edo *Levenshtein distantziak*; hots, soinu distantzia den neurrian, *string distance* eta *edit distance* bezala ere ezagutua, datu fonetikoaren arteko distantzia neurtzeko oso baliagarria da: neurketa soinu baten agerrera, galtze zein aldatzearen arabera egiten du.

Cluster analisiaren ondorioz sortzen den dendrograma hiru multzotan banatzea erabaki da; horretarako clusterreko adarren luzera hartu da kontuan; izan ere, hizkuntza distantzia edo desberdintasuna adarren luzerarekin adierazten da dendrograman. Lehen hiru adar luze horiek adierazten dute ordezkatzeko dituzten hiru multzoen artean dagoen hizkuntza desberdintasuna handia dela eta benetako desberdintasun dialektala ordezkatzeko dutela.

## 6. Ondorioak

Hizkerak modu zientifikoan sailkatzeko erabiltzen diren tresna konputerizatuak erabiliz izen morfologiari dagokion sailkapen dialektala egiteko gai izan gara. Sailkapenerako erabili diren tresna eta prozedura objektiboek gainditu egiten ditu orain artean isoglosa metodoak erabiliz modu intuitiboan egin diren dialektoen banaketa.

Emaitzetan agertzen da izen deklinabideari dagozkion ezaugarriak erabiliz hiru multzotan banatzen direla hizkerak; beraz, hiru euskalkitan banatzen da euskara ezaugarri hauek erabiliz.

Emaitzan hauek ez datoz bat orain artean isoglosen metodoa erabiliz lortutako banaketekin. Aztertzeke dago desberdintasunen izaera eta emaitzen konparaketa zehatza egitea behar-beharrezkotzat jotzen da.

Ez dugu zalantzarik, etorkizunean metodo hauetan sakonduz eta gainerako hizkuntza alorretako datuekin sailkapenak egiten jarraitu behar dela, euskara bere osotasunean hartuz behin betiko sailkapena lortu arte.

## 7. Erreferentziak

- Aurrekoetxea, G. (1992). "Nafarroako Euskara: azterketa dialektometrikoa", *Uztaro* 5, 59-109.
- Aurrekoetxea, G. (2009). "Euskalkien dialektotasunaz", in Lakarra, J. / R. Etxepare / R. Gómez (arg.), *Beñat Oihartzabali gorazarre - Festschrift for Bernard Oyharçabal*, *ASJU* XLIII:1-2, 119-135.
- Aurrekoetxea, G. (2013). "Is a Scientific Measurement of Linguistic Boundaries Possible?", in Ernestina Carrilho, Catarina Magro and Xosé Álvarez (arg.), 2013, *Current Approaches to Limits and Areas in Dialectology*, Cambridge: Cambridge Scholars Publishing, 123-142.
- Aurrekoetxea, G. & Videgain, X. (2009). "Le projet Bourciez: traitement géolinguistique d'un corpus dialectal de 1895", *Dialectologia* 2, 81-111.
- Azkue, R.M. (1905-1906). *Diccionario vasco-español-francés*.
- Bonaparte, L. L. (1863-1869). *Carte des sept provinces basques...*, Londres.
- Euskaltzaindia (G. Aurrekoetxea & X. Videgain ard.; A. Arejita zuz.) (2013). *Euskararen Herri Hizkeren Atlas*, V, *Izen morfologia*, 1027-1272 mapak, Bilbo.
- Goebel, H. (1976). "La dialectométrie appliquée à l'ALF (Normandie)". In *Atti del XIV Congresso internazionale di linguistica e filologia romanza (Napoli 1974)*, Neapel, Amsterdam, vol. II, 165-195.
- Goebel, H. (1991). "Una classificazione gerarchica di dati geolinguistici tratti dall'ALS. Saggio di dialettometria dendrografica", *Linguistica* 31, 341-351.

- Camino, I. (2004). «Irizpide metodologikoak egungo euskal dialektologian», *Euskera* XLIX, 67-102.
- Clua, E. (2010). “Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal”. In G. Aurrekoetxea & J. L. Ormaetxea (arg.), *Tools for Linguistic Variation*, Bilbo: UPV-EHU, 151-166.
- Heeringa, W. (2004). *Measuring dialect pronunciation Differences using Levenshtein Distance*, Ph D dissertation, Rijksuniversiteit Groningen, Groningen.
- Zuazo, K. (1998). “Euskalkiak, gaur”, *FLV* 78, 191-233.
- Zuazo, K. (2003). *Euskalkiak herriaren lekukoak*, Donostia: Elkar.
- Zuazo, K. (2008). *Euskalkiak euskararen dialektoak*, Donostia: Elkar.



## SUBJECT PRONOUN EXPRESSION IN BASQUE: DESCRIPTION AND PEDAGOGICAL IMPLICATIONS<sup>7</sup>

Lorena Sainzmaza-Lecanda eta Itxaso Rodríguez-Ordóñez  
The Ohio State University eta Southern Illinois University  
[sainz-maza-lecanda.1@osu.edu](mailto:sainz-maza-lecanda.1@osu.edu), [rodrig52@illinois.edu](mailto:rodrig52@illinois.edu)

### Laburpena

Lan honek bi xede ditu: lehenik, euskarazko subjektuaren mintzagaiaren bariazioa deskribatzea du helburu, hots, subjektuaren aipua edota isilpena nola adierazten den aztertzen dugu euskara-gaztelera elebidun ezberdinen artean (jatorrizko elebidunak eta euskara umetatik eta nerabezarotik ostera ikasitako euskaldunberriak). Bigarrenik, subjektuaren adierazpena ondorengo euskaldunberriei irakasteko idea xume batzuk aipatzen ditu.

**Gako-hitzak:** Euskara, euskrazko subjektua, elebitasuna

### Abstract

The present study has two general goals: first, it aims at providing an account of the patterns of use of Subject Pronoun Expression (henceforth SPE) in Basque among different groups of bilinguals (native, early sequential speakers, and advanced and intermediate second language speakers of Basque). Second, it discusses a few pedagogical implications regarding how to deal with the variation between null and overt SPE in the classroom for future learners of Basque.

**Key words:** Basque, subject expression, bilingualism, pedagogy

## 1. Introduction

Basque, is a null subject language (NS), that is, a language that can leave the subject of a sentence unexpressed (Rizzi, 1986; D'Alessandro, 2015:201). For instance, example (1) shows a Basque utterance with a phonologically unrealized subject whereas example (2) shows the same example with an overtly expressed pronominal subject:

- (1) Ø [= *pro*] etxera noa
- (2) Ni etxera noa

One of the earliest proposals as to why a language allows null subjects has been introduced in the Government & Binding Theory within the generativist framework (Chomsky, 1981). Such framework was primarily interested in investigating whether the null subject parameter correlated with other syntactic phenomena or its interfaces (such as pragmatics or lexicon). As such, it has been proposed that the study of null subjects is an ideal candidate to study human cognition and syntax and its interfaces (Camacho, 2013).

Early proposals, such as those made by Rizzi (1986), postulated that null subjects are primarily licensed by verbal morphology and suggested that those languages showing richer verbal morphology (like Basque) will have the feature *pro*, represented

---

<sup>7</sup> The present paper is a modified version of the paper delivered at the EUDIA 5 workshop on *Linguistic Variation in the Basque Language and Education* so that it better addresses the focus of such workshop.

by an empty pronoun in canonical subject position that allows subject omission. Although this proposal has been challenged, including for Basque (Duguine, 2008), it allowed researchers to classify languages according to five different types of NS languages. The most studied type has been the *Canonical* type,<sup>8</sup> which covers those languages where a full referential subject can remain unexpressed, or null, and virtually any clause can be uttered leaving the subject unexpressed (D'Alessandro, 2015:203). Basque, like Italian, Spanish, Greek, Hausa, Turkish and Berber, has been considered one of those languages.

Such classifications may aid at grouping languages according to similar syntactic behaviors. However, studies on NSLs from a sociolinguistic perspective, whose objective is to study the linguistic and sociolinguistic factors that condition SPE (i.e. the overt and null realization of pronoun expression), have demonstrated that there is abundant variation with regards to the behavior of this phenomenon even within the same language. One of the most studied canonical NS language, for instance, is Spanish. Sociolinguistic research on the Spanish-speaking world has shown that speakers drastically vary both quantitatively and qualitatively in their pronominal subject usage. For instance, Puerto Rican Spanish shows the highest rates of SPE (45%; Cameron, 1992) followed by Dominican Spanish (41%), Cuban Spanish (33%), Ecuadorian Spanish (27%), Colombian Spanish (24%) (Otheguy et al., 2007), Madrid Spanish (21%; Cameron, 1993), Mexican Spanish (19%; Otheguy et al., 2007) and Spanish spoken in the Basque Country (15.3%; Sainzmaza-Lecanda, 2012). Qualitatively speaking, there seems to be a set of common factors that are relevant in conditioning the use of SPE. These include grammatical person and number, switch reference, distance from the previous mentioned referent, structural priming, tense-aspect-mood morphology and the lexical semantics of the verb (Carvalho et al., 2015). Different sociolinguistic factors have also been shown to be relevant in the use of SPE in Spanish. These include whether the speaker is bilingual (Carvalho and Bessett, 2015; Michnowicz, 2015; de Prada-Pérez, 2015), bilinguals' age of acquisition of the second language (Shin and Erker, 2015) or dialect and language contact (Otheguy et al., 2010; Otheguy and Zentella, 2012), among others.

These two views on NSLs collide in a very important way: whereas formal linguists within the generative framework have focused on the obligatory contexts in which subjects can be omitted, sociolinguists have been mainly interested in the variable use of SPE when the basic meaning of the utterance remains unchanged, also known as the *linguistic variable* (Labov, 1966). This latter approach regards the patterns of use that gives rise to linguistic variation as an important aspect of speakers' grammatical knowledge, which is not only constrained by linguistic factors, but also social ones. In this paper, we take this second approach and we examine the variable use of SPE in Basque through the lens of bilingualism. Our rationale is three-fold: first, the study of subject pronoun expression in Basque has been primarily described in formal theoretical terms (Ortiz de Urbina, 1989; Duguine, 2008); second, the increasing research on Basque-Spanish bilingualism has mainly relied on the development of the

<sup>8</sup>The other 4 types are: *Radical NSLs* (Chienes, Korean, Japanese, Vietnamese and Thai); *Partial* (Finnish, Marathi, Russian, Icelandic and Hebrew) (Holmberg, 2005, 2010); *Expletive* (Dutch, German, Afrikaans, Berbice Dutch Creole, Cape Verdean Creole, Jamaican Creole, Haitian Creole, Papiamentu, and Krio); *Non-NSL* (English, French, Swedish and Sindhi).

null subject parameter in child L1 and L2 acquisition, using native adult speakers of Basque as control groups (Ezeizabarrena, 2003, 2013; Iraola, 2015; Iraola and Ezeizabarrena, 2012), and has paid little to no attention to the actual variability with its overt counterpart. And, third, little is known about the linguistic patterns of adult Basque L2 speakers, a population that has rapidly grown in the last decades and continues to do so (Eusko Jaurlaritza, 2012). Importantly, the study of Basque SPE among different Basque-Spanish bilinguals has relevant pedagogical implications for Basque, especially because it is not a structure that is typically taught through instruction.

The remainder of this paper is structured as follows: section 2 presents a brief overview of previous studies looking at the Basque null subject parameter and subject pronoun expression. In section 3, we present the results of the present study. Finally, section 4 addresses the implications of the present study for the instruction of future Basque learners.

## 2. What do we know about Basque SPE?

Basque is considered a canonically consistent Null Subject language (Holmberg et al., 2009) showing a 3-way agreement that allows up to three pro-drop arguments (Ortiz de Urbina, 1989) because these arguments get encoded in the finite auxiliary verb through pronominal clitics. This is exemplified in (3a-b):

- (3) a. Harek niri asignaturak konbalideu eitzen (do)ztiez  
 b.  $\emptyset_i \emptyset_j$   $\emptyset_k$  konbalideu eitzen (do)-zt<sub>j</sub>-ie<sub>i</sub>-z<sub>k</sub>

(native, Gernika, 27)

It has been shown that Basque null subjects are not only subject to omission with inflected verbs, but can also appear in non-inflected forms such as nominalized t(z)ea contexts as in (4) (Duguine, 2008: 344). Note that in (4), the null subject in the embedded clause ( $\emptyset$ ) can have either a generic reading ('it is favorably seen to read books') or be interpreted as if some specific individual were performing the action of reading ('it is favorably seen that as I /you /he /she /we /you (pl.)/ they read books').

- (4) Begi honez ikusten dute [ $\emptyset$  liburuak irakurtzea]

While research has shown that although Basque null subjects are typically the preferred option, Basque overt SPE should not be considered an "idiosyncratic option" (Ezeizabarrena, 2013). In fact, there are certain contexts in which Basque null subjects are mandatory as it is the case of impersonal sentences (5) and weather constructions (6) (examples from Ezeizabarrena, 2013:314-315). Similarly, the overt realization of Basque subjects interacts with its phonological interface (i.e. intonation), in the sense that when the SPE is pronounced, it instigates either a topicalized reading (e.g. (6)) or a contrastive focus one (e.g. (7)):

- (5)  $\emptyset$  esan dute pretroleoa garestizera doala  
 (6)  $\emptyset$  bart gauean izotza bota du  
 (7) **Harek** egunero urten bier d-a-be!  
 (8) Karo, ezta ba  $\emptyset$  pozik bizi!**Nibe** biziko nintzen pozik han!

Most experimental research on Basque SPE comes from child acquisition studies (Ezeizabarrena, 2003, 2013; Iraola, 2015; Iraola and Ezeizabarrena, 2012). In one of her latest work, Ezeizabarrena (2013) studied the development of subject expression in Basque among two Basque monolingual children (L1) and two early balanced Basque-Spanish bilinguals (2L1). Her results showed that overt subject pronouns appear from early on 45% of the time and remain quite constant until the age of 3:6 in Basque. Interestingly, it was shown that different bilinguals did not show clear differences with respect to Basque SPE rates, but differences were found with respect to the type of subject: demonstrative subjects were the ones with higher rates of overt expression (31.5%), followed by one-word DPs (29.5%) and personal pronouns (25.4%). Based on these results, it was concluded that the pro-drop parameter is set since birth for Basque, and that the children must learn *when* and *for what communicative purpose* they should use the subject overtly.

Adult L2 Basque speakers may be presented with the same challenge. However, in the absence of any empirical study that demonstrates how the variable use of Basque SPE is conditioned, it would be a hard task to find a productive way to incorporate an explanation of this grammatical construction into teaching. In fact, pedagogical materials on the teaching of Basque as an L2 have rarely addressed the use of overt or null subjects. In those cases, some resources only mention that Basque ‘can omit subjects’ (Olaizola and King, 1996: 11) whereas others briefly explain ‘how to use the subject’ or mention its obligatory uses. For instance, in Euskaltzaindia’s series of *Euskal Grammatika – Lehen urratsak*, it is mentioned that overt subjects should be used for emphatic purposes (Euskaltzaindia, 1991: 362) or that it is obligatory when the referent is new information (Euskaltzaindia, 1991: 363). Similarly, teaching materials developed by HABE show that subject expression needs to be used with *ere*, *behintzat*, *behinik behin* and similar constructions.<sup>9</sup>

These grammatical accounts can be helpful in understanding the use of Basque SPE in certain contexts, but they tell us little about the existing variation that Basque experiences with respect to this phenomenon. Therefore, the present study aims at providing an account of the patterns of variable use of SPE in Basque among different bilinguals with the humble hope that it will bring further implications for teaching Basque to future adult learners.

### 3. The study

In pursuing a comprehensive description of the linguistic and extra-linguistic factors influencing the variability between overt and null pronominal subjects in Basque, a quantitative corpus-based investigation of naturally-occurring data was conducted. To this end, the spontaneous speech of 25 adult Basque-Spanish bilinguals, gathered through semi-guided sociolinguistic interviews, was analyzed in the quantitative computation.

#### 3.1. Data collection

<sup>9</sup> HABE Materials developed explaining how to use pronoun in Basque [http://www.ikasbil.eus/web/ikasbil/hizkuntza-arauak-fitxa?p\\_p\\_id=56\\_INSTANCE\\_Oz2o&p\\_p\\_lifecycle=0&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_col\\_id=column-1&p\\_p\\_col\\_count=1&articleId=444167&groupId=10138](http://www.ikasbil.eus/web/ikasbil/hizkuntza-arauak-fitxa?p_p_id=56_INSTANCE_Oz2o&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&articleId=444167&groupId=10138) [last accessed July 10, 2016].

Sociolinguistic interviews are one of the core methodologies for data collection used in the study of language variation. As opposed to reading tasks, these interviews aim at eliciting a more casual speech from the subjects. They usually constitute recorded one-to-one conversations between the interviewer, who is conducting the study, and the informant, who is the interviewee. While casual and spontaneous speech is the target of this elicitation procedure, these interviews tend to be semi-guided conversational exchanges, wherein the researcher asks specific questions in order to trigger the unconscious production of specific phonetic or (morpho-)syntactic variables.

Within the current project, Basque speakers were interviewed following this technique and constitute part of a larger subject pool studied in Rodríguez-Ordóñez (2016). Informants were asked to talk about themselves, their childhood memories, their future summer plans as well as more controversial topics such as the current national economic crisis. Before the interviews were recorded, all participants completed a language background questionnaire soliciting information about language use, education, self-reported proficiency and language background of family members in addition to a 24-item multiple-choice test in order to measure their Basque proficiency, which consisted of questions selected from multiple levels of the standardized *Euskararen Gaitasun Agiria* (EGA) test. Through the scores obtained in the latter, these 25 speakers were divided into 4 bilingual groups according to their Basque proficiency scores as shown in table 1.

<b>Intermediate (Bilbao area)</b>	<b>Advanced (Bilbao area)</b>	<b>Early-sequential (Bilbao area)</b>	<b>Native (Bilbao area)</b>	<b>TOTAL</b>
7 speakers	6 speakers	6 speakers	6 speakers	25 speakers

Table 1. Distribution of participants

The intermediate group was made up of those subjects who, out of 24 possible points, received a score between 12 and 16 points. The advanced group consisted of individuals whose scores ranged between 17 to 21 points. And, finally, those who scored 21 points or more were categorized as early sequential or native, depending on their age of acquisition. On the one hand, speakers were considered native if they began learning Basque from birth, while they were classified as early sequential bilinguals if they began learning Basque through emersion programs at the age of 3. All native speakers came from the rural Gernika area and, therefore, spoke the local Basque variety. Conversely, early sequential bilinguals and second language learners (L1-Spanish, L2-Basque) were chosen from surrounding areas of the capital of the Bizkaian province, Bilbao (i.e. Getxo, Las Arenas, Algorta, Barakaldo and Portugaleta) and spoke *Euskara Batua*, that is, the standard variety promoted by *Euskaltzaindia*.

### 3.2. Methodology: data transcription and quantification

The recorded interviews were first transcribed by one of the authors. The researcher produced a word-by-word written document of each interview using the annotation system offered by ELAN, a professional tool for the creation of complex annotations on video and audio resources. Once all 25 transcriptions were compiled, the data scrutiny and quantification were carried out. In order to analyze under which linguistic contexts subject pronouns were produced, from each transcribed interview, the researchers noted in a separate excel document approximately 100 finite verbs which co-occurred with either an expressed or an unexpressed pronominal subject,

yielding a total of 2,121 extracted tokens.

In order to account for *bona fide* variation, the *principle of accountability* couched within the Sociolinguistics tradition was put to practice. Hence, tokens were only selected if they constituted part of the variable context; that is, a context that could equally be expressed by overt and null pronominal forms as shown in (10).

(10) Ødeseroso sentitu nintzan igual etxean lehenengo egunatan eta garesti zen eta Øjoan nintzen etxera. Ni nire etxean ondo nau, egia esan.

In this scenario, all three tokens could have been equally expressed through the use of either a null or an overt subject pronoun (*ni*). The speaker chose to use the null form in the first two tokens and employed the overt *ni* in the third occurrence. However, not every context exhibits the pronominal variability observed in (10). In fact, certain contexts contain mandatory overt and null subjects as explored in section 2. Following traditional accountability practices, these non-variable contexts were excluded from the analysis. Weather expressions such as *Øeuria ari zuen*, for instance, were discarded from the analysis because these verbs never take pronominal subjects in Basque. Similarly, sentences that contained focalized subjects were also excluded because the latter can never be null.<sup>10</sup>

### 3.3. Methodology: Coding criterion

After circumscribing all variable and non-variable contexts, it was coded whether these 2,121 extracted tokens occurred with a null or overt subject. Subsequently, we explored the linguistic contexts in which these expressed and unexpressed pronouns occurred. First, we looked at the properties of the finite verb forms and we noted the person and number (e.g. 1<sup>st</sup> sing., 2<sup>nd</sup> sing., etc.), the semantic class (i.e. stative: *izan, eduki, egon*; mental: *uste, pentsatu, jakin*; dynamic: *atera, jarraitu, joan*) and tense-aspect-mood configurations of each finite form. Additionally, we also considered the effects of discourse-pragmatic factors such as the expressed or unexpressed nature of the preceding subject (i.e. ‘structural priming’) as well as its referent; that is, whether the preceding subject had the same or different person and number (i.e. ‘switch reference’). To illustrate an example, let’s take the second token in (10): ...*eta Øjoan nintzen etxera*. This token was coded as follows: first, it was labeled as a “null” token; next, the verb was identified as a dynamic verb form with a perfective past tense configuration in the 1<sup>st</sup> person singular form. Finally, the preceding form in the discourse utterance, represented by *Ødeseroso sentitu nintzan*, was specified as a “null” subject containing the “same” referent (i.e. 1<sup>st</sup> person singular).

### 3.4. Results: What factors condition the choice between an overt or null subject pronoun?

After conducting a series of statistical analyses, our results revealed that overt vs.

<sup>10</sup>Other excluded contexts are: existential *egon*, subject relative clauses and imperative verbs. Likewise, contexts that could potentially lead to skewing effects were also excluded: verbs with non-personal subjects, demonstrative subjects (*hau* ‘this’, *hori* ‘that’, etc.), set phrases (e.g. *ez dakit* ‘no sé’), false starts and repetitions.

null SPE in Basque does not behave homogeneously across all 4 bilingual groups.<sup>11</sup> In other words, the proportion of subject used and the linguistic factors that influence the presence or absence of a subject are different across groups and, hence, heavily dependent on the level of Basque proficiency.

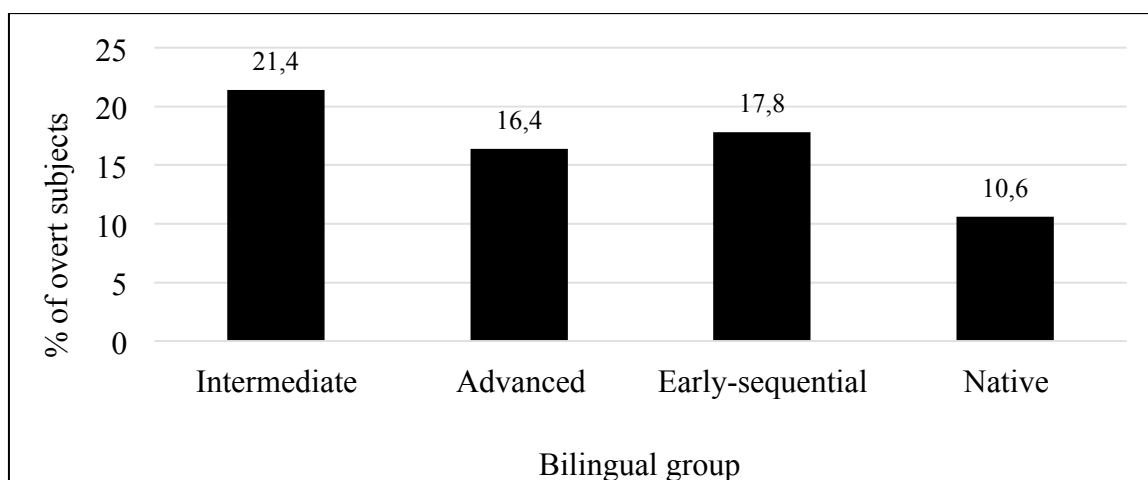


Figure 1: Frequency rates across bilingual groups

As displayed in Figure 1, the proportion of overt subjects increases as Basque proficiency decreases: intermediate learners and early sequential bilinguals use significantly more overt subjects than Basque native speakers. This frequency rate differences may be attributable to a processing cost effect, as has been suggested for other languages (Sorace, 2011): less proficient speakers use more overt subjects in order to alleviate processing costs (e.g. keep track of the subject referent, make accurate subject-verb agreement patterns, etc.).

Disparities across groups are further reflected in the kinds of linguistic factors that constrain the use of overt and null subjects in Basque. Very importantly, the observed variation between overt and null pronominal forms is not free or *ad hoc*; rather, it is shaped by particular linguistic constraints as summarized in Table 2.

Intermediate	Advanced	Early-sequential	Native
<i>Verb semantics:</i> Stative + Mental	<i>Verb semantics:</i> Stative + Mental	<i>Verb semantics:</i> Stative	<i>Verb semantics:</i> Stative
<i>Structural priming:</i> Overt > Overt Null > Null	<i>Structural priming:</i> Overt > Overt	<i>Reference:</i> Switch > Overt	<i>Reference:</i> Switch > Overt
<i>Reference:</i> Switch > Overt	<i>Reference:</i> Switch > Overt	<i>Person &amp; Number:</i> 1 <sup>st</sup> sing, 3 <sup>rd</sup> sing	
<i>Person &amp; Number:</i> 1 <sup>st</sup> sing, [+spec]	<i>Person &amp; Number:</i> 1 <sup>st</sup> sing, 2 <sup>nd</sup>		

<sup>11</sup> A series of statistical analyses were performed using R statistical software. First, the data were submitted to a hierarchical analysis of constraints and, subsequently, four individual mixed effects model for each bilingual group were generated. For details, see Rodríguez-Ordóñez & Sainzmaza-Lecanda (under review).

2<sup>nd</sup> sing

sing

Table 2: Linguistic constraints conditioning overt vs. null subject expression

Among intermediate Basque learners, overt pronominal subjects are more likely to be used when verbs convey a stative or mental meaning and take either the 1<sup>st</sup> or 2<sup>nd</sup> person singular form as shown in example (11). Regarding discourse-pragmatic factors, the presence of a subject with a different referent in the preceding clause increases the probability of producing an overt subject. Likewise, we found a structural priming effect such as that in (12), where a preceding overt pronominal form (i.e. *nik*) primes the realization of a subsequent overt subject (e.g. *ni*) and, similarly, a preceding null subject triggers another null subject (e.g. (13)).

- (11)  $\emptyset_{3rd-sg}$ ausartia da, **ni** harrituta nago neska honekin.  
 (12) ...**nik**<sub>1st</sub> ulertzen dut baina **ni**<sub>1st-sg</sub> ikasi nuen, batua... eta karo  
 (13) Eta gero,  $\emptyset_{3rd-pl}$ joan ziren etxera eta  $\emptyset_{1st-sg}$ geratu nintzen nire lagunarekin...

As for advanced learners, the same number of linguistic factors condition subject pronoun expression. Yet, the internal arrangement of these constraints is different. On the one hand, the likelihood of expressing an overt pronoun is significantly higher with mental verbs and verbs in the 1<sup>st</sup> person singular form. Both the referent of the preceding verb form and the null or overt nature of the previous subject have an effect on subject expression. Yet, unlike the intermediate group, our findings suggest that the odds of expressing a null subject are higher if the preceding finite verb holds the same referent and it is also null, as indicated in (14). This *null > null* effect seems to disappear if the preceding form has a different referent. In effect, as shown in (15), if the preceding form is null but holds a different person and number configuration, the subsequent pronominal mention tends to be overt.

- (14) ...gero  $\emptyset_{1st-sg}$  Zarautzera joan ginen. Hondartzara ere,  $\emptyset_{1st-sg}$  hondartzan egon ginen.”  
 (15) ... $\emptyset_{3rd-pl}$  horregaitik digitalizatu nahi dituzte. **Nik** egun batean eskatzen dut

As we move on towards more proficient speakers, sharper differences emerge. First and foremost, unlike intermediate and advanced L2 Basque learners, neither early sequential bilinguals nor Basque natives show any signs of a structural priming effect; hence, the overt or null form of the previous subject does not condition the form of the subsequent pronominal subject. For early sequential bilinguals, overt subject pronouns are more probable if 1) the verb expresses a state, 2) if the verb is the 1<sup>st</sup> or 3<sup>rd</sup> person singular or 3) if the previous verb carries a different referent. And, finally, native Basque speakers only prefer the overt pronominal variant if there is a switch in the referent of the subject or if the verb is stative as in (16). Interestingly, person and number specifications of the finite verb form do not seem to condition subject expression among native speakers according to our results.

- (16) **gu**<sub>1st-pl</sub> ikasliek izen gara, irakasliek izen baino arinau eta **zuk**<sub>2nd-sing</sub> ikusten zunien..”

In a nutshell, while native Basque speakers use overt pronominal subjects



minimally (e.g. 10,6% overt subjects) and few linguistic factors constrain the variability, as Basque proficiency decreases, not only does the amount of overt subjects raise but also the number of linguistic factors constraining their realization. This trend is especially visible among L2 Basque learners, whose pronominal variation is led by a more complex set of constraints.

#### 4. Summary and pedagogical implications for the Basque classroom

While Basque is indeed a null subject language, this study has shown that overt pronominal subjects are also frequent in Basque and that the observed variation between null and overt subjects is neither random nor unconstrained. Specifically, it has been revealed that less proficient Basque speakers use the overt pronominal variant significantly more often than native speakers who have acquired the language since birth. The higher proportion of overt subjects among L2 Basque speakers is also constrained by a larger set of linguistic constraints such as person and number specifications of the finite verb and the overt or null shape of a previous subject. Despite these differences, variable SPE in Basque seems to be constrained by one common factor among all four groups: if the previous subject has a different referent, Basque speakers tend to produce a subsequent overt mention.

In our view, these findings pose at least two questions from a pedagogical standpoint: should subject expression be addressed in the classroom? And, if so, how should we do it? We believe that the fact that L2 students use subjects ‘more than necessary’ in casual speech deserves some attention, just like lack of ergative marking does. While increased rates of overt subject expression seem to be the norm among less proficient L2 speakers, it might be worth showing these L2 learners that in Basque, beyond mandatory contexts such as focalization, overt SPE primarily respond to discourse-pragmatic effects such as switch reference. In the long run, by making students aware of these patterns of use, we might be able to facilitate a more native-like usage of this grammatical component of language.

Yet, in the possibility of explicitly teaching SPE in Basque classrooms, one should cautiously notice that in early stages of L2 development it is quite unrealistic to expect native-like use of Basque SPE. The acquisition of any language is not abrupt but gradual and it is contingent upon an interaction between cognitive processes and social experience. With respect to SPE, it has been shown that bilinguals learning a null subject language tend to ‘overuse’ overt pronominal subjects in order to reduce processing costs, by virtue of keeping the referent in mind (Sorace, 2011), and that way, achieve communicative goals in the most efficient manner. When it comes to input, research also shows that in early classroom teaching, teachers often times ‘modify’ their input, also known as ‘teacher talk’ in order to provide a more ‘comprehensible input’. The use of such input is of paramount importance for the acquisition of L2 grammar (Krashen, 1982, 1985; Ellis, 1985) because it sets a model for the learning process. Therefore, it is quite possible that teachers also ‘overuse’ the SPE in providing their input to their students. While teachers’ input could be helpful in early stages of the acquisition of Basque SPE, we think that this input could likewise be complemented with explicit explanations on the role that pronominal subjects play with regards to discourse-pragmatic functions (e.g. switch reference) once students have achieved a threshold in which they no longer need to repeat the subject to reduce processing costs.

At what point in the acquisition process such threshold occurs, however, is a matter of further research.

## 5. References

- Camacho, J. (2013). *Null Subjects*. Cambridge: Cambridge University Press.
- Cameron, R. (1992). Pronominal and null subject variation in Spanish: Constraints, dialects and functional compensation. (Ph.D. Dissertation, University of Pennsylvania).
- Cameron, R. (1993). Ambiguous agreement, functional compensation and non-specific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change*, 5, 305-334.
- Carvalho, A. M., & Bessett, R. M. (2015). Subject pronoun expression in Spanish in contact with Portuguese. In A. M. Carvalho, R. Orozco & N. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 143-165). Georgetown: Georgetown University Press.
- Carvalho, A. M., Orozco, R., & Shin, N. (2015). Introduction. In A. M. Carvalho, R. Orozco & N. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. xviii-xxvi). Georgetown: Georgetown University Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- D'Allesandro, R. (2015). The null subject parameter: Where are we and where are we headed? In A. Fábregas, J. Mateu & M. Putnam (Eds.), *Contemporary linguistic parameters* (pp. 201-226). London / New York: Bloomsbury.
- Duguine, M. (2008). Silent arguments without *pro*: The case of Basque. In T. Biberauer (Ed.), *The limits of syntactic variation* (pp. 311-330). Amsterdam / Philadelphia: John Benjamins.
- Ellis, R. (1985). *Understanding Second Language Acquisition*. Shanghai: Shanghai Foreign Language Education Press
- Euskaltzaindia. (1991). *Euskararen gramatika - lehen urratsak I*. Bilbao: Euskaltzaindia.
- Eusko Jaurlaritzak. (2013). *V. inkesta soziolinguistikoa*. Vitoria-Gasteiz: Eusko Jaurlaritzaren Argitalpen Zerbitzu Nagusia.
- Ezeizabarrena, M. J. (2003). Null subjects and optional infinitives in Basque. In N. Müller (Ed.), *(In)vulnerable domains in multilingualism* (Hamburg Studies on Multilingualism ed., pp. 83-106). Amsterdam / Philadelphia: John Benjamins.
- Ezeizabarrena, M. J. (2013). Overt subjects in early Basque and other null subject languages. *International Journal of Bilingualism*, 17(3), 309-336.
- Holmberg, A. (2005). Is there a little *pro*? evidence from Finnish. *Linguistic Inquiry*, 36, 533-64.
- Holmberg, A. (2010). Control into finite clauses in partial null-subject languages. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (Eds.), *Parametric variation: Null subjects in minimalist theory* (pp. 125-152). Cambridge: Cambridge University Press.
- Holmberg, A., Nayudu, A., & Sheehan, M. (2009). Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica*, 63(1), 59-97.
- Iraola, M. (2015). Anaphora resolution in children and adults: An experimental study of mature speakers and learners of Basque. (Ph.D. Dissertation, University of the Basque Country).
- Iraola, M., & Ezeizabarrena, M. J. (2012). La resolución anafórica en la adquisición del euskera. *Empiricism and Analytical Tools for 21st Century Applied Linguistics*.

- Selected Papers from the XXIX International Conference of the Spanish Association of Applied Linguistics*, pp. 65-76.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. New York: Longman
- Labov, W. (1966). *The social stratification of English in New York city*. Washington D.C.: Center for Applied Linguistics.
- Michnowicz, J. (2015). Subject pronoun expression in contact with Maya in Yucatan Spanish. In A. M. Carvalho, R. Orozco & N. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 101-119). Georgetown: Georgetown University Press.
- Olaizola, B., & King, A. R. (2010). *Colloquial Basque: A complete language course (colloquial series)*. New York: Routledge.
- Ortiz de Urbina, J. (1989). *Parameters in the grammar of Basque*. Dordrecht: Foris.
- Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
- Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language contact in Spanish in new york: Toward the formation of a speech community. *Language*, 83, 770-802.
- Otheguy, R., Zentella, A. C., & Livert, D. (2010). Generational differences in pronominal usage in spanish reflecting language and dialect contact in a bilingual setting. In M. Norde, B. de Jonge & C. Hasselblatt (Eds.), *Language contact: New perspectives* (pp. 45-62). Amnsterdam: John Benjamins.
- Prada-Pérez, A. d. (2015). First person singular subject pronoun expression in Spanish in contact with Catalan. *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 121-142). Georgetown: Georgetown University Press.
- Rizzi, L. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry*, 17(3), 501-557.
- Rodríguez-Ordóñez, I. (2016). Differential object marking in Basque: Grammaticalization, attitudes and ideological representations. (Ph.D. Dissertation, University of Illinois at Urbana-Champaign).
- Sainzmaza-Lecanda, L. (2012). Rates and constraints on subject expression in Basque Spanish. (M.A. Thesis, The Ohio State University).
- Shin, N., & Erker, D. (2015). The emergence of structured variability in morphosyntax: Childhood acquisition of Spanish subject pronouns. In A. M. Carvalho, R. Orozco & N. Shin (Eds.), *Subject pronoun expression in Spanish: Cross-dialectal perspectives* (pp. 169-190). Georgetown: Georgetown University Press.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1-33.

## BARIAZIO PROSODIKOAREN ERAGINA TESTU IRAKURRIEN ULERMENEAN<sup>12</sup>

Aintzane Etxebarria, Asier Romero, Iñaki Gaminde eta Urtza Garay  
Euskal Herriko Unibertsitatea

[aintzane.etxebarria@ehu.eus](mailto:aintzane.etxebarria@ehu.eus), [a.romero@ehu.eus](mailto:a.romero@ehu.eus), [inaki.gaminde@ehu.eus](mailto:inaki.gaminde@ehu.eus),  
[urtza.garay@ehu.eus](mailto:urtza.garay@ehu.eus)

### Laburpena

Lan honetan bariazio prosodikoak testu irakurri bat ulertzeari zer neurritan eragiten dion aztertu da. Horretarako, Bilboko Irakasleen Unibertsitate Eskolako 151 ikaslek bi galderatako ulermen proba egin dute Senpereko, Urdiango, Oiartzungo eta Bermeoko neska gazteek euskaraz irakurritako berriaren grabazioa erabiliz. Lehenengo, estimulu horien ezaugarriak zehazten dira eta talde prosodikoen muga tonuak, eten nahiz lapsus motak adierazten dira Praat softwarearen bitartez egiten diren transkripzioetan, eta ondoren, ulermen galderen erantzunak aztertzen dira SPSS programa estatistikoa erabilita. Azkenik, ondorioetan irakur daitekeen bezala, estimuluen ezaugarri suprasegmentalak oso antzekoak izan dira, ez ordea, iparraldekoaren ezaugarri segmentalak, horrek ez du eragin ezberdintasun esanguratsurik erantzun zuzenen kopuruan, baina ikusten da alde txiki bat.

**Hitz-gakoak:** prosodia, bariazioa, testu irakurriak

### 1. Sarrera teorikoa

Prosodia ezaugarri suprasegmentalen multzoa da, osagai horiek azentua, intonazioa, tonua, erritmoa, melodia, etenak, lokuzio abiadura eta ahots kalitatea dira (Llisterri, 2016). Elementu horien bidez ahozko mezua zenbait alderdiren berri ematen zaigu, besteak beste, hiztunaren adina, generoa, jatorria eta maila soziala (Mozziconacci, 1998: 3). Hori dela eta, hiztunak ahotsa era koherentean, adierazkortasunez eta testuingurura egokituta erabili behar du (Mozziconacci, 1998; Grau i Tarruell, M. eta Vilá i Santasusana, M., 2009; Saá, 2011; Garay y Etxebarria, 2010, Gaminde et al. 2011, 2012), ezaugarrietako batean hutsuneren bat badago, Saáren arabera gaixotasun edota urritasun bat dagoela pentsa dezakegu:

“[La prosodia] Comprende la melodía, tono, rapidez, pausas e inflexiones conferidas al discurso. También ayuda a llevar estados emocionales o afectivos en las oraciones. En las lesiones del HD, el lenguaje es anormalmente plano y monótono, sin matices o pausas adecuadas. No logran transmitir el afecto apropiado al discurso.” (Saá, 2011)

Era berean, gaitasun prosodikoa komunikazio gaitasunaren alderik garrantzitsuenetakoa da ahozko mezua ondo ulertzeko nahiz adierazi nahi ditugun jarrerak, emozioak, esaldi motak adierazteko, besteak beste. Ildo horretatik Hizkuntzen Europako Erreferentzi Marko Bateratuak zehazten dituen gaitasun linguistikoarekin, diskurtsiboarekin, soziokulturalarekin eta pragmatikoarekin lotura du; izan ere, hizkuntza aldetik azentuaren, intonazioaren, fokuaren eta perpausaren egituraren inguruko nondik norakoak ematen dizkigu; gaitasun pragmatikoari dagokionez, hizketa funtzioaren berri ematen du eta Fujisakik (2004) zehaztu duen informazio paralinguistikoaz zenbait xehetasun jakin ditzakegu, hala nola, hiztunaren jarrera eta

<sup>12</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

intentzioa, hizketaldian parte hartzen dutenen distantzia soziala eta botere erlazioa (Garay eta Etxebarria, 2010, Gaminde et al. 2011, 2012). Horretaz gain, gaitasun prosodikoak tonuaren eta etenen bidez egituratzen du ahozko testua, zentzu horretan gaitasun diskurtsiboarekin du lotura. Azkenik, gaitasun soziokulturalarekin ere erlazionatzen dugu gaitasun prosodikoa, bariazio linguistikoaren arabera ezaugarri ezberdinak baititu (Gaminde et al. 2011). Era honetan, prosodia gaitasuna gainerako en gainetik eta lotura duela esan dezakegu ahozko testuari dagokionez.

Zenbait ikerketak bai hizkuntzalaritzaren (Gaminde eta lank., 2014), bai komunikazioaren (Rodero, 2013), bai psikologiaren alorretatik (Klauda & Guthrie, 2008; Miller & Schwanenflugel, 2006, 2008; Rasinski, Rikli & Johnson, 2009; Veenendaal et. al., 2014; Schwanenflugel et. al. 2004; Benjamin & Schwanenflugel, 2010) gaitasun prosodikoa garatzeko beharrezko azpimarratu dute, batez ere ahoz gora irakurtzen dena ulertzeko eta irakaskuntzaren eremura mugatuta, haurren garapen kognitiboari laguntzeko (Etxebarria et. al., 2011; Cova, 2004). Zenbait ikerketak ulermen-maila egokia izateko F0k esaldien bukaerako norabide aldaketak eta etenek garrantzia dutela ondorioztatu dute (Miller eta Scwanenflugel, 2006, 2008; Schwanenflugel et. al., 2004; Benjamin eta Schwanenflugel, 2010), hala ere, talde prosodikoenezagarriek ere (etenek, muga-tonuek, abiadak, luzerak, tonu-motak, bariazio linguistikoa) ulermenari nahiz pertzepzioari nabarmenki eragiten diotelako ustea dugu.

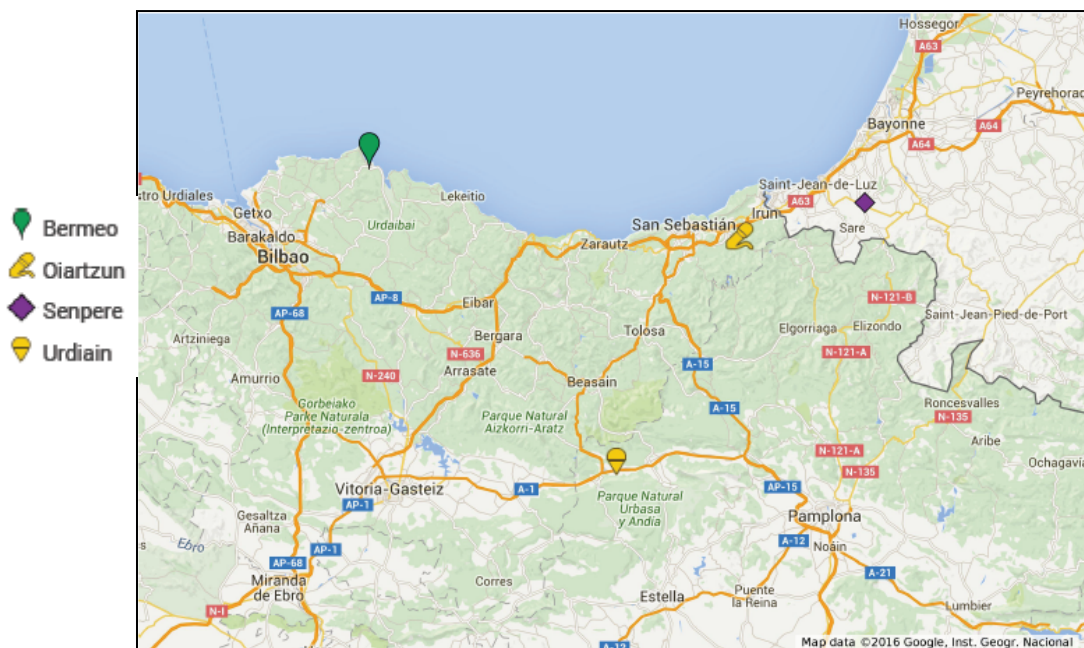
Beraz, lan honen helburua hau da: bariazio prosodikoak, erarik eta zabalenean ulertuta eta talde prosodikoen aipatutako ezaugarriak kontuan izanik, ahoz gorako irakurketaren ulermenari nola eragiten dion aztertzea. Horretarako, sarrera teoriko honen ondoren, metodoaren nondik norakoak azaltzen dira bi ataletan banatuta, batetik, estimuluei dagokien atala, eta bestetik, ulermen probei dagokiena. Azkenik, aurreko ataletako bi motatako emaitzetan oinarritutako ondorioen berri ematen da.

## **2. Metodoa**

Atal honetan estimulutzat erabili den corpusaren ezaugarriak, ulermen probaren nondik norakoak nahiz aplikazioa, eta emaitza nagusien berri ematen da.

### *2.1. Corpusa*

Bariazio prosodikoak testu irakurrien ulermenean eraginik duen ala ez ikusteko grabatutako ahozko corpusa erabili zen, corpus hau euskaldun zaharrak diren lau neska elebidun gaztek (hemezortzi eta hogeita bat urte bitartekoak) irakurritako ahoz gorako testuek osatzen dute, jatorriari dagokionez, ondoko mapan (1. irudia) agertzen diren Bermeo, Oiartzun, Urdiain eta Senpere herrietakoak dira:



1. irudia: Corpusa osatu duten parte-hartzaileen jatorria.

Parte hartzaileek irakurritako testua Berria egunkaritik ateratako kirol albistea zen:

“Urte asko pasatu beharko dira beste pilotari batek Aimar Olaizolari pilotari onenaren titulua kentzeko. Zazpi final jokatu ditu, zazpi txapel jantzi eta kaiolan etorkizuneko belaunaldiek bideoan miretsiko duten abilezia lortu. Hori ez du lortu bakarrik jokatu, aurkari gogorreirei aurre eginez baizik. Horixe falta zitzaion txapeladun handiari, gogoan hartzeko moduko final bat.” (<http://www.berria.eus/>)

Parte hartzaileen ahozko testuei dagokienez, honako ezaugarriok dituzte:

A) Bermeo

Bermeoko parte-hartzaileak hamar talde prosodiko egiten ditu, silaba kopuruari dagokionez batez besteko kopurua hamalau silabakoa da. Denborari dagokionez, irakurketaren denbora osoa 25000 ms-koa da, talde prosodikoen denbora 21060 ms-koa, etenen denbora 3500 ms-koa, lapsusen denbora 440 ms-koa. Talde prosodikoen mugako tonuei dagokienez, talde prosodikoen bukaeran tonu baxua egiteko joera du (L eta HL), tonu altua behin baino ez du egiten (H). Tonuarekin batera egiten diren etenei dagokienez, +Isilunedun etena / -bokal-txertaketa edo luzapena (%) erakoa da gehien agertzen dena. Isilunerik ez dagoen etena bokal-txertaketarekin (V\$) nahiz txertaketarik gabe (\$), kasu banatan agertzen da, lehengoa H muga tonuarekin eta bigarrena HL erakoarekin. Lapsusei dagokienez, ondorengo hitzaren errepikapena (aβijeş) egiten du, ez du etenik egiten banaezinak diren sintagmetan. Adierazitakoa, 1. taulako transkripzioan ikus daiteke:

“urte aşko paşatu βearko đira βeşte pilotari βatek aimar olaişolari pilotari (L%) (eh)onenaren titulua kenđtfeke (L%) (eh) şaşpi final đokatu đitu (L%) (eh)şaşpi tşapel đanđfi (v) (H)(\$) eta kajolan etorkişuneko βelaunaldieđ bideolan miretşiko đuten (HL)(\$) (lap)(eh)aβijeşia lortu (L%) (eh) ori ez du lortu βakarig đokatus(L%) (eh) aurkari yoşoređ aurre yiņez baışik (L%) (eh)orişe falta şitşajonđ

tʃapeɫdun aŋdiari (L%) (eh) goyoan artʃeko moðuoko final bat (L%)”.

1. taula: Bermeoko parte-hartzailearen transkripzio fonetikoa.

#### B) Oiartzun

Oiartzuneko irakurketak hamazazpi talde prosodiko ditu, zortzi silabakoa da talde prosodikoen batez besteko kopurua, guztira 23660 ms irauten du eta ms horietatik 20080s talde prosodikoei dagokie, etenen denbora 3580 ms-koa da eta ez du lapsusik egiten, ezta sintagmen nahiz hitzen barruko etenik. Talde prosodikoen mugetan goranzko tonuak nahiz beheranzkoak ia neurri berean egiten ditu, hamar eta zazpi hurrenez hurren. Tonuekin batera agertzen diren etenak direla eta, gehienak % erakoak dira (hamairu), hau da, isilunedunak eta bokal-txertaketarik gabeak. Gainerakoak, isilunedunak eta bokal-txertaketadunak (hiru) eta isilunerik ez duen baina bokal txertaketarik ere ez duen (\$) eten bakarra dago. Ezaugarriok, 2. taulako transkripzioan ikus daitezke:

“urte aʃko paʃatu bearko ðia beʃte pelotari batek aɪmar olaiʃolari (H%) (eh) piɫotari (v) (H%) (eh) onenaren titulua kentʃeko (L%) (eh) ʃaʃpi final xokatwitu (L%) (eh) ʃaʃpi tʃapel xaŋʃi (H%) (eh) eta (L%) (eh) kajolan (H\$) (eh) etorkiʃuneko belauñaldiek (H) (\$) biðeoan (H%) (eh) miretʃiko ðuten (v) (H%) (eh) aβileʃia lortu (L%) (eh) ori ez ðu lortu βakarik xokatuʃ (L%) (eh) auɾkari ʋoʋoreɪ (v) (H%) (eh) auɾeʋinez βaiʃik (H%) (eh) oriʃe faɫta ʃiʃaʃjon<sup>j</sup> tʃapeɫdun aŋdiari (L%) (eh) goyoan (H%) (eh) artʃeko moðuoko final bat (L%)”.

2. taula: Oiartzuneko parte-hartzailearen transkripzio fonetikoa.

#### C) Urdiain

Urdiaingo parte-hartzaileak hamar talde prosodiko egiten ditu, 13 silabako batez bestekoa egiten du irakurleak, testuak 20110 ms-ko iraupena du, eta talde prosodikoen denbora osoa 17070 ms-koa da, tartean egiten dituen etenena, berriz, 3040 ms-koa. Talde prosodikoen mugetan beheranzko tonoa egiteko joera aurkezten du L eta HL dira gehien agertzen direnak, sei eta bi kasu hurrenez hurren, goranzko tonu bi soilik agertzen dira mugetan. Tonoekin batera doazen eten motak direla eta, isilunedunak eta bokal-txertaketarik gabeak (%) dira guztiak, horietatik bokal txertaketa edo luzapena agertzen du goranzko tonoarekin batera. Azkenik, etenak ez daude banaezinak diren egituretan eginak. Honen guztiaren berri 3. taulak ematen du:

“urte aʃko paʃatu bearko ðira beʃte pilodari batek (HL%) (eh) aɪmar olaiʃolari pilota onenaren titulua kentʃeko (HL%) (eh) ʃaʃpi final xokatu ðitu (L%) (eh) ʃaʃpi tʃapel xaŋʃi (v) (H%) (eh) eta kajolan etorkiʃuneko belauñaldiek (H%) (eh) biðeoan miretʃiko ðuten aβileʃia lortu (L%) (eh) ori ez du lortu βakarig xokatuʃ (L%) (eh) auɾkari ʋoʋoreɪ auɾe ʋinez βaiʃik (L%) (eh) oriʃe faɫta ʃiʃaʃjon<sup>j</sup> tʃapeɫdun aŋdiari (L%) (eh) goyoan artʃeko moðuoko final bat (L%)”.

3. taula: Urdiaingo parte-hartzailearen transkripzio fonetikoa.

#### D) Senpere

Senpereko ikasleak 16 talde prosodiko egiten ditu, zortzi silabakoak batez beste, eta euron denbora 19757 ms-koa da, guztira 28596 ms-ko irakurketa da, taldeen artean egiten dituen etenek 5765s-ko iraupena dute eta lapsusarena 472 ms-koa da, auɾkari hitzaren errepikapena egiten du. Talde prosodikoen mugetan goranzko tonoa (H) egiten

du gehienbat (hamar kasutan), beheranzkoa ere sei kasutan agertzen da (L= 5 eta HL= 1). Mugako tonuen ondoan egiten dituen etenik gehienak isilunedunak eta bokala txertatuta egiten ditu (zortzi), bokala txertatu gabe egiten dituen eten isilunedunak ere asko dira (sei); isilunerik ez dituen eten bi egiten ditu, bata bokal txertaketarekin (v\$) eta bestea hori barik (\$). Ez du etenik egiten banaezinak diren egituretan. Informazio honen berri 4. taulako transkripzioan ematen da.

“urte aško pasatu bear̃ko dira (v) (H%) (eh) beste pilotari batek (v) (H%) (eh) aimarolaĩsolari (v) (H%) (eh) pilotari onenaren titulua kentzeko (v) (HL%) (eh) saşpi final<sup>j</sup> jokatu ditu (v) (H%) (eh) saşpi tşapel<sup>j</sup> jantzi (v) (H%) (eh) eta kaşolan (v) (H%) (eh) etorkişuneko belaunaldiek (H%) (eh) bideoan miretşiko duten (v) (H%) (eh) aβileşia lortu (L%) (eh) ori ez du lortu bakarrik (H\$) şokatuş (L%) (eh) (lap) (eh) aũrkari govoreĩ (v) (H\$) aũreginez baişik (L%) (eh) orişe falta şitşayjon<sup>j</sup> tşapelđun andia.ri (L%) (eh) gogoan artşeko moduko final bat (L%)”.

4. taula: Senpereko parte-hartzailearen transkripzio fonetikoa.

### 2.1. Part-hartzaileen ulermen probak

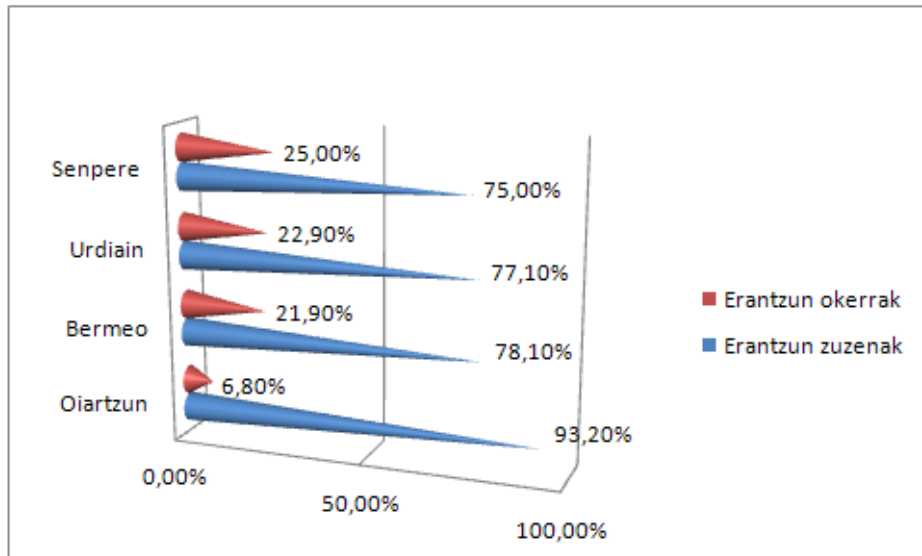
Corpusean zehaztutako testuaren inguruko ulermen-galdera bi (5. taula) erantzun zituzten Bilboko Irakasleen Unibertsitate Eskolako 151 ikaslek, horietatik 44k (% 29,1) Oiartzungo irakurleak irakurritakoa entzun zuten, 40k (% 26,5) Senperekoaren testua, 35ek (% 23,2) Urdiaingoak irakurritakoa eta 32k (% 21,1) Bermeokoaren testua entzun zuten.

- 1- Zertarako pasatu behariko dira urte asko?
  - a) Aimar Olaizolak pilotari onenari txapela kentzeko.
  - b) Aimar Olaizolaren zazpi finalak bideoan miresteko.
  - c) Zazpi txapel jantzi dituen pilotari onenari titulua kentzeko.
  - d) Jokatutako zazpi finalak bideoan miresteko.
- 2- Nola lortzen da pilotari onena izatea?
  - a) Azken urteetan egin diren zazpi finaletan bakarrik jokatu.
  - b) Hainbat final eta txapel jantzi ondoren, ez ahazteko moduko finala jokatu.
  - c) Etorkizuneko belaunaldietako jokalaria gogor hartuta.
  - d) Bideoen bidez mirestuko den abilezia lortuta.

5. taula: Parte hartzaileek erantzundako ulermen galderak.

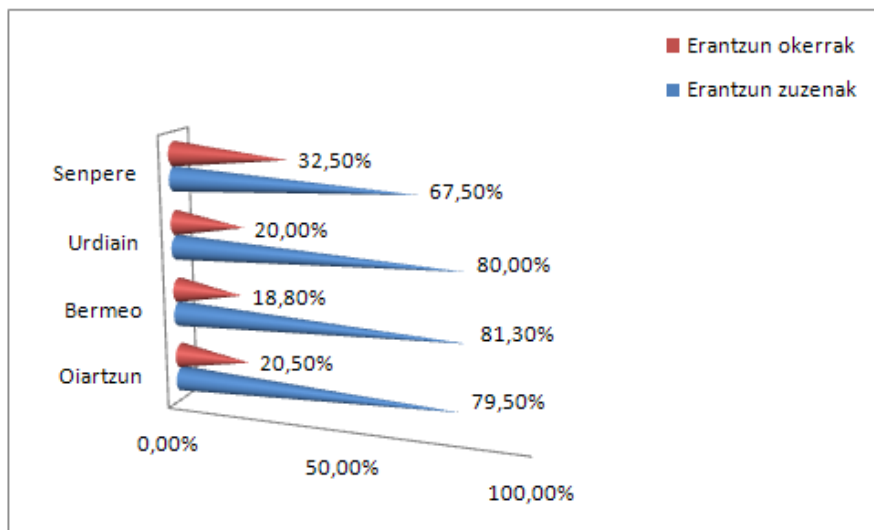
Ikasle gehienen ama hizkuntza gaztelera zen (% 65,6) eta gainerakoena euskara (% 34,4). Ulermen proba egin zuten gehienak Bizkaikoak ziren (% 93,4), sei ikasle gipuzkoakoak eta beste bi Arabakoak ziren, Iparraldeko eta Nafarroako ikaslerik ez zegoen parte-hartzaileen artean. Galderak orokorrean ondo erantzun zituzten ikasleek lehena % 81,5ek zuten erantzun zuen eta bigarrean % 76,8k. Erantzun zuzenei eta okerre dagokienez, lehen galdera ondoen erantzun dutenak Oiartzungo testua entzun dutenak dira (%93,2 erantzun zuten) eta Senpereko testua entzundakoek egin dituzte oker gehien lehen galdera erantzuterakoan (% 25 erantzun okerra). Urdian (% 77,1 zuten eta % 22,9 oker) eta Bermeoko (% 78,1 zuten eta % 21,9 oker) testuko ulermen galderak bete dituztenek antzera egin dute (1. grafikoan ikus daitezke ehuneko hauek).





1. grafikoa: Lehen galderaren erantzun motak testuaren arabera.

Bigarren erantzunari dagokionez, erantzun zuzen gehien jaso dituen testua Bermeokoa izan da, baina ez dago alde handirik erantzun zuzenen kopuruan Oiartzun (% 79,5), Bermeo (% 81,3) eta Urdiainen (% 80,0) artean, Senpereko (% 32,5) testuak izan ditu erantzun oker gehien, hau guztia 2. grafikoa ikus daiteke:



2. grafikoa: Bigarren galderari dagozkion erantzun zuzenak eta okerrak.

Estimuluen eta erantzun zuzenen/okerren emaitzen ez dagoela alde handirik ikus daiteke, nahiz eta rango promedioaren balioan ikus daitekeen bezala, Oiartzungo estimuluak izan dituen galdera asmatu gehien, hau ondorengo 6. taulan ikus daiteke:

	Euskalkia	N	Rango promedio
Galdera asmatuak	Oiartzun	44	82.39
	Bermeo	32	76.50
	Urdiain	35	76.03
	Senpere	40	68.55
	Total	151	

6. taula: Estimulu bakoitzean asmatutako galderak.

Estimuluaren arabera asmatutako galderetan ezberdintasun esanguratsurik dagoen ikusteko, Kurskal Wallis proba aplikatu da, eta ikusi da ez dagoela, izan ere,  $\chi^2(3, n = 2,918), p = 0,404$  da.

### 3. Ondorioak

Lan honetarako erabili diren estimuluaren ezaugarriak oso antzekoak izan dira bariazio linguistikoa kenduta, batez ere ahoskeran azaleratzen dena, Bermeok eta Urdiainek hamar talde prosodiko izan dituzte eta Oiartzunek eta Senpererek hamazazpi eta hamasei hurrenez hurren. Emaitzetan Bermeoko eta Urdiaingo estimuluek antzeko emaitzak izan dituzte, iraupena (talde prosodikoena, etenena, eta irakurketa osoarena) nahiz talde prosodikoen ezaugarriak ere oso antzekoak izan dira kopuruan ez ezik, muga tonuetan (L eta HL) nahiz eten motetan (% gehienak edo guztiak Urdiaingo kasuan) ere antzekotasun handiak egon dira, esan beharra dago Bermeoko grabazioan lapsus bat entzuten dela. Beste estimulu biek, Oiartzungoak eta Senperekoak emaitzarik onenak eta txarrenak lortu dituzte, hurrenez hurren, talde prosodikoen ezaugarriak oso antzekoak dira, luzera, iraupena, muga tonuak (H gehien) eta etenak (% gehien) kontuan izanik, Senperekoak lapsus bat egiten du, hitz bat errepikatzen baitu, baina ezberdintasun nagusia irakurlearen bariazio diatopikoa da gehienbat ahoskeran azaleratzen dena eta ez ezaugarri suprasegmentaletan, estatistikaren aldetik ezberdintasun esanguratsurik azaleratu ez arren, ehunekoetan erantzun zuzenen kopuruan badaude aldeak.

Bukatzeko, badirudi irakurleak duen bariazio diatopikoak ez diola era esanguratsuan txarrerako eragiten entzulearen ulermenari, ikusi beharko litzateke pertzepzio mailan nola hartzen duen entzuleak. Gainera, bariazio diatopikoa irakurlearen gaitasun prosodikoaren ezaugarri bat baino ez da, eta badaude beste zenbait osagai manipula daitezkeenak ulermenari nola eragiten dioten ikusteko edota irakurleak entzutean zer motatako informazio paralinguistikoa edota ez linguistikoa perzibitzen duen ikertzeko, hain zuzen ere, alde horiek izango dira etorkizunera begira landu beharrekoak.

### 4. Bibliografia

- Benjamin, R. G. y Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly*, 45, 388-404.
- Cova, Y. (2004). La práctica de la lectura en voz alta en el hogar y en la escuela a favor de niños y niñas. *Sapiens*, 002, 53-66. Hemendik berreskuratua <http://redalyc.uaemex.mx/pdf/410/41050205.pdf>

- Etxebarria, A., Garay, U., Gaminde, I., Romero, A. eta Larrea, K. (2011). Prosodiaren ebaluazioa irakurketa ozenean. En Aurrekoetxea, G. y Gaminde, I. (eds.) *Prosodiaz eta Hezkuntzaz I. Jardunaldiak/I. Jornadas sobre Prosodia y Educación*.
- Fujisaki, H. (2004). Information, Prosody, and Modeling. *Proceedings of Speech Prosody*. Japón: Nara.
- Gaminde, I., Aurrekoetxea, G., Etxebarria, A., Garay U., & Romero, A. (2015). *Ahoskera lantzeko argibideak eta jarduerak. Laguntzarako materiala: teoria eta praktika*. Leioa: UPV/EHU.
- Gaminde, I., Etxebarria, A., Garay, U. eta Romero, A. (2011). Bokatihoa eta gaitasun prosodikoa: lehen urratsak. *Uztaro*, 79, 5-23.
- Gaminde, I., Etxebarria, A., Garay, U. eta Romero, A. (2012). Komunikazio gaitasunetik kultura arteko komunikazio gaitasunera: Baliabide didaktikoak. En *Hizkuntzaz Jabetzen* (pp. 209-251). Bilbao: Mendebalde Kultura Alkartea.
- Garay, U. eta Etxebarria, A. (2010). Irakasleen komunikazio-gaitasuna: hausnarketarako gaia. *Uztaro*, 73, 39-48.
- Garay, U. eta Etxebarria, A. (2010). Irakasleen komunikazio-gaitasuna: hausnarketarako gaia. *Uztaro*, 73, 39-48.
- Grau i Tarruell, M. eta Vilà i Santasusana, M. (2009). La competencia prosódica y la comunicación no verbal. En Vilà i Santasusana, M. (coord.), *El discurso oral formal. Contenidos de aprendizaje y secuencias didácticas* (pp. 89-100). Barcelona: Graó.
- HABE (itzulpena) (2005). *Hizkuntzen ikaskuntza, irakaskuntza eta ebaluaziorako Europako Erreferentzi Marko Bateratua*. Donostia: HABE.
- Klauda, S.L. eta Guthrie, J.T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, 100 (2), 310-321.
- Llisterri, J. (2016). *Los elementos suprasegmentales*. Hemendik berreskuratua [http://liceu.uab.es/~joaquim/phonetics/fon\\_prosod/suprasegmentales.htm](http://liceu.uab.es/~joaquim/phonetics/fon_prosod/suprasegmentales.htm)
- Miller, J. eta Schwanenflugel, P.J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, 98 (4), 839-853. doi:10.1037/0022-0663.98.4.839
- Miller, J. eta Schwanenflugel, P.J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, 43(4), 336-354. doi:10.1598/RRQ.43.4.2
- Mozziconacci, S. (1998). *Speech Variability and Emotion: Production and Perception*. Eindhoven: Proefschrift.
- Rasinski, T.V., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction*, 48 (4), 350-361.
- Rodero, E. (2013). The Perception of a Broadcasting Voice. *US-China Education Review*, 4, 225-230.
- Saá, N. (2001). Lenguaje y hemisferio derecho. *Cuadernos de neurología*, XXV. Hemendik berreskuratua <http://escuela.med.puc.cl/publ/cuadernos/2001/06.html>
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M.R., Wisenbaker, J. M., eta Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96 (1), 119-129.
- Veenendaal, N. J., Groen M. A. eta Verhoeven, L. (2014). The role of speech prosody and text Reading prosody in children's reading comprehension. *British Journal of Educational Psychology*, 84, 521-536.

## DOS DÉCADAS DE DIALECTOMETRÍA ENTONATIVA

**Paolo Roseano**  
**Universitat de Barcelona**  
[paolo.roseano@ub.edu](mailto:paolo.roseano@ub.edu)

### Abstract

This article summarizes the most important developments of intonational dialectometry during the last two decades (1998-2016). After a smooth start at the very end of the 20th century, intonational dialectometry has become more and more popular during the last ten years. During the last decade researchers interested in Romance languages have developed a few different computer programs and packages that can be used to carry out dialectometrical analysis of intonational data. The most recent application of this kind, called ProDis, was developed at the Phonetics Laboratory of the University of Barcelona.

**Keywords:** dialectometry, intonation, prosody, ProDis

### Introducción

La preocupación por la necesidad de medir de manera objetiva las distancias entre variedades lingüísticas surgió en los años 70 del siglo XX y ha llevado al nacimiento de la disciplina conocida como dialectometría (v. entre otros, Séguy, 1971; Goebel, 1982), que consiste en aplicar varios tipos de técnicas de análisis estadístico a bases de datos muy extensas (normalmente atlas lingüísticos). Las bases de datos sobre las cuales se suelen aplicar las técnicas dialectométricas incluyen datos fonético-fonológicos segmentales, morfológicos, léxicos o sintácticos, pero no incluyen información acerca de la entonación. De hecho, sólo en la última década se han creado las bases de datos entonativas – los atlas prosódicos como AMPER (Contini, 1992; Contini, Rouillet, Romano y Lai, 2003) o IARI (Prieto, Roseano y Borràs-Comes, 2010-2014) – que constituyen el material imprescindible para el análisis geolingüístico de la entonación. Este retraso en la creación de bases de datos prosódicos ha comportado que hoy en día sean aún muy escasos los estudios que aplican métodos estadísticos para valorar las diferencias y semejanzas entre la entonación de diferentes dialectos o lenguas (Hermes 1998a,b; Romano, 1999; Moutinho, Coimbra, Rilliard y Romano, 2011; Rilliard y Lai, 2008; Romano y Miotti, 2008; Fernández Planas, Roseano, Martínez Celdrán y Romera, 2011; Sullivan, 2011; Romano, Rilliard, Lai y Contini, 2011; Roseano, 2012; Fernández Rei, Moutinho y Coimbra, 2013; Prieto y Cabré, 2013). Este artículo, presenta una breve historia de la evolución de la que podríamos definir la dialectometría entonativa. El objetivo es, por lo tanto, presentar el estado de la cuestión del desarrollo de técnicas e instrumentos para el análisis dialectométrico de datos prosódicos.

### 1. La dialectometría tradicional, sus datos y técnicas de análisis

Los atlas lingüísticos representan una fuente importante de informaciones sobre los rasgos fonéticos, morfológicos, sintácticos y, sobre todo, léxicos de una lengua y de sus variedades. Además, constituyen un punto de partida para la detección de similitudes entre dialectos, lo que permite clasificarlos y agruparlos. Este proceso de clasificación se ha desarrollado, en los estudios tradicionales de dialectología, de forma cualitativa, es decir que cada investigador selecciona una serie de rasgos relevantes, les

otorga menor o mayor importancia según criterios que el investigador mismo (o la tradición) considera importantes, y los utiliza para establecer las agrupaciones dialectales.

En los años 70 del siglo XX surgió una nueva tendencia metodológica en el ámbito de los estudios dialectológicos, que se llamó dialectología cuantitativa o también dialectometría, término propuesto por Séguy (1971). Esta disciplina, según la definición de Goebel (1981: 349), representa una alianza entre la geolingüística y la taxonomía numérica como rama de la matemática. La dialectometría ofrece algunas ventajas claras en comparación con la dialectología tradicional. En primer lugar, permite gestionar una cantidad de datos muy grande con un esfuerzo relativamente limitado. Justamente el hecho de trabajar con bases de datos extensas hace posible que las conclusiones a las que llega el análisis dialectométrico tengan un valor estadístico que va más allá de la intuición de un investigador. En segundo lugar, las técnicas de análisis dialectométrico están menos expuestas al riesgo de apriorismos, ya que no otorgan una importancia menor o mayor a las diferentes variables que conforman la base de datos, sino que cada una de ellas tiene el mismo peso que las otras. Aun así, hay que recordar que, justamente por este motivo, estos métodos no han estado exentos de críticas desde la dialectología tradicional, ya que no tienen en cuenta que ciertas diferencias lingüísticas son cualitativamente más relevantes que otros o también porque puede ser discutible la medida de similitud a partir de la cual se establecen las distancias, tal como apunta Clua (1999). Una última ventaja técnica de las técnicas dialectométricas en comparación con la dialectología tradicional consiste en la forma de presentación de las correlaciones estadísticas entre los datos. De hecho, para evitar que la interpretación de los resultados del tratamiento estadístico de los datos lingüísticos (mediante análisis como el *cluster analysis* y el *multidimensional scaling*) se convirtiera en una barrera infranqueable para muchos lingüistas, a lo largo de las últimas cuatro décadas se han ensayado diversas formas de representación gráfica, entre ellas los dendrogramas y los mapas MDS, que actualmente se utilizan de manera habitual en este tipo de trabajos.

Los análisis dialectométricos tradicionales se realizan exclusivamente a partir de bases de datos de tipo léxico, fonético segmental y morfológico (entre los muchos ejemplos, podemos citar para el catalán Clua, 2004; Perea, 2010; Valls, Nerbonne, Prokic, Wieling, Clua y Lloret, 2012; Goebel, 2013, entre otros). Solo a partir de los primeros años del siglo XXI se empezó a disponer de las bases de datos suprasegmentales que, a partir de la primera aproximación de Romano (1999), se analizaron mediante técnicas dialectométricas. Desde el punto de vista metodológico, como se verá más adelante, hay que destacar que las bases de datos suprasegmentales difieren radicalmente de las bases de datos dialectales tradicionales, ya que las segundas son alfabéticas (p.e. las transcripciones de la pronunciación de una palabra), mientras que las primeras son numéricas (pe los valores numéricos, en Hz o st, del F0 en diferentes puntos de un contorno entonativo). Esta diferencia en el tipo de datos, aparentemente banal, es en realidad crucial para entender por qué razón en el marco de la dialectología entonativa con datos numéricos no se ha podido utilizar ninguna de las herramientas dialectométricas más conocidas, es decir *Visual DialectoMetry* (Haimerl, 2006), *Gabmap* (Nerbonne, Colen, Gooskens, Kleiweg y Leinonen, 2011) y *DiaTech* (Aurrekoetxea Fernandez-Aguirre, Rubio, Ruiz y Sánchez, 2013).

En definitiva, el proceso de la dialectometría tradicional se puede representar esquemáticamente en las tres fases que aparecen en la Figura 1: la de recogida de datos

alfabéticos, la de creación de una matriz de distancias numéricas a partir de los datos alfabéticos, y la de transformación de la matriz en cuestión en una (o más de una) representación gráfica de más fácil interpretación.

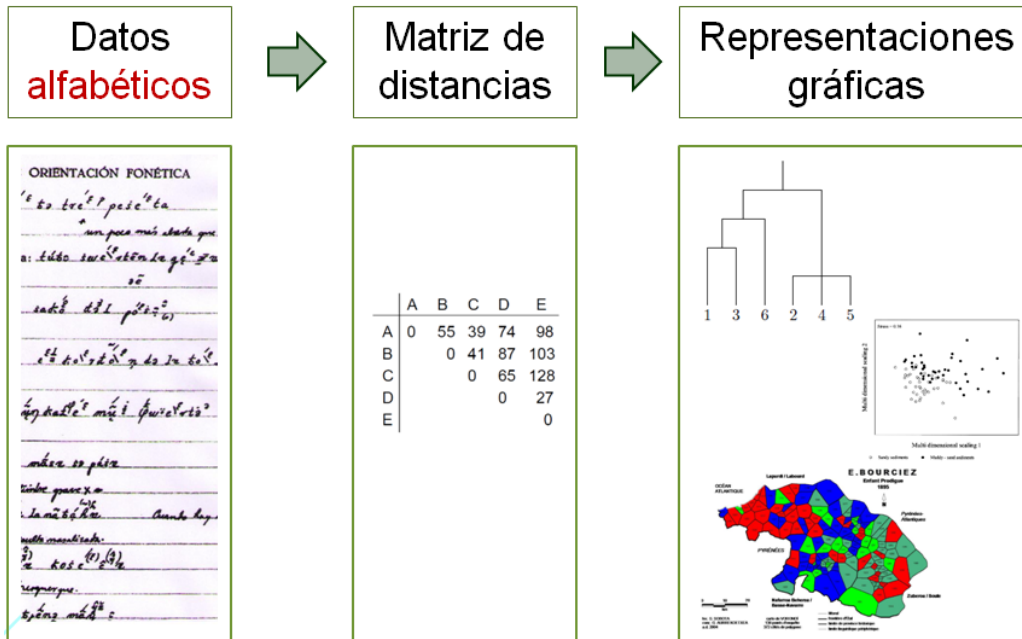


Figura 1. Fases de análisis en dialectometría tradicional

## 2. La dialectometría prosódica: objetivo general

Los estudios de prosodia (y su variación dialectal) se pusieron de moda solo en las últimas décadas, y las grandes bases de datos de ese tipo se han ido recogiendo solo en los últimos años. Cuando los entonólogos quisieron aplicar técnicas dialectométricas a sus bases de datos (que contenían sobre todo datos numéricos), se encontraron con que no existían instrumentos adecuados, ya que –tal y como se ha adelantado en la Sección anterior– los programas de análisis dialectométrico que existían estaban pensados para analizar solo (o casi solo) datos alfabéticos. Por esta razón, uno de los objetivos metodológicos más urgentes para los investigadores que se ocupaban de dialectología entonativa era el de crear un instrumento dialectométrico adecuado, que fuera capaz de reproducir las tres etapas de la dialectología tradicional, pero teniendo en cuenta que los datos de base son numéricos. En otras palabras, el proceso que se quería implementar en el marco de la dialectometría entonativa se puede representar esquemáticamente en las tres fases que aparecen en la Figura 2 y que son muy parecidas a las que se han contemplado en la Figura 1: la de recogida de datos numéricos (básicamente, de valores de F0 en distintos puntos de una curva entonativa), la de creación de una matriz de distancias numéricas a partir de esos datos, y la de transformación de la matriz en una representación gráfica transparente.

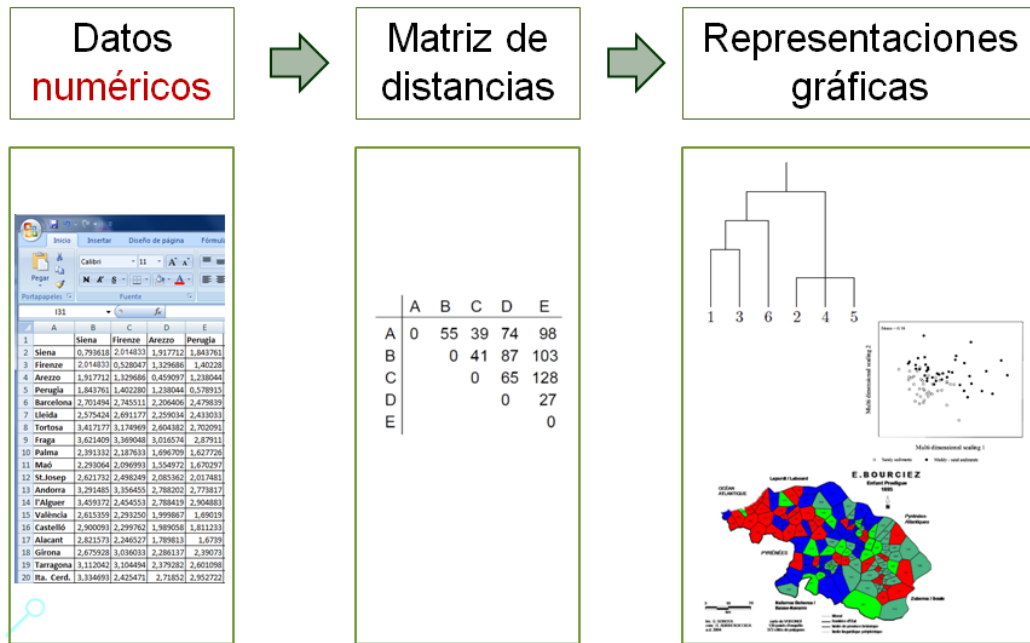


Figura 2. Fases de análisis en dialectometría entonativa

### 3. La dialectometría entonativa: un esquema de su historia

La dialectometría entonativa es, de hecho, una disciplina con una historia relativamente breve, que empieza a finales de los años noventa y que se puede representar con un esquema como el de la Figura 3, que se irá comentando y desarrollando a lo largo de los próximos apartados.

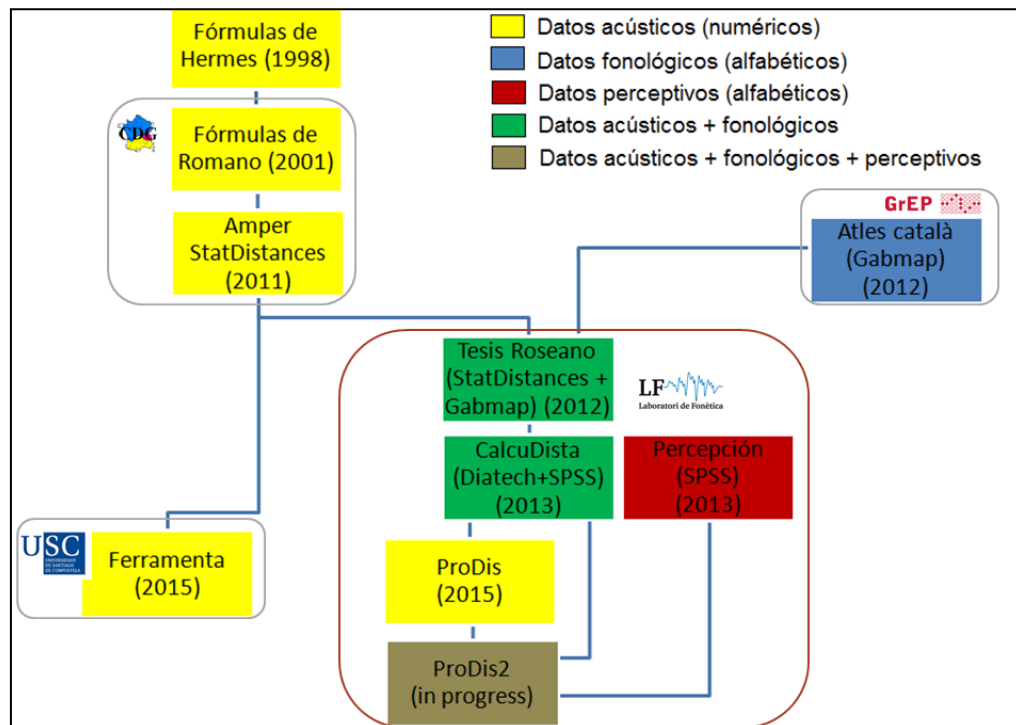


Figura 3. Fases históricas de la dialectometría entonativa. Los colores indican los diferentes tipos de datos que se someten a análisis dialectométrico y los recuadros identifican las aportaciones de diferentes grupos de estudio.

#### 4. La dialectometría entonativa: sus bases de datos

Para entender el camino que ha seguido el desarrollo de los programas dialectométricos entonativos es necesario tener en cuenta las bases de datos entonativas de las que se dispone hoy en día (en concreto para las lenguas románicas, que son las únicas a las que se han aplicado las técnicas de dialectometría entonativa). Las bases de datos en cuestión son, fundamentalmente, dos, o, si queremos, de dos familias: la familia AMPER (Contini, 1992) y la familia IARI (Prieto et ál., 2010-2014; Frota y Prieto, 2015). Ambos proyectos no solo son similares en sus objetivos, sino también en su organización y en sus resultados. Sin embargo, muestran diferencias en los datos que se analizan. En lo que se refiere a la organización, ambos proyectos tienen un comité central de dirección y grupos de investigación para cada idioma (y, a veces, en el caso de lenguas geográficamente extensas, grupos de investigación regionales). Los datos se ponen a disposición del público en línea en el sitio web central de cada proyecto, con muestras de todos los idiomas, y en las páginas web de los grupos de investigación que recopilan datos de lenguajes específicos (como Martínez Celdrán y Fernández Planas, 2003-2016a y Prieto y Cabré, 2007-2012 catalán; Prieto y Sichel-Bazin 2014 por el occitano; Prieto y Roseano, 2009-2013 y Martínez Celdrán y Fernández Planas, 2003-2016b de español; Frota y Cruz, 2012-2015 y de Castro Moutinho, 2008-2013 para el portugués; Fernández Rei, 2016 para el gallego; Roseano y Fernández Planas, 2009-2013 para friulano, Turculeţ, 2010-2013 para el rumano, entre otros).

La diferencia quizás más importante entre AMPER y IARI es el método de análisis que, a su vez, determina el tipo de información que está contenida en las bases de datos mencionadas anteriormente. Mientras que en la familia IARI el método predominante de análisis es auditivo y fonológico, en la familia AMPER el enfoque es principalmente fonético y numérico. Eso tiene consecuencias importantes sobre el tipo de datos que constituyen la base de datos final, la que se puede utilizar para análisis de tipo dialectométrico. Mientras que las bases de datos en línea IARI suelen ofrecer el contorno F0 y el etiquetaje fonológico de un solo ejemplo de un gran número de modalidades oracionales por punto de encuesta (desde 18 a 47, dependiendo del idioma), las bases de datos AMPER ofrecen los valores numéricos (en Hz st, ms y dB) de la prosodia de varias grabaciones (de 378 a 567, dependiendo de la lengua) de dos tipos de oraciones (declarativas neutras e interrogativas totales neutras). La Figura 4 contiene un ejemplo de datos entonativos que se ofrecen en los atlas en línea en español de la familia IARI (Prieto y Roseano, 2009-2013).



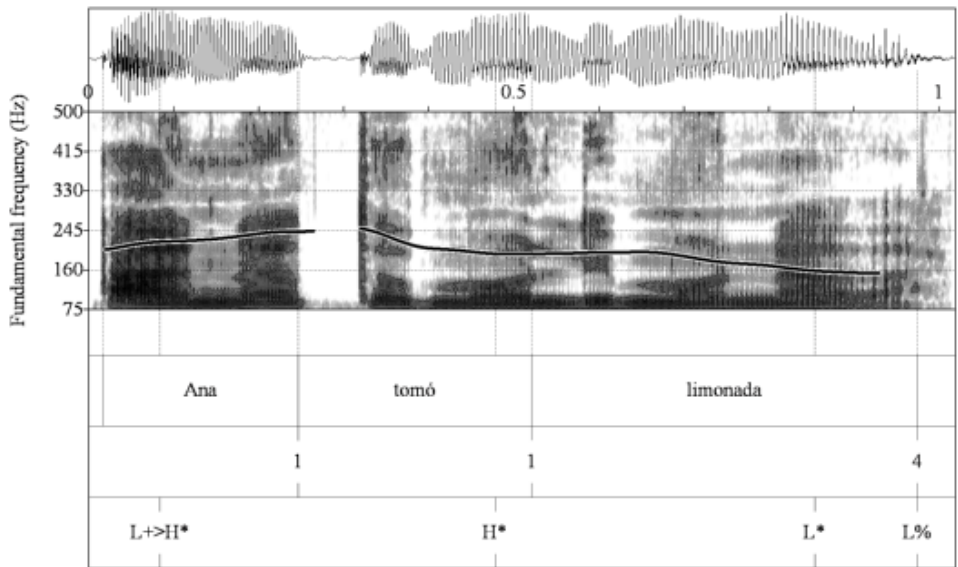


Figura 4. Ejemplo de datos en la familia IARI

Por otra parte, las bases de datos AMPER contienen valores numéricos seleccionados que resumen la prosodia de cada oración; básicamente, para cada sílaba de la frase, la base de datos contiene tres valores F0 (en Hz y st), tres valores de intensidad (en dB) y la duración (en ms). Basándose en estos valores numéricos, un transcriptor recientemente desarrollado *ad hoc* llamado *AmperEti* (Roseano y Fernández Planas, 2013) puede añadir una transcripción entonativa fonética estrecha del contorno de cada frase. Esto significa que, desde hace muy pocos años, la base de datos AMPER puede contener, además de sus valores fonéticos prosódicos numéricos habituales, también un conjunto de transcripciones fonéticas (es decir, de datos alfabéticos) de los contornos entonativos. La Figura 5 resume el proceso de análisis en el marco AMPER, en el que se puede destacar la presencia de una base de datos numérica.

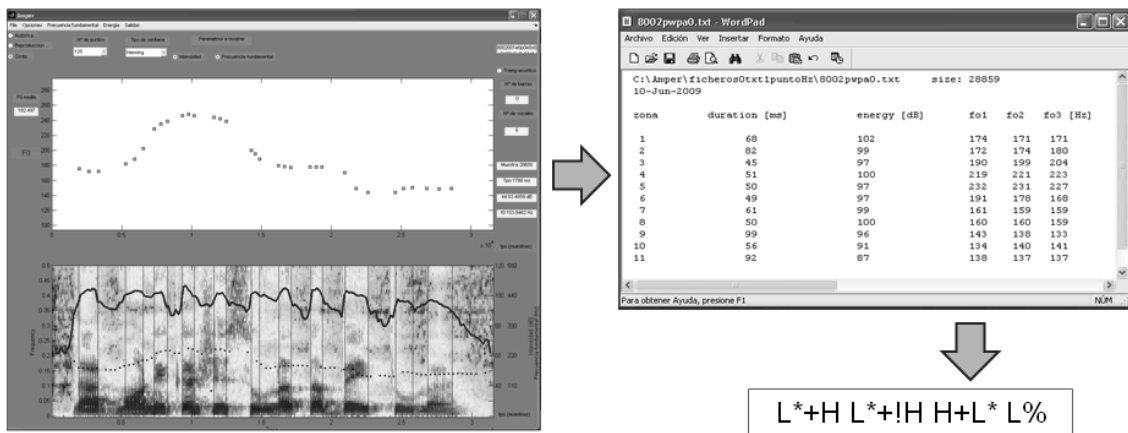


Figura 5. Ejemplo de datos en la familia AMPER.

En definitiva, mientras que la base de datos IARI contiene datos alfabéticos (es decir, las etiquetas entonativas) que se pueden intentar analizar con los instrumentos de la dialectometría tradicional, la base de datos AMPER contiene típicamente datos

numéricos para cuyo análisis los instrumentos como *VisualDialectoMetry*, *Gabmap* o *DiaTech* no resultan adecuados.

### 5. La dialectometría entonativa numérica: su origen

Los primeros intentos de calcular la distancia prosódica entre dos curvas entonativas se deben a Hermes (1998a,b), que propuso las primeras fórmulas que tenían esa finalidad. Su propuesta, sin embargo, no llegó a implementarse inmediatamente en el sector dialectológico. Se tuvo que esperar algunos años para que eso sucediera. A principios del siglo XXI, de hecho, es cuando se empiezan a aplicar las fórmulas de Hermes a bases de datos de variación diatópica de la entonación. En esta primera fase el centro impulsor de la dialectometría entonativa es el grupo de trabajo que tiene como punto de referencia en Centro de Dialectología de Grenoble. En concreto, es con los trabajos de investigación relacionados con la tesis doctoral de Romano (1999) que se empieza a vislumbrar la posibilidad alentadora de aplicar técnicas dialectométricas a la base de datos prosódicos que se estaba empezando a construir en aquellos años: AMPER (Contini, 1992; Contini et ál., 2003).

Los dos programas que se utilizan en esta primera fase utilizan, básicamente, las fórmulas de Hermes (1998a,b) o adaptaciones de las mismas. El primero de ellos, conocido como *Amper StatDistances* fue creado por Albert Rillard y Jean-Pierre Lai en el marco del proyecto AMPER (Rillard y Lai, 2008; Moutinho et ál., 2011), específicamente con el objetivo de dialectometrizarse los datos numéricos de la F0 de las frases que se recogen de acuerdo con la metodología del proyecto. Este aplicativo, a partir de los datos prosódicos que corresponden a tres valores de F0 (en Hz) en cada vocal de la frase (inicial, central y final), más un valor de intensidad en cada una de ellas, ofrece una representación gráfica de las proximidades entre dialectos y de las agrupaciones de los mismos a partir de un dendrograma y un mapa MDS. Las figuras 6 y 7 contienen un ejemplo de dendrograma y mapa MDS, ambos realizados con *Amper StatDistances* y adaptados de Roseano (2012).

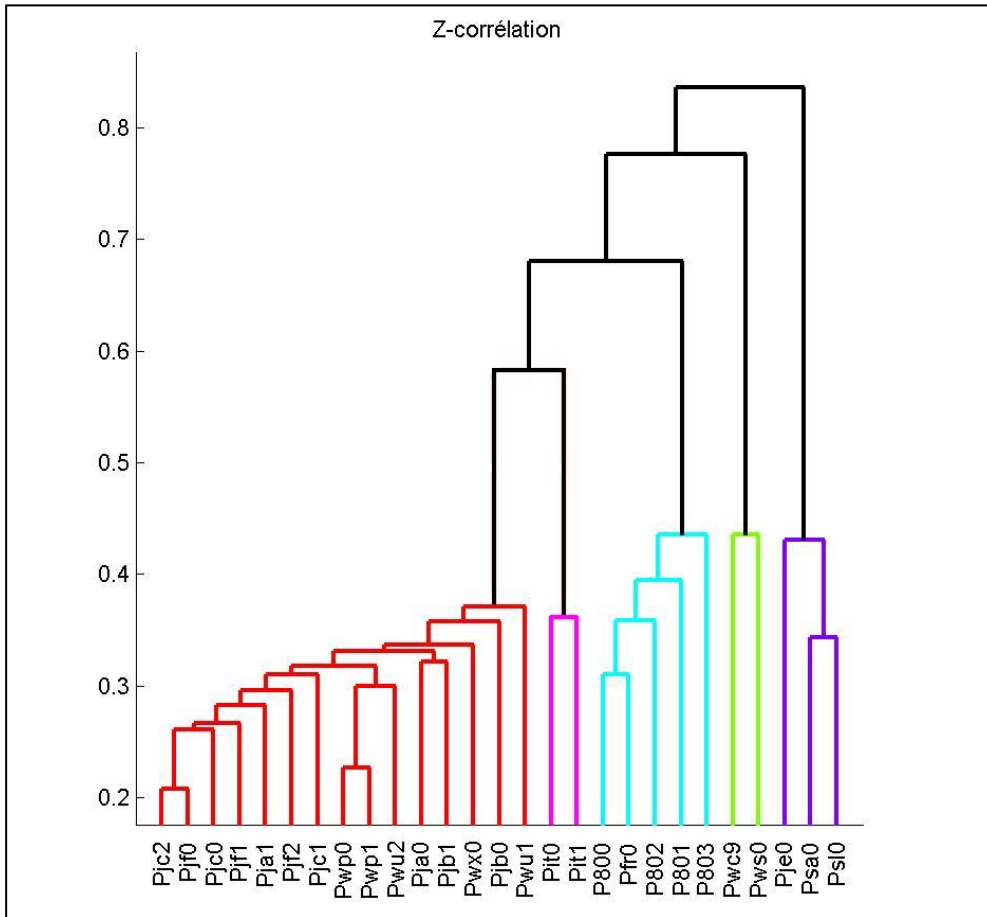


Figura 6. Dendrograma que representa las distancias entonativas entre distintas variedades románicas realizado con *Amper StatDistances*

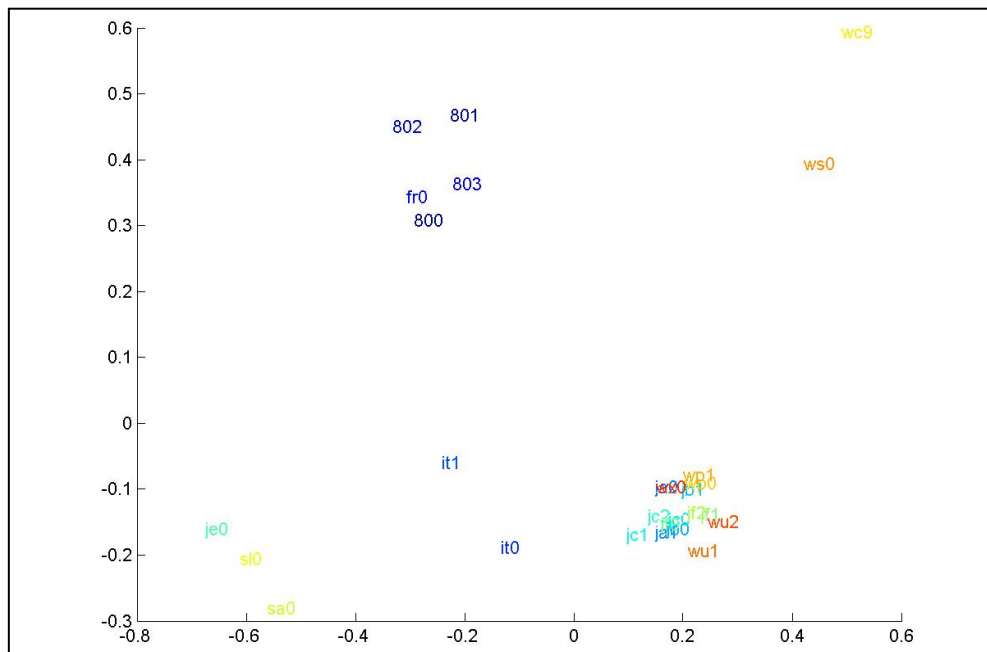


Figura 7. Mapa MDS que representa las distancias entonativas entre distintas variedades románicas realizado con *Amper StatDistances*

Además, mediante unos gráficos de cajas, *Amper StatDistance* ofrece también la posibilidad de valorar la variabilidad observada entre diferentes repeticiones de un

mismo informante y entre las repeticiones de los diferentes informantes de un punto de encuesta. La Figura 8 contiene un ejemplo de gráfico de cajas realizado con *Amper StatDistances* adaptado de Roseano (2012).

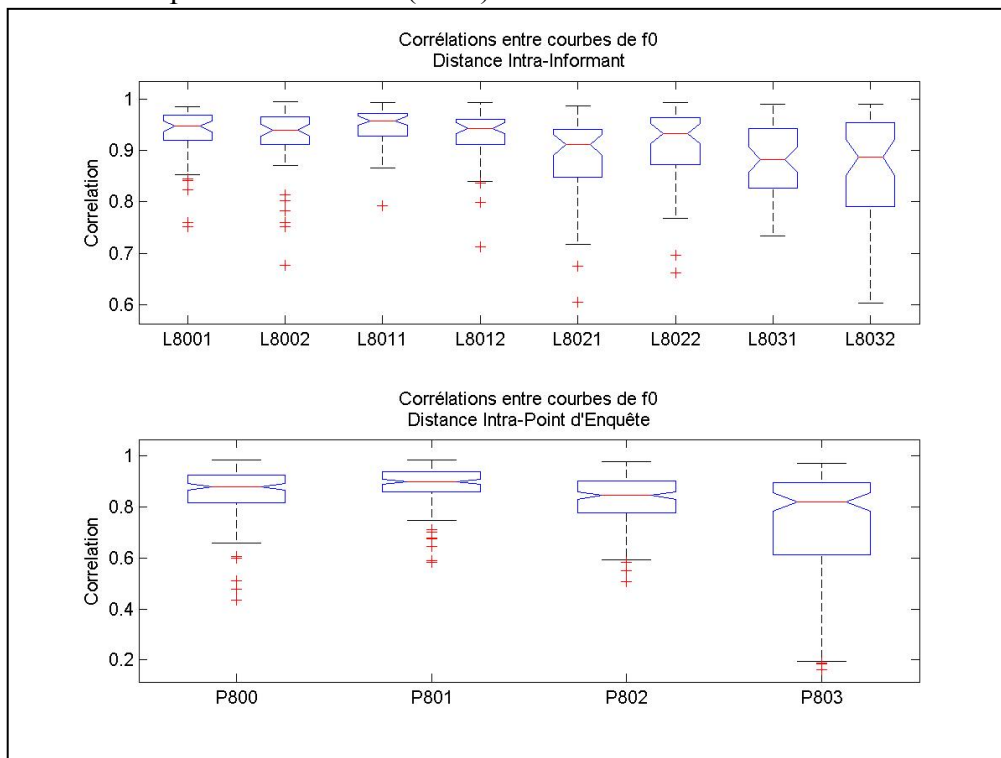


Figura 8. Mapa MDS que representa las distancias entonativas entre distintas variedades románicas realizado con *Amper StatDistance*

La limitación fundamental de *Amper StatDistance* consistía en el hecho de que no permitía visualizar la matriz de distancias ni escoger los métodos de conglomeración. Por esa razón en el Laboratorio de Fonética de la Universitat de Barcelona, centro desde el que se lidera la parte catalana y española del mismo proyecto AMPER, se elaboró – pocos años después de la aparición de *Amper StatDistances*– una rutina capaz de subsanar el inconveniente en cuestión.

La rutina elaborada por el laboratorio de fonética (Roseano, Fernández Planas, Elvira-García y Martínez Celdrán, 2015) toma como base los datos el *output* del análisis numérico que se suele llevar a cabo en el marco AMPER y crea, mediante un script de Praat, una base de datos en un formato Excel que se puede analizar con SPSS. Una vez que los datos numéricos en cuestión se han importado en SPSS, es posible aplicar técnicas de *clustering* equivalentes a las que se implementan en la dialectometría tradicional. La ventaja, en comparación con *Amper StatDistances*, consiste en el hecho de que, así haciendo, es posible seleccionar los algoritmos adecuados y ofrecer los indicadores estadísticos (como el *stress* o el RSQ) que en el programa anterior quedaban ocultos. La rutina del Laboratorio de Fonética, de acuerdo con la propuesta clásica de Hermes (1998a,b), utiliza como índice de la distancia entonativa entre dos frases, que podemos llamar  $x$  e  $y$ , la media cuadrática de la diferencia entre los valores de F0 de la frase  $x$  y de la frase  $y$  en cada uno de los puntos de medición. Para los dos conjuntos  $x$  e  $y$  de valores de F0  $\{f0x1, f0x2, \dots, f0xN\}$  y  $\{f0y1, f0y2, \dots, f0yN\}$ , donde  $N$  es el número de puntos de medición de F0 en cada una de las dos frases, mientras que  $f0xi$  y  $f0yi$  son los valores de F0 en  $st$  en cada uno esos puntos, la media cuadrática en cuestión viene dada por la siguiente fórmula:

$$RMS = \sqrt{\frac{\sum_{i=1}^N (f0x_i - f0y_i)^2}{N}}$$

Esta fórmula proporciona la distancia entre dos frases. Para determinar la distancia general entre todas las frases de dos puntos de encuesta, la rutina utiliza la mediana de las RMS calculadas por cada pareja de frases x e y. A partir de las medianas de las distancias entre cada par de puntos de encuesta se construye una matriz de distancias que aparece en la Figura 9 y que corresponde a la fase 2 de la Figura 2.

	Siena	Firenze	Arezzo	Perugia	Barcelona	Lleida	Tortosa	Fraga	Palma	Maó	St.Josep	Andorra	l'Alguer	València	Castelló	Alacant	Girona	Tarragona	Ita.Cerd.	Sardo
Siena	0.79362																			
Firenze	2.01483	0.52805																		
Arezzo	1.91771	1.32969	0.45910																	
Perugia	1.84376	1.40228	1.23804	0.57892																
Barcelona	2.70149	2.74551	2.20641	2.47984	1.27458															
Lleida	2.57542	2.69118	2.25903	2.43303	2.60293	2.09842														
Tortosa	3.41718	3.17497	2.60438	2.70209	2.56334	2.89857	1.68506													
Fraga	3.62141	3.36905	3.01657	2.87911	2.51865	2.95265	3.04267	1.71903												
Palma	2.39133	2.18763	1.69671	1.62773	2.29823	2.34280	2.16346	2.66916	0.92233											
Maó	2.29306	2.09699	1.55497	1.67030	2.40566	2.26436	2.67430	2.74954	1.59460	1.12967										
St.Josep	2.62173	2.49825	2.08536	2.01748	2.19261	2.22254	3.73926	2.29165	1.63698	1.56400	1.19501									
Andorra	3.29149	3.35646	2.78820	2.77382	1.98106	2.57778	2.56256	2.43157	2.38082	2.47538	2.07799	1.81683								
l'Alguer	3.45937	2.45455	2.78842	2.90488	3.93404	3.53782	3.79562	4.30739	2.95567	2.97357	3.30629	3.99666	1.81205							
València	2.61536	2.29325	1.99987	1.69019	1.81788	2.20153	1.84788	1.94496	1.47872	1.67567	1.45762	1.92488	3.43892	1.43678						
Castelló	2.90009	2.29976	1.98906	1.81123	2.03270	2.10225	2.10483	2.08824	1.43910	1.72195	1.59209	2.13889	3.41749	1.43678	1.17111					
Alacant	2.82157	2.24653	1.78981	1.67390	2.09568	2.25583	2.11356	2.23795	1.44771	1.65968	1.62390	2.15361	3.17047	1.41619	1.40801	1.05075				
Girona	2.67593	3.03603	2.28614	2.39073	2.12698	2.61069	2.56769	2.70127	2.19328	2.22967	1.94563	2.21256	3.78602	1.80547	1.94542	2.05511	1.76715			
Tarragona	3.11204	3.10449	2.37928	2.60110	1.91436	2.54337	2.61550	2.62002	2.48882	2.27095	1.90090	1.99143	3.41376	1.53402	1.88835	1.94739	2.11682	1.16163		
Ita.Cerd.	3.33469	2.42547	2.71852	2.95272	3.76104	3.73926	3.93832	4.36039	3.19864	3.01855	3.38387	4.14699	2.30358	3.40952	3.44873	3.17457	3.82572	3.45911	1.20007	
Sardo	3.14025	2.36387	2.67281	2.66999	3.58960	3.47936	3.62814	4.02838	3.03382	2.62801	3.09136	3.82080	2.33841	3.27006	3.18927	2.96695	3.59627	3.31371	1.86951	1.14617

Figura 9. Matriz de distancias

La matriz de distancias constituye a su vez la base para la fase final del proceso de análisis, la tercera, que se efectúa con SPSS y consiste en un análisis de clúster cuya finalidad es clasificar los puntos de encuesta en grupos a partir de la diferencia entre sus características entonativas, que es el objetivo último de un análisis dialectométrico entonativo (en decir, corresponde a la fase 3 de la Figura 2). La Figura 10 contiene un ejemplo (adaptado de Roseano et ál., 2015) de representación gráfica, en forma de dendrograma, de los resultados del análisis dialectométrico. Cabe destacar que para el gráfico en cuestión se ha podido escoger y explicitar el método de conglomeración (en este caso, el valor más lejano) y la medida (intervalo distancia euclidiana; valores no transformados), cosa *Amper StatDistances* que no permitía.

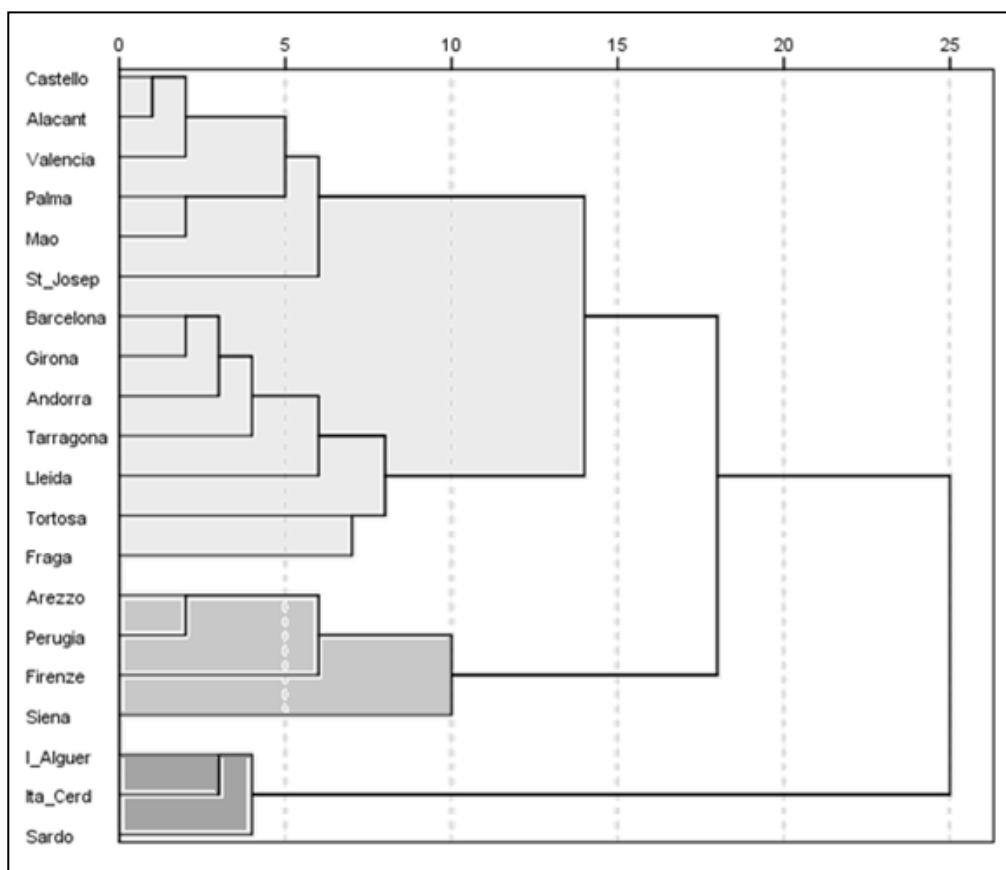


Figura 10. Dendrograma del análisis de clúster de las distancias entonativas entre variedades románicas (adaptado de Roseano et ál., 2015)

A pesar de cumplir con el requisito de explicitar y flexibilizar las tres fases imprescindibles de una dialectometría entonativa que se han ilustrado en la Figura 2, la rutina del Laboratorio de Fonética que se acaba de presentar seguía presentando algunos defectos. El más importante de ellos es, sin duda, su laboriosidad. De hecho, la realización de un análisis dialectométrico con esos instrumentos implica una serie de operaciones informáticas (que en este trabajo no se detallan por razones de espacio) que representan un coste muy alto en términos de tiempo. Por esta razón, tal y como se verá más adelante, en los años siguientes surgirán rutinas más sencillas.

## 6. La dialectometría entonativa alfabética: su origen

En un segundo momento, aproximadamente hacia el año 2012, en el marco del proyecto *Atles interactiu de l'entonació del català* (Prieto y Cabré, 2007-2012) del *Grup d'Estudis de Prosòdia* de la Universitat Pompeu Fabra, se lleva a cabo el primer análisis dialectométrico de datos entonativos a partir de datos alfabéticos (en este caso, transcripciones fonológicas de configuraciones nucleares de diferentes modalidades oracionales). El instrumento utilizado es *Gabmap* (Nerbonne et ál., 2011), un instrumento capaz de procesar datos de tipo alfabético<sup>13</sup>. Los resultados del análisis en cuestión se publicaron en forma de mapa en Prieto y Cabré (2013) y se reproducen en la Figura 11. De ella, cabe destacar una conclusión a la que llegan los autores del trabajo

<sup>13</sup>Debido al hecho de que *Gabmap* no acepta símbolos comúnmente utilizados en la transcripción de la entonación (como \* + - % > ! i), se han tenido que convertir en símbolos alfabéticos convencionales. Por ejemplo: L+H\* → LpHs, L% → Lb, etc.

en cuestión y que parece ser una constante de todos los trabajos de dialectometría entonativa: que los grupos dialectales que se pueden formar a partir de la entonación no coinciden con los que propone la dialectología tradicional. En el caso del catalán, por ejemplo, los grupos dialectales entonativos no siguen la tradicional división entre variedades orientales y occidentales de la lengua, que en el mapa de la Figura 11 se representa con una línea de rayas.

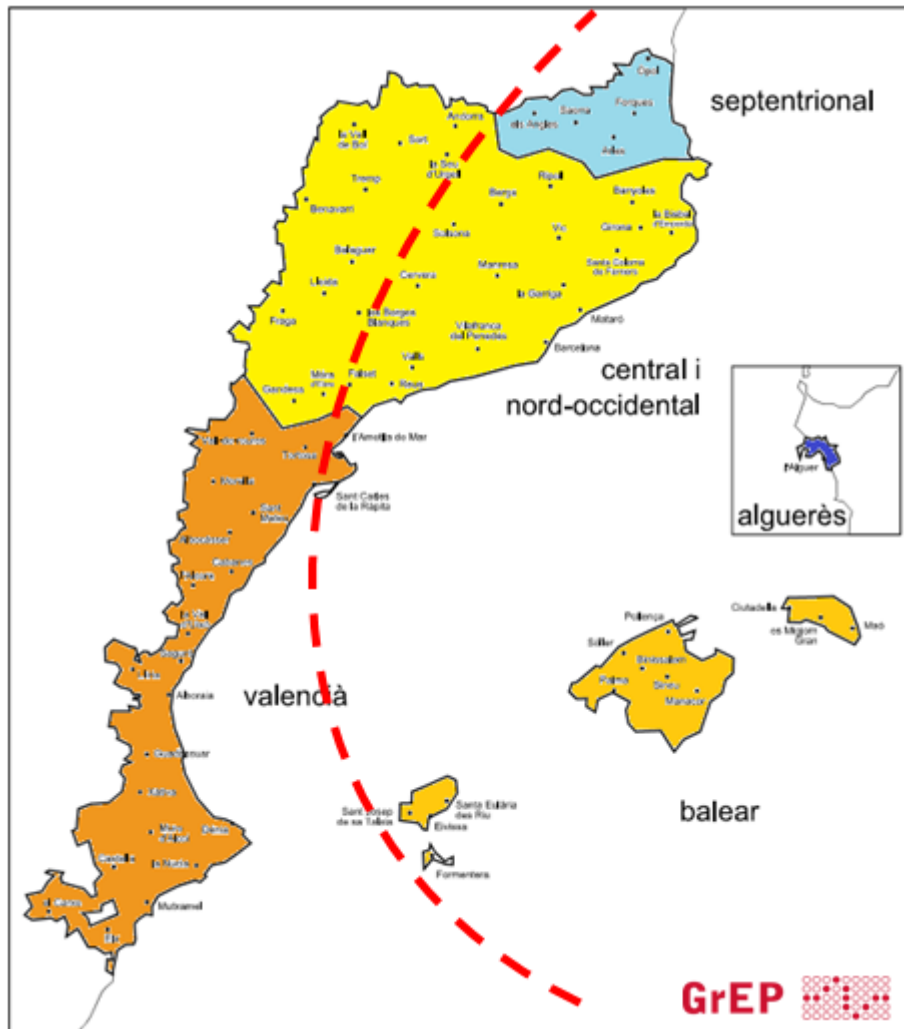


Figura 11. Representación en forma de mapa de una dialectometrización de datos entonativos del catalán (adaptada de Prieto y Cabré, 2013)

## 7. La dialectometría entonativa alfabética y numérica: primeros intentos

Si en los años entre el 2008 y el 2012 las dos líneas de análisis dialectométrico entonativo (la numérica y la alfabética) se desarrollaban por separado, a partir del 2012 se observan los primeros intentos de compaginar los dos métodos en cuestión. En la tesis doctoral de Roseano (2012) sobre la entonación del friulano se empiezan a utilizar, a la vez, datos numéricos acústicos y datos alfabéticos fonológicos, que se someten a análisis dialectométrico con instrumentos diferentes. Los resultados se comparan entre sí y con los resultados de análisis dialectométricos tradicionales (es decir, de datos no entonativos).

En esta fase, que es la que se caracteriza en verde en la Figura 3, se utilizan paquetes de instrumentos diferentes para analizar, por separado, los datos acústicos numéricos y los datos fonológicos alfabéticos. El primero de ellos, que se utilizó en Roseano (2012), utiliza dos instrumentos que ya existían: *Amper StatDistances* y *Gabmap*. El segundo conjunto de instrumentos es el que se emplea en Fernández Planas, Dorta, Roseano, Díaz, Elvira-García, Martín Gómez y Martínez Celdrán (2015) y que combina un software existente (*DiaTech*) con una rutina creada *ah hoc* en el Laboratorio de Fonética de la UB y que se denomina *Calcu-Dista*. Mientras que, en el marco del estudio en cuestión, *DiaTech* se utiliza para llevar a cabo el análisis de los datos alfabéticos (es decir, de las transcripciones fonéticas realizadas con *AmpeEti*), *Calcu-Dista* se ocupa de dialectometrizarse los datos numéricos y es un conjunto de scripts de Praat (Elvira-García, 2014) que predisponen una matriz de distancias que se analiza mediante SPSS. En este sentido, *Calcu-Dista* representa la evolución de las rutinas del mismo tipo que se habían creado en el mismo laboratorio unos años antes y que se describen en la Sección 6 de este trabajo. No sorprende, por lo tanto, descubrir que *Calcu-Dista* utiliza el cálculo de distancias prosódicas de Hermes (1998a,b) y la fórmula de RMS para crear una matriz como la que se propone en la Figura 12, con el objetivo de proporcionar representaciones gráficas e indicadores estadísticos como los que aparecen en las Figuras 13 y 14, todas ellas extraídas de Fernández Planas et ál. (2015).

**Matriz de distancias**

Caso	distancia euclídea												
	1: Tenerife	2: Gran_Ca naria	3: La_Palm a	4: Lanzarote	5: Palencia	6: Salaman ca	7:Madrid	8: Barcelon a	9:Lleida	10: La_Haba na	11: Santiago _Cuba	12:Bullas	13: Mallorca
1:Tenerife	,000	,865	1,813	,687	2,587	3,789	2,734	2,049	2,699	1,681	,797	2,785	2,911
2:Gran_Canaria	,865	,000	1,775	,853	2,139	3,659	2,873	2,394	2,738	1,434	,921	2,715	2,950
3:La_Palma	1,813	1,775	,000	1,339	3,123	3,541	2,024	2,033	2,034	2,476	1,586	2,420	2,656
4:Lanzarote	,687	,853	1,339	,000	2,662	3,720	2,504	2,066	2,491	1,815	,760	2,674	2,837
5:Palencia	2,587	2,139	3,123	2,662	,000	3,827	3,784	3,449	3,530	1,916	2,503	3,398	3,562
6:Salamanca	3,789	3,659	3,541	3,720	3,827	,000	2,125	2,549	2,089	3,032	3,451	1,394	1,328
7:Madrid	2,734	2,873	2,024	2,504	3,784	2,125	,000	1,274	,888	2,783	2,410	1,273	1,094
8:Barcelona	2,049	2,394	2,033	2,066	3,449	2,549	1,274	,000	1,288	2,411	1,962	1,685	1,447
9:Lleida	2,699	2,738	2,034	2,491	3,530	2,089	,888	1,288	,000	2,832	2,493	1,317	1,140
10:La_Habana	1,681	1,434	2,476	1,815	1,916	3,032	2,783	2,411	2,832	,000	1,330	2,325	2,558
11:Santiago_Cuba	,797	,921	1,586	,760	2,503	3,451	2,410	1,962	2,493	1,330	,000	2,435	2,618
12:Bullas	2,785	2,715	2,420	2,674	3,398	1,394	1,273	1,685	1,317	2,325	2,435	,000	,807
13:Mallorca	2,911	2,950	2,656	2,837	3,562	1,328	1,094	1,447	1,140	2,558	2,618	,807	,000

Esta es una matriz de disimilaridades

Figura 12. Matriz de distancias creada a partir de datos entonativos numéricos (Fernández Planas et ál., 2015)



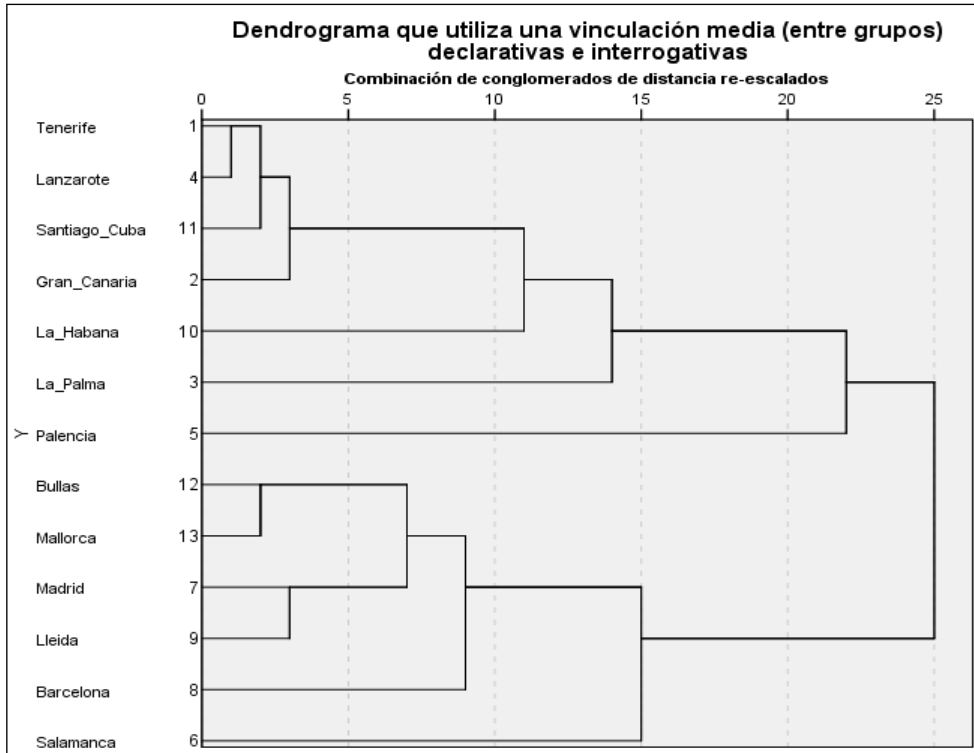


Figura 13. Dendrograma (con indicación de los algoritmos utilizados) creado en SPSS a partir de datos entonativos numéricos (Fernández Planas et ál., 2015)

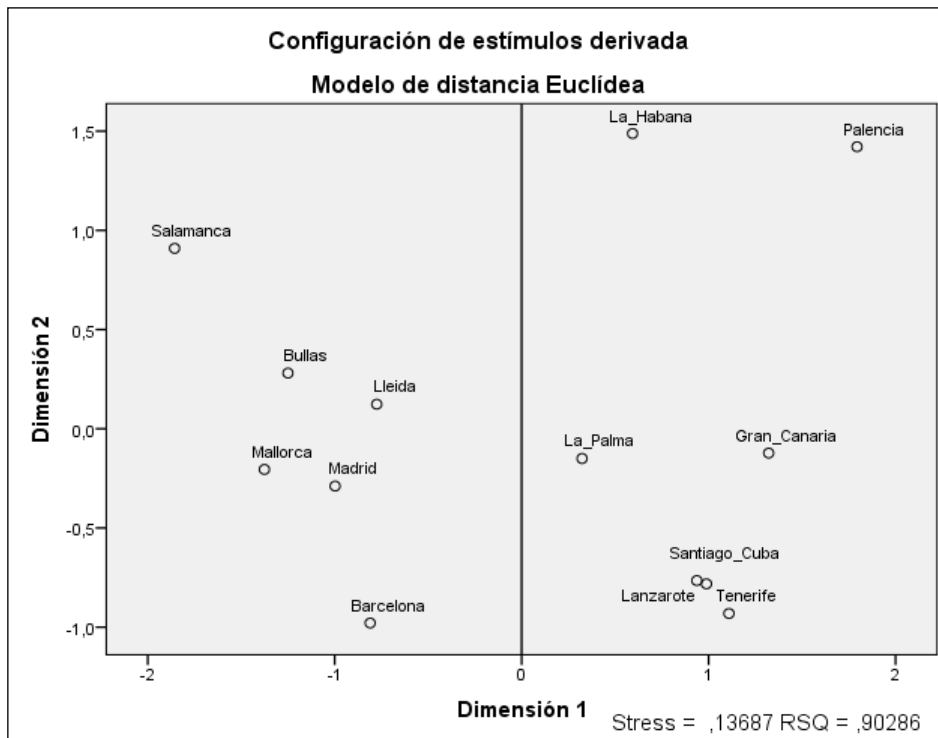


Figura 13. Mapa MDS (con indicación de *stress* y *RSQ*) creado en SPSS a partir de datos entonativos numéricos (Fernández Planas et ál., 2015)

En comparación con las rutinas anteriores del mismo tipo creadas por el equipo del Laboratorio de Fonética de la Universitat de Barcelona, *Calcu-Dista* presenta la ventaja de agilizar ligeramente los procedimientos. Aun así, sigue siendo un sistema poco práctico, en el sentido de que, para realizar una dialectometrización de datos

entonativos acústicos, sigue siendo necesario implementar una serie de acciones con programas diferentes (Praat y SPSS). Además, si bien es cierto que con *Calcu-Dista* se obtiene una representación gráfica de los resultados, cabe decir que esa representación es relativamente básica y que, sobre todo, falta uno de los elementos tradicionales de las representaciones dialectométricas: la trasposición en un mapa geográfico de los resultados del análisis de clúster. Antes de explicar cómo, en los últimos años, se ha intentado superar esas limitaciones (Sección 10), es oportuno presentar brevemente los primeros resultados de una línea de investigación dialectométrica diferente (Sección 9).

### 8. La dialectometría entonativa perceptiva: primeras aproximaciones

Los estudios dialectales se centran a menudo en los datos de producción, es decir en las diferencias y similitudes que existen entre la manera de realizar un elemento léxico o entonativo en distintos puntos de encuesta. Sin embargo, en las últimas décadas también se ha planteado la cuestión de la percepción de las distancias dialectales (v. Long, 1999; Montgomery y Beal, 2011, entre otros).

En el sector de los estudios suprasegmentales, los intentos de aplicar técnicas dialectométricas a datos perceptivos son escasos. Entre ellos, cabe destacar el de Fernández Planas, Roseano, Elvira-García, Carrera Sabaté y Román Montes de Oca (en prensa), que aplica técnicas de ese tipo a los resultados de diferentes test de percepción de contornos entonativos de lenguas románicas. Desde el punto de vista metodológico, para los experimentos en cuestión se utilizó Praat para la suministración de los test de percepción y la creación de unas matrices de confusiones (como la que se reproduce en la Figura 14, adaptada de Fernández Planas et ál. en prensa) y, a continuación, SPSS para la elaboración en clave dialectométrica en forma de dendrogramas o mapas MDS (como el que se propone en la Figura 15, adaptado de la misma fuente. En ese sentido, el método es comparable al que se ha descrito para *Calcu-Dista*.

	Barcelona	L'Alguer	PortuTurre	Biddanoa	Siena
Barcelona	92,325	50,8125	34,3	11,3125	60,875
L'Alguer	60,0625	91,1375	45,9625	16,95	35,8875
PortuTurre	40,7125	41,1375	92,6375	33,322222	43,3
Biddanoa	32,675	36,75	59,25	95,575	25,8
Siena	61,2875	39,1125	41,5375	15,725	92,3375

Figura 14. Matriz de confusiones creado a partir de datos de percepción obtenidos con Praat (Fernández Planas et ál. en prensa)

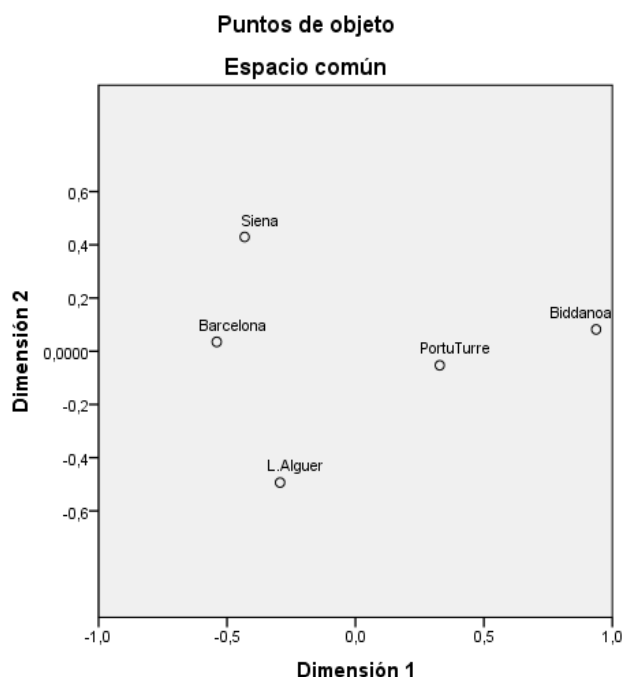


Figura 15. Mapa MDS creado en SPSS a partir de datos de percepción (Fernández Planas et ál. en prensa)

## 9. La dialectometría entonativa numérica: su desarrollo

En los apartados anteriores se ha visto como los años entre 2008 (cuando se publicaron los primeros resultados obtenidos con *Amper StatDistances*) y 2015 (cuando aparece *Calcu-Dista*) han conocido un desarrollo bastante rápido de la dialectometría entonativa. El 2015 es, justamente, el año que seguramente marca un punto de inflexión en este camino, ya que se presentan oficialmente dos instrumentos dialectométricos elaborados de forma independiente en dos equipos relacionados con el proyecto AMPER: la *Ferramenta* de la Universidad de Santiago de Compostela (Martínez Calvo y Fernández Rei, 2015) y *ProDis* (Elvira-García, Balocco, Fernández Planas, Roseano y Martínez Celdran, 2015), el nuevo aplicativo dialectométrico del Laboratorio de Fonética de la Universitat de Barcelona.

A pesar de presentar diferencias interesantes, los dos instrumentos responden a unas exigencias comunes y presentan unas soluciones comparables. Por un lado, de hecho, se proponen agilizar el análisis dialectométrico de los datos entonativos contenidos en la base de datos de AMPER proponiendo un único *software* que permita llevar a cabo la totalidad de las operaciones que, en fases anteriores, requerían programas distintos (v. Secciones 6 y 8). Por otra parte, amplían el abanico de resultados estadísticos que se podían obtener con SPSS y añaden la posibilidad de representar los resultados mismos en mapas geográficos. Además, ambas se dirigen a un público que no tiene conocimientos de programación informática.

### 9.1. Ferramenta

La *Ferramenta* presentada por Martínez Calvo y Fernández Rei (2015) es una herramienta informática desarrollada con el *software* estadístico R que permite al usuario realizar un análisis dialectométrico del corpus AMPER. El análisis realizado se fundamenta en las fórmulas de la distancia prosódica propuestas por Hermes (1998a,b)

en los años noventa y que se habían utilizado también en *Amper StatDistances* y *Calcu-Dista*. A partir de allí, calcula la variabilidad observada entre los diferentes informantes de un punto de encuesta y la representa con un gráfico de cajas, al igual que *Amper StatDistances*. La Figura 16 ofrece un ejemplo de ello, adaptado de Martínez Calvo y Fernández Rei (2015).

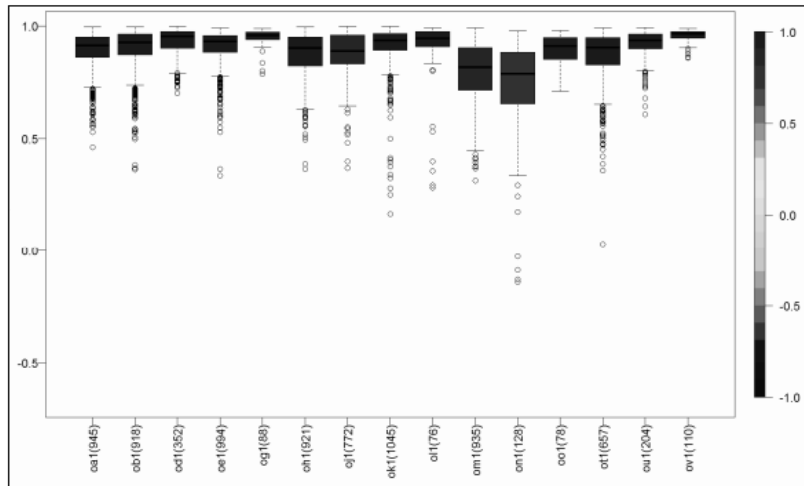


Figura 16. Representación gráfica de la variabilidad en los diferentes puntos de encuesta (adaptado de Martínez Calvo y Fernández Rei, 2015)

A continuación, construye una matriz de correlaciones entre puntos de encuesta y la representa gráficamente de una manera que es mucho más transparente e intuitiva: en vez de generar una matriz numérica, el programa propone una matriz con colores, en la que las correlaciones se indican con gradaciones diferentes que van del rojo al azul. La Figura 17, extraída de Fernández Rei, Moutinho y Coimbra (2014) contiene un ejemplo de la representación en cuestión de la matriz de correlaciones.

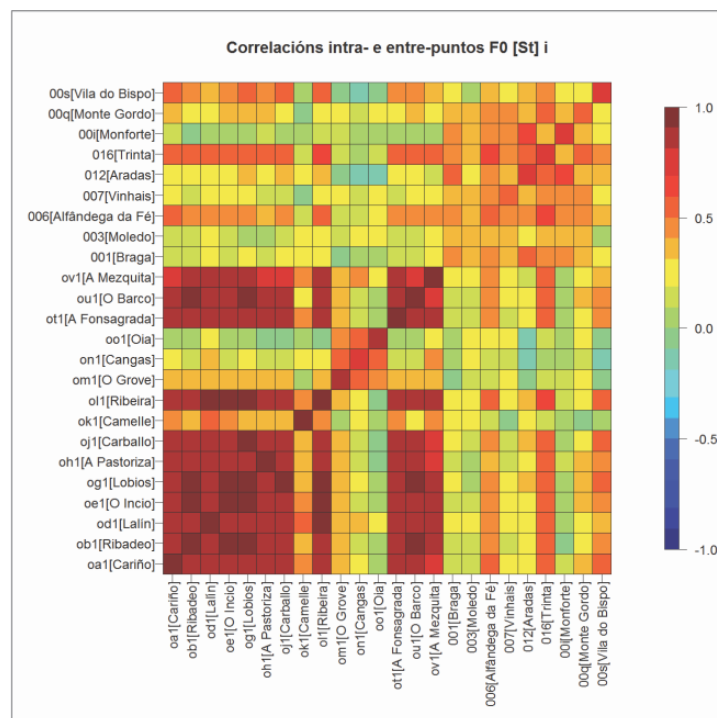


Figura 17. Representación gráfica de las correlaciones entre puntos de encuesta obtenida con *Ferramenta* (adaptado de Fernández Rei, Moutinho y Coimbra, 2014)

A partir de allí, y mediante la aplicación de unos algoritmos (ilustrados en Martínez Calvo y Fernández Rei, 2015), la *Ferramenta* pasa a proporcionar las representaciones gráficas habituales en dialectometría, que son fundamentalmente tres: dendrograma (Figura 18), mapa MDS (Figura 19) y mapa geográfico (Figura 20).

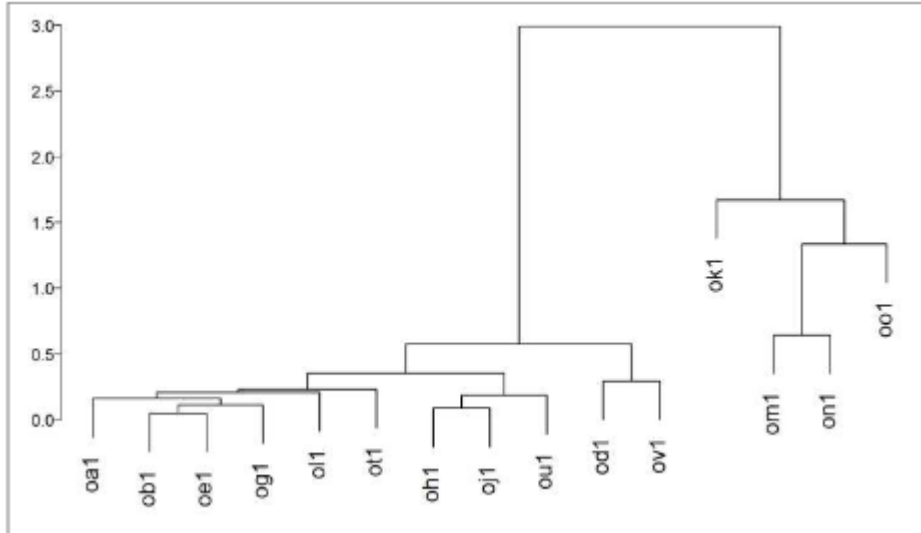


Figura 18. Representación gráfica de los resultados de un análisis dialectométrico con *Ferramenta* en forma de dendrograma obtenido con *Ferramenta* (adaptado de Martínez Calvo y Fernández Rei 2015)

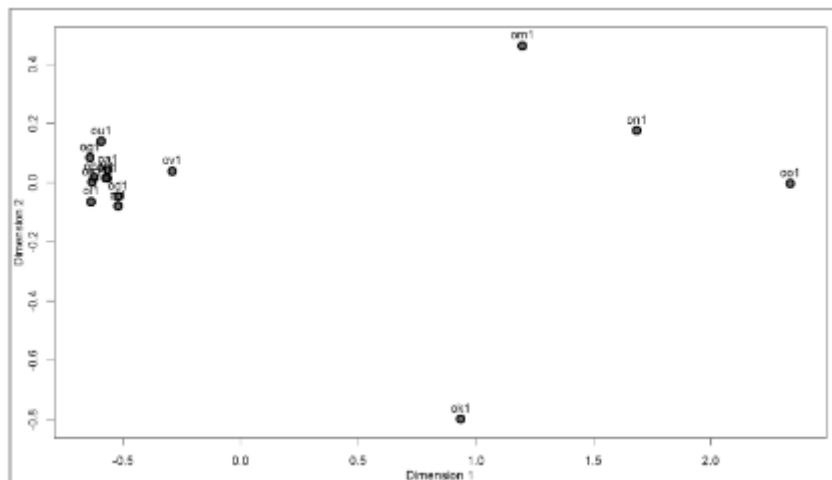


Figura 19. Representación gráfica de los resultados de un análisis dialectométrico con *Ferramenta* en forma de mapa MDS obtenido con *Ferramenta* (adaptado de Martínez Calvo y Fernández Rei 2015)

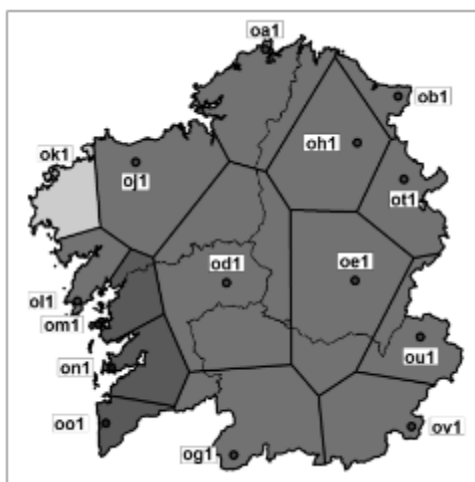


Figura 20. Representación gráfica de los resultados de un análisis dialectométrico con *Ferramenta* en forma de mapa geográfico (adaptado de Martínez Calvo y Fernández Rei 2015)

## 9.2. *ProDis*

En 2015 el Laboratorio de Fonética de la Universitat de Barcelona, gracias a la cooperación con la Facultad de Matemáticas de la misma universidad, anuncia la creación de *ProDis*, un nuevo programa, realizado en MatLab, para el análisis de datos prosódicos numéricos del proyecto AMPER. *ProDis*, al igual que sus predecesores, calcula la distancia prosódica entre distintos puntos de encuesta utilizando las fórmulas de Hermes (1998a,b). Además, ofrece la posibilidad de ponderar los datos de F0 (el principal parámetro prosódico), por otros parámetros acústicos, como la intensidad y la duración<sup>14</sup>.

*ProDis*, tal y como sus antecesores, calcula la variabilidad entonativa (es decir, si y cuánto los informantes utilizan siempre el mismo patrón entonativo) en los diferentes puntos de encuesta. La diferencia con los programas más antiguos reside en el hecho de que *ProDis* presenta la información en cuestión de forma gráfica, es decir utilizando una matriz en escala de colores en las que los colores fríos indican una alta congruencia de los patrones entonativos y los colores cálidos la alternancia entre patrones entonativos diferentes en el mismo sujeto o punto de encuesta (para más detalles ver Elvira-García et ál., 2015). La Figura 21 ofrece un ejemplo de la representación gráfica en cuestión.

<sup>14</sup>Mientras que la ponderación por la intensidad se contemplaba también en los trabajos de Hermes y en la *Ferramenta*, la posibilidad de ponderar por la duración es una de las innovaciones de *ProDis*.

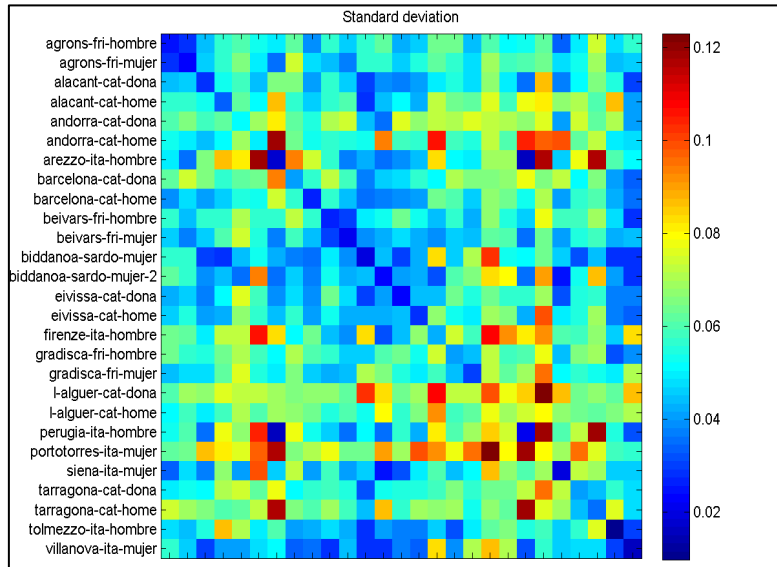


Figura 21. Representación gráfica de la variabilidad en diferentes sujetos y entre diferentes sujetos con *ProDis*

El programa, al igual que la *Ferramenta* descrita en la Sección 10.1, prefiere evitar presentar la matriz de distancias de forma numérica (aunque es una opción que el usuario puede escoger) y utiliza un sistema gráfico basado en gradaciones de colores que van del blanco al rojo oscuro, donde el blanco representa la correlación máxima y el marrón oscuro la más baja. La figura 22 contiene un ejemplo de matriz de correlaciones calculada con *ProDis*.

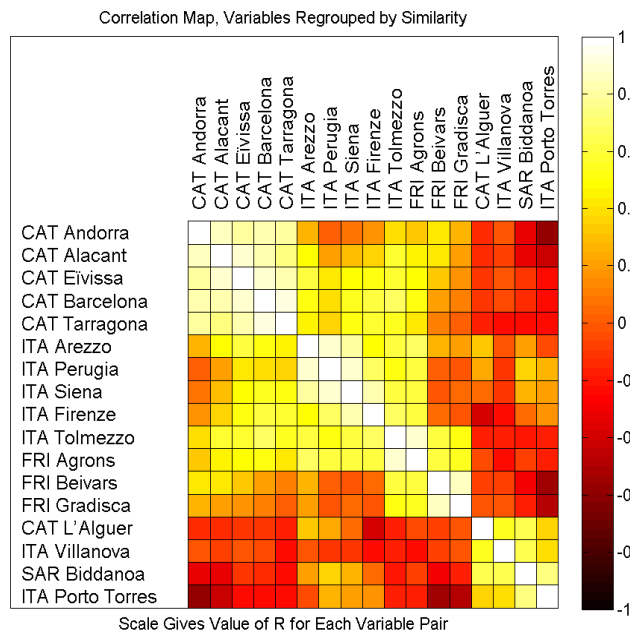


Figura 22. Representación gráfica de las correlaciones entre puntos de encuesta calculads con *ProDis*

A partir de la matriz de correlaciones, y gracias a unos algoritmos que se detallan en Elvira-García et ál. (2015), *ProDis* genera las representaciones gráficas típicas de la dialectometría: dendrograma (Figura 23), mapa MDS (Figura 24) y mapa geográfico (Figura 25).





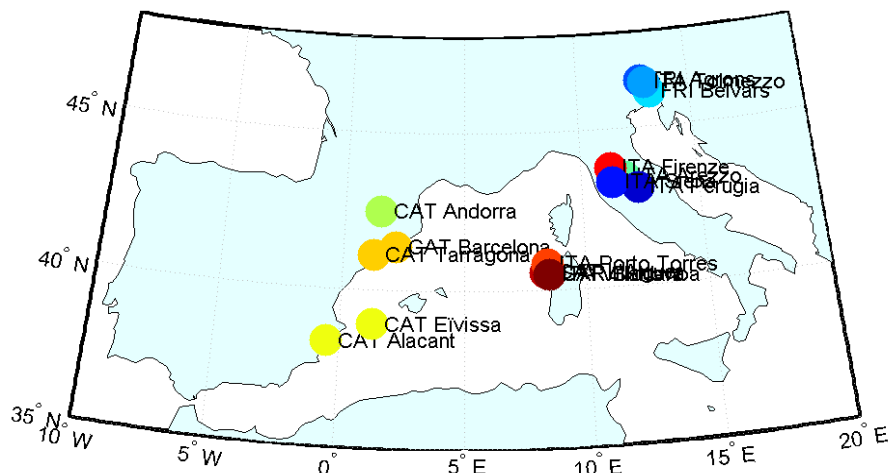


Figura 25. Representación gráfica de los resultados de un análisis dialectométrico con *ProDis* en forma de mapa geográfico

Finalmente, cabe destacar que actualmente (agosto de 2016) *ProDis* se encuentra en fase de ampliación, con el objetivo de añadir otros tipos de análisis estadísticos y *output* gráficos. El objetivo más a largo plazo es la realización de un programa integrado de análisis dialectométrico en el sector prosódico (*ProDis2*) que pueda dialectometrizar, al mismo tiempo, datos entonativos numéricos y alfabéticos, y – en un futuro más lejano – también datos perceptivos.

## 10. Conclusiones

La dialectometría entonativa es, sin duda, una disciplina muy reciente. Aun así, en pocos años (2008-2015) hemos pasado de no tener ningún instrumento para la dialectometría entonativa a disponer de varios programas, lo que demuestra el interés que suscita el tema en cuestión. Sin embargo, a pesar de los avances muy rápidos, incluso los programas de dialectometría entonativa más avanzados (como *Ferramenta* y *ProDis*) no llegan a tener toda la madurez que caracteriza los aplicativos que se utilizan en dialectometría no entonativa (como, por ejemplo, *DiaTech*).

El primer desafío que se tendrá que superar es de carácter técnico y es la portabilidad de los programas en cuestión, en el sentido de que actualmente ambos se pueden aplicar solo a datos entonativos numéricos que se hayan recogido y analizado de acuerdo con la metodología AMPER. Posibilitar que se puedan analizar con esos instrumentos los datos recogidos con otros métodos representa, sin duda, uno de los retos técnicos más importantes de los próximos años.

El segundo desafío que se tendrá que recoger, más a largo plazo, es el de la integración de los resultados de los dos tipos de dialectometría: la tradicional y la entonativa. Los resultados de la aplicación de los distintos programas de análisis dialectométrico a las bases de datos entonativas de las dos familias (AMPER y IARI) no coinciden, en general, con las clasificaciones dialectométricas de las variedades geográficas a partir de elementos segmentales. En otras palabras, la dialectometría tradicional y la dialectometría entonativa no siempre concuerdan. Esa discrepancia conlleva una serie de dificultades, que se pueden dividir en a) desafíos técnicos y b) retos metodológicos en sentido más amplio. Desde el punto de vista técnico, la dificultad principal parece estar relacionada con la integración, en una misma base de

datos analizable conjuntamente, de datos prosódicos (numéricos) y datos segmentales o léxicos (alfabéticos). Por otra parte, cuando se consiga superar ese reto, se tendrá que establecer qué peso tendrá la componente entonativa en las nuevas bases de datos, pero también otras componentes que no siempre han sido el foco de atención de las grandes bases de datos dialectales, como por ejemplo los aspectos sintácticos.

## 11. Referencias bibliográficas

- Aurrekoetxea, G., Fernandez-Aguirre, K., Rubio, J., Ruiz, B. y Sánchez, J. (2013). DiaTech: A new tool for dialectology. *Literary and Linguistic Computing*, 28(1), 23-30.
- Clua, E. (2004). El mètode dialectomètric: aplicació de l'anàlisi multivariant a la classificació de les varietats del català. En M.P. Perea, (Ed.), *Dialectologia i recursos informàtics* (pp. 59-88). Barcelona: PPU.
- Contini, M. (1992). Vers une géoprosodie romane. En G. Aurrekoetxea y X. Videgain(Eds.), *Nazioarteko dialektologia biltzarra: Agiriak* (pp. 83-109). Bilbo: Euskaltzaindia.
- Contini, M., Lai, J.P., Romano, A. y Rouillet, S. (2003). Vers un Atlas prosodique des variétés romanes. En J.C. Bouvier, J. Gourc y F. Pic (Eds.), *Sempre los camps auràn segadas resurgantas, Mélanges offerts a Xavier Ravier* (pp. 73-84). Toulouse: CNRS.
- Elvira-García, W., Balocco, S., Fernández Planas, A.M., Roseano, P. y Martínez Celdran, E. (2015). Presentació d'una aplicació informàtica per a l'anàlisi dialectomètrica de dades prosòdiques en el marc de l'Atles Multimèdia de la Prosòdia de l'Espai Romànic. *Presentación en el Workshop sobre la prosòdia del català*. Barcelona, junio, 22.
- Elvira-García, Wendy, 2014, *Calcu-Dista scripts package*. Praat script. <http://stel.uib.edu/labfon/en/praat-scripts> (data de consultación: 29 junio 2016.)
- Fernández Planas, A.M., Dorta, J., Roseano, P., Díaz, X., Elvira-García, W., Martín Gómez, J.A. y Martínez Celdrán, E. (2015). Distancia y proximidad prosódica entre algunas variedades del español: Un estudio dialectométrico a partir de datos acústicos. *Revista de Lingüística Teórica y Aplicada*, 53(2), 13-45.
- Fernández Planas, A.M., Roseano, P., Elvira-García, W., Carrera Sabaté, J. y Román Montes de Oca, D. (en prensa). From a perceptual point of view, is there continuity between languages in contact?. *Spanish in Context*, 14(2).
- Fernández Planas, A.M., Roseano, P., Martínez Celdrán, E. y Romera, L. (2011). Aproximación al análisis dialectométrico de la entonación en algunos puntos del dominio lingüístico catalán. *Estudios de Fonética Experimental*, 20, 141-178.
- Fernández Rei, E., Moutinho de Castro, L. y Coimbra Lúcia, R. (2013). Abordagem dialectométrica das variedades prosódicas do Galego o do Português. *Presentación en las II Jornadas de Ciências da Linguagem*. Aveiro, junio, 12.
- Fernández Rei, E., Moutinho de Castro, L. y Coimbra Lúcia, R. (2014). As entoacións galega e portuguesa: a fronteira á luz da dialectometría e da percepción. En X. Sousa, M. Negro y R. Álvarez (Eds.), *Lingua e identidade na fronteira galego-portuguesa* (pp. 115-141). Santiago de Compostela: Consello da Cultura Galega.
- Fernández Rei, Elisa (Ed.), 2016, *AMPER-Galicia, Atlas Multimedia Prosódico do Espazo Románico*. <http://ilg.usc.es/amper/> (data de consultación: 29 junio 2016.)
- Frota, S. y Prieto, P. (eds.). (2015). *Intonation in Romance*. Oxford: Oxford University Press.

- Frota, Sónia y Marisa Cruz (Eds.), 2012-2015, *Interactive Atlas of the Prosody of Portuguese*. <http://labfon.letras.ulisboa.pt/InAPoP/> (data de consultación: 29 junio 2016.)
- Goebel, H. (1981). Eléments d'analyse dialectométrique (avec application à l' AIS). *Revue de Linguistique Romane*, 45, 349-420.
- Goebel, H. (1982). *Dialektometrie: Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*, Wien: Verlag der Öst. Akademie der Wissenschaften.
- Goebel, H. (2013). La dialectometrització dels quatre primers volums de l'ALDC: Una breu presentació. *Estudis Romànics*, 35, 87-116.
- Haimerl, E. (2006). Database design and technical solutions for the management, calculation, and visualization of dialect mass data. *Literary & Linguist Computing*, 21(4), 437-444.
- Hermes, D. J. (1998a). Auditory and visual similarity of pitch contours. *Journal of Speech, Language and Hearing Research*, 41, 63-72.
- Hermes, D. J. (1998b). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language and Hearing Research*, 41, 73-82.
- Long, D. (1999). Mapping non-linguists' evaluations of Japanese language variation. En D. Preston (Ed.), *Handbook of Perceptual Dialectology* (pp. 199-226). Amsterdam: John Benjamins.
- Martínez Calvo, A., y Fernández Rei, E. (2015). Unha ferramenta informática para a análise dialectométrica da prosodia. *Estudios de Fonética Experimental*, 24, 289-303
- Martínez Celdrán, Eugenio y Ana Ma. Fernández Planas (Eds.), 2003-2016a, *AMPER-CAT. Atlas Multimèdia de la Prosòdia de l'Espai Romànic*. [http://stel.uib.edu/labfon/amper/index\\_ampercat\\_cat.html](http://stel.uib.edu/labfon/amper/index_ampercat_cat.html) (data de consultación: 29 junio 2016.)
- Martínez Celdrán, Eugenio y Ana Ma. Fernández Planas (Eds.), 2003-2016b. *AMPER España e Iberoamérica. Atlas multimedia de la prosodia del espacio románico*. <http://stel.uib.edu/labfon/amper/cast/index.html> (data de consultación: 29 junio 2016.)
- Montgomery, C. y Beal, J. (2011). *Perceptual Dialectology*. En W. Maguire y A. McMahon (Eds.), *Analysing Variation in English* (pp. 121-148). Cambridge, NY: Cambridge University Press.
- Moutinho de Castro, L., Coimbra Lidia, R., Rilliard, A. y Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental*, 20, 33-56.
- Moutinho de Castro, Lurdes (Ed.), 2008-2013, *AMPER-POR. Atlas Multimèdia Prosòdico do Espaço Românico - Língua Portuguesa*. <http://pfonetica.web.ua.pt/AMPER-POR.htm> (data de consultación: 29 junio 2016.)
- Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P. y Leinonen, T. (2011). Gabmap: A web application for dialectology. *Dialectologia, Special issue II*, 65-89.
- Perea, M.P. (2010). La dialectometría y su aplicación en el estudio de las variedades dialectales del catalán. *Revista de Filología Asturiana*, 9(1), 109-130.
- Prieto, P. y Cabré, T. (eds.). (2013). *L'entonació dels dialectes catalans*. Barcelona: Publicacions de l'Abadia de Montserrat
- Prieto, Pilar y Paolo Roseano (Eds.), 2009-2013, *Atlas interactivo de la entonación del español*. <http://prosodia.upf.edu/atlasentonacion/> (data de consultación: 29 junio 2016.)
- Prieto, Pilar y Rafèu Sichel-Bazin (Eds.), 2014, *Atlàs interactiu de l'intonacion de l'occitan*. <http://prosodia.upf.edu/atlasintonacion/> (data de consultación: 29 junio 2016.)

- Prieto, Pilar y Teresa Cabré (Eds.), 2007-2012, *Atles interactiu de l'entonació del català*. <http://prosodia.upf.edu/atlesentonacio/> (data de consultación: 29 junio 2016.)
- Prieto, Pilar y Teresa Cabré (Eds.), 2007-2012, *Atles interactiu de l'entonació del català*. <http://prosodia.upf.edu/atlesentonacio/> (data de consultación: 29 junio 2016.)
- Prieto, Pilar, Joan Borràs-Comes y Paolo Roseano (Eds.), 2010-2014, *Interactive Atlas of Romance Intonation*. <http://prosodia.upf.edu/iari/> (data de consultación: 29 junio 2016.)
- Rilliard, A. y Lai, J.P. (2008). La Base de Données AMPER et ses interfaces: Structure et formats de données, exemple d'utilisation pour une analyse comparative de la prosodie de différents parlars romans». En L. Moutinho de Castro y R. Lídia Coimbra (Eds.), *Actas das I Jornadas Científicas AMPERPOR* (pp. 127-139.). Aveiro: Universidade de Aveiro.
- Romano, A. (1999). *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento (Italie): Approche linguistique et instrumentale*, Tesis doctoral, Université Stendhal Grenoble 3.
- Romano, A. y Miotti, R. (2008). Distancias prosódicas entre variedades románicas. En A. Turculeț (Ed.), *La variation diatopique de l'intonation dans le domaine roumain et roman* (pp. 231-249). Iași: Editura Universității Alexandru Ioan Cuza.
- Romano, A., Contini, M., Lai, J.P., y Rilliard, A. (2011). Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. *Revista Internacional de Lingüística Iberoamericana*, 9(1), 13-25.
- Roseano, P. (2012). *La prosòdia del friülà en el marc de l'Atles Multimèdia de Prosòdia de l'Espai Romànic*, Tesis doctoral, Universitat de Barcelona.
- Roseano, P. y Fernández Planas, A.M. (2013). Transcripció fonètica i fonològica de l'entonació: Una proposta d'etiquetatge automàtic. *Estudios de Fonètica Experimental*, 22, 275-332.
- Roseano, P., Fernández Planas, A.M., Elvira-García, W. y Martínez Celdrán, E. (2015). Contacto lingüístico y transferencia prosódica: El caso del alguerés. *Dialectologia et Geolinguistica*, 23, 95-123.
- Roseano, Paolo y Fernández Planas, Ana Maria (Eds.), 2009-2013, *Atlant multimediàl de prosodie des varietâts romanichis*. <http://stel.ub.edu/labfon/ampere/friul/index.html/> (data de consultación: 29 junio 2016.)
- Seguy, J. (1971). La relation entre la distance spatiale et la distance lexical». *Revue de Linguistique Romane*, 35, 335-357.
- Sullivan, J. N. (2011). *Approaching intonational distance and change*, Tesis doctoral, University of Edinburgh.
- Turculeț, Adrian (Ed.), 2010-2013, *AMPROM. Atlas multimedia prozodic român*. <http://amprom.uaic.ro/indexamprom.php> (data de consultación: 29 junio 2016.)
- Valls, E., Nerbonne, J., Prokic, J., Wieling, M., Clua, E. y Lloret, M.R. (2012). Applying the Levenshtein distance to Catalan dialects: A brief comparison of two dialectometric approaches. *Verba*, 39, 35-61.

### Agradecimientos

Proyecto “Tecnologías derivadas de AMPER-CAT y análisis de corpus complementarios” (FFI2015-64859-P MINECO/FEDER), IP Ana Ma. Fernández-Planas.

## EUSKALDUN BERRI ETA ZAHARREN ULERMENA JARREREN INGURUAN

**Nerea Asensio, Haizea Barrios, Amaia Lázaro eta Irantzu Sáez  
Euskal Herriko Unibertsitatea**

### **Laburpena**

Euskaldun zaharrak eta berriak barnebiltzen dituen ikerketa bat egiten da, hizkuntzaren jarrerak interpretatzerako eran desberdintasunik dagoen ikusteko. Lanean jarrera desberdinak kontuan hartu dituzte, konkretuki, abegikortasuna, haserrea, ironia, eta neutroa. Planteatutako hipotesiei dagokienez, bideoa eta ahotsa dauden galderetan, bai ama hizkuntza gaztelania dutenek eta baita euskara ama hizkuntza dutenek ere, bideoan ageri den jarrera erraztasun handiagoz iragarriko dutela uste dute.

**Gako-hitzak:** emozioak, ama hizkuntza, euskara, gaztelania

### **1. Sarrera**

Esku artean dugun lan honetan euskaldun zaharrak eta berriak barnebiltzen dituen ikerketa bat egin nahi izan dugu, hizkuntzaren jarrerak interpretatzerako eran desberdintasunik dagoen ikusteko.

Lan hau burutu baino lehen hipotesi batzuk genituen eta lana egindakoan hipotesi horiek baieztatu edo ezeztatu egin ditugu. Gure lanean jarrera desberdinak kontuan hartu ditugu, konkretuki, abegikortasuna, haserrea, ironia, eta neutroa.

Gure hipotesiei dagokienez, bideoa eta ahotsa dauden galderetan, bai ama hizkuntza gaztelania dutenek eta baita euskara ama hizkuntza dutenek ere, bideoan ageri den jarrera erraztasun handiagoz iragarriko dutela uste dugu. Ahotsa besterik ez dagoen galderetan, aldiz, euskaldun berriek arazo handiagoak izango dituztela uste dugu, eta euskaldun zaharrei dagokienez, berriz, irudirik aurrean ez izan arren, euskaldun berriek baino hobeto ulertuko dutela espero dugu. Orokorrean, euskaldun zaharrek erraztasun handiagoa izango dutela uste dugu, bideoak beren ama hizkuntzan daudelako eta beraien ohiko hizkuntza izanik, euskaldun berriak baino hobeto moldatuko direla pentsatzen dugu.

Jarrerren inguruko hipotesiei dagokienez, uste dugu hoberen antzemango den jarrera orokorrean haserrea izango dela, gero abegikortasuna, ondoren ironia eta, azkenik, neutroa.

Horretaz gain, aipatu beharra dugu euskaldun berri guztiek, B1 eta EGA mailaren artean kokatzen direla, beraz, inkesta egiteko ez dute ulermen arazorik izan.

Azkenik aipatu behar da ikerketa honen helburu nagusia izan dela euskaldun zaharren eta euskaldun berrien artean jarrerak antzemateko gaitasuna desberdina den jakitea.

### **2. Marko teoriko eta kontzeptuala**

Gaminderen (2010) esanetan, jarrerak eta emozioak estuki erlazionaturik dauden arren, ez dira gauza bera. Izan ere, emozioek pertsona baten egoera islatzen duten heinean, kontrolaezinak dira eta jarrerei dagokienez aldiz, kontrolagarriak izateaz gain, mintzatzaileak aukeratzen du zer, nola eta noiz adierazi. Horregatik, azken kasu horretan, interakzioa egotea ezinbestekoa da eta emozioetan berriz, ez da beharrezkoa.

Egia da, adituak esaten duenez (2010), adierazle akustiko berdinak direla emozioak eta jarrerak adierazterakoan erabiltzen ditugunak eta, hori horrela, bien artean dagoen erlazioa ukaezina da. Horren harira, korrelazio positibo batez mintzaten da poza eta jarrera ironikoaren —tristura eta jarrera abegikorraren— eta haserrea eta jarrera errietaliaren artean. Oinarrizko jarrerak, beraz, abegikorra, errietalaria, errikiarra eta ironikoa dira (Aurrekoetxea, Gaminde, Garay eta Romero, 2014).

Horretaz gain, aipagarritzat jotzen du jarrerak, emozioekin gertatzen den antzera, pertsona bakoitzaren hizkuntzaren erabileraren ezaugarritzat jotzea (Gaminde, 2010). Gainera, ikerketa batzuek frogatu dutenez, jarrerak estrategia batzuk erabiliz adierazten dira eta erabiltzen ari garen hizkuntzaren arabera, aldakorrak omen dira.

Jarrerak eskaintzen duten informazio paralinguistikoa hizkuntzen arau gramatikaetik at dago, egoera horietan igorleak hartzaileari eragiten diolako (Aurrekoetxea *et al.*, 2014). Informazio paralinguistikoa hori hiztunok daukagun gaitasun pragmatikoarekin lotuta dago, aritzen garen egoera desberdinetara gure hizkera egokitzeko dugun gaitasunarekin, alegia. Izan ere, desberdina da hiztun batek lagun batekin erabiltzen duen hizkera edo irakasle batekin erabiltzen duena, esaterako. Jarreraren adierazpenean, beraz, pragmatismo hori kontuan izaten dugu hitz egiterakoan, gure jarrera hoberen adierazten duen era aukeratuz.

Aurrekoetxea *et al.* (2014) intentzioa, jarrera eta hizkera estiloak informazio paralinguistikotzat hartzen dituzte, hiztunaren kontrolpean dauden ezaugarriak direlako eta interaktuatzen ari garen pertsonarengan eragiteko erabiltzen ditugulako.

### 3. Metodologia

Lan honetan erabili dugun metodologia bi zatitan banatu dugu. Alde batetik, gure ikerketarako beharrezkoak izan diren datuak batzen duen metodologia eta, bestetik, ikerketatik jasotako datuak aztertzeke erabili dugun metodologia.

#### 3.1. Datuak batzeko metodologia

Erantzule euskaldun berriek eta euskaldun zaharrek jarrera konkretu batzuk ulertzeko duten abilezia neurtu eta konparatu nahi izan dugunez, pausu batzuk eman behar izan ditugu.

Hasteko, landu nahi ditugun jarrerak zeintzuk diren erabaki dugu, oinarrizko jarrerak hautatuz: errietalaria edo haserrea, ironikoa eta abegikorra. Jarrera horiei neutroa gehitu diogu, erantzuleei bidea pixka bat zailtzeko asmoz.

Ondoren, jarrera horiek adierazteko, karga semantikorik ez duen esaldi bat aukeratu dugu. Esaldi hori “irakaslea etorri da” izan da, motza eta zehatza dena eta

jarrera desberdinetan sartzeko aukera ematen duena, arrunta delako eta karga emozionalik ez daukalako.

Jarraitzeko, jarreraren adierazpena bideo eta audioen bitartez egitea erabaki dugu, erantzuleei eskainiko diegun inkestan sartzeko. Bideo eta audio horien grabaketetarako, 22 urteko emakume euskaldun zahar bat aukeratu dugu aktore gisa, bere intonazioa eta keinuak jarrerak ahoz eta keinuz interpretatzean, euskaldun berri batenak baino egokiagoak izango direla uste dugulako. Boluntario horri jarreraren interpretazioa errazteko asmoz, egoera konkretu batzuk planteatu dizkiogu, jarrera bakoitzarekin zerikusia daukatenak:

- > Errietalaria edo haserrea: Klasean zaude, irakaslea joan da eta klasekide batek zarata egiten du. Bere erruz irakaslea klasera bueltatzen da eta esaten diozu: “Irakaslea etorri da”
- > Abegikorra: Jolastokian ume bat negarrez ikusten duzu min hartu duelako eta bere irakaslearekin egon nahi du. Orduan zu bere irakaslea zara eta berarengana hurbiltzen zarenean esaten diozu: “Irakaslea etorri da”.
- > Ironikoa: Klasean zaude eta askojakin bat den klasekide bat sartzen da. Orduan zuk, ironiarekin, esaten diozu beste bati: “Irakaslea etorri da”.
- > Neutroa: Ez dugu egoera konkreturik erabili, inolako jarrera adierazi gabe interpretatzeko esan diogu.

Grabaketei dagokienez, gure laguntzailea lau aldiz grabatu genuen bideoz eta beste lau audioz. Horren arrazoia bideoa galdetegian formatu ezberdinetan sartu nahi izan dugula da, audioa, bideoa edo audioa+bideoa erabiltzean jarreraren pertzepzioa aldatzen den edo ez ikusteko. Bideoak eta horren barianteak galdetegi batean sartu ditugu, eta galdetegian lau erantzun daude galdera bakoitzarentzako: abegikorra, ironikoa, errietalaria, neutroa.

Galdetegirako 15 estimulu atera ziren, jarrera bakoitzeko 3 formatu (audioa, bideoa eta audio+bideo) eta beste hiru estimulu non jarrera batzuk errepikatzen diren. Horren harira aipatzekoa da jarrerak eta formatuak desordenaturik egon direla.

Behin inkesta eginda, hautatutako erantzuleei *Drive* aplikazioaren bidez lortutako esteka bidali diegu. Erantzule horiek ondorengo baldintzak bete dituzte: emakumeak, gazteak (18-25 urte bitartekoak) eta euskaldunak, haien arteko desberdintasun bakarra ama- hizkuntza izanik: gaztelania edo euskara. Emaizten kopuru borobila izatearren, 40 pertsonen erantzunak izan ditugunean inkesta itxi dugu (20 euskaldun berri eta 20 euskaldun zahar).

### 3.2. Datuak aztertzeke metodologia

Datuak aztertzeke *Drivetik* pasatutako inkestetatik jasotako datuak hiru esparrutan banatu ditugu. Alde batetik, ama hizkuntzaren arabera informanteek jarrera bakoitzean izan duten egokitasuna. Horretarako jarrera neutroa alboratu dugu, horretan ateratako emaitzak ez direlako estatistikoki esanguratsuak izan. Bestetik, formatuaren arabera (audio, bideo eta audio+bideo) jarrera bakoitzean informanteek izan duten egokitasuna. Azkenik, orokorrean, ama-hizkuntza kontuan izan gabe, egon den egokitasuna aztertu dugu.

#### 4. Datuen azterketa

Jasotako datuen azterketa egin eta gero, lan honetan estatistikoki esanguratsuak suertatu diren emaitzak azalduko ditugu. Horretarako atal honetan, hiru jarreretan (abegikorra, errietalaria eta ironikoa) ateratako emaitzak desberdinduak ditugu, emaitza orokorrekin bukatuz.

##### 4.1. Abegikorra

Lehen jarrera aztertzean 1. taulan erakusten diren datuak lortzen dira. Bertan erakusten den moduan, euskara taldeko informatzaileek kopuru handiagoak erabiltzen dituzte gaztelania taldeko informatzaileek baino. Alde hori estatistikoki esanguratsua da Khi-karratu probaren arabera ( $\chi^2 =$  (a.m.: 1) 6,522;  $p = 0,011$ ). Formatuei erreparaturaz, ama hizkuntza kontuan izan gabe, egokitasun handiagoa ikusi da audio formatuan.Formatuen artean ikusitako aldeak estatistikoki esanguratsuak dira Khi-karratu probaren arabera ( $\chi^2 =$  (a.m.: 2) 28,050;  $p = 0,000$ ).

	Ondo		Txarto	
	N	%	N	%
Euskara	63	% 63	37	% 37
Gaztelania	45	% 45	55	% 55
Audioa	29	% 72,5	11	% 27,5
Bideoa	55	% 68,75	25	% 31,25
Audioa + Bideoa	26	% 32,5	54	% 67,5

1. taula: Jarrera abegikorrean lortutako emaitzak

##### 4.2. Errietalaria

Errietalaria jarrera aztertzean 2. taulan agertzen diren datuetan erakusten den moduan, euskara taldeko informatzaileek egokitasun handiagoa lortu dute gaztelania taldeko informatzaileek baino. Hala ere, datu hori ez da estatistikoki esanguratsua. Formatuei erreparaturaz, ama hizkuntza kontuan izan gabe, egokitasun handiagoa ikusi da audio formatuan.Formatuen artean ikusitako aldeak estatistikoki esanguratsuak dira Khi-karratua izeneko probaren arabera ( $\chi^2 =$  (a.m.: 2) 25,653;  $p = 0,000$ ).

	Ondo		Txarto	
	N	%	N	%
Euskara	46	% 76,67	14	% 23,33
Gaztelania	42	% 70	18	% 30
Audioa	18	% 45	22	% 55
Bideoa	33	% 82,5	7	% 17,5
Audioa + Bideoa	37	% 92,5	3	% 7,5

2. taula: Jarrera errietalarian lortutako emaitzak

##### 4.3. Ironikoa

Hirugarren jarrera aztertzean 3. taulan erakusten diren datuak azaltzen dira, zeinetan gaztelania taldeko informatzaileek kopuru handiagoak erabiltzen dituzteneuskara taldeko informatzaileek baino. Alde hori estatistikoki esanguratsua da Khi-karratu probaren arabera ( $\chi^2 =$  (a.m.: 1) 4,103;  $p = 0,043$ ). Formatuei erreparaturaz, audioa + bideoa izan da egokitasun handiena jaso duena.Formatuen artean ikusitako



aldeak estatistikoki esanguratsuak dira Khi-karratu probaren arabera ( $\chi^2 =$  (a.m.: 2) 33,447;  $p = 0,000$ ).

	Ondo	Ondo	Txarto	Txarto
	N	%	N	%
Euskara	20	% 25	60	% 75
Gaztelania	32	% 40	48	% 60
Audioa	9	% 11,25	71	% 88,75
Bideoa	20	% 50	20	% 50
Audioa + Bideoa	23	% 57,5	17	% 42,5

3. taula: Jarrera ironikoan lortutako emaitzak

#### 4.4. Orokorrean jarrerak eta motak

Emaitzak orokorrean aztertzean 4. taulan erakusten diren datuak lortzen dira. Bertan ikusten den moduan, jarrera errietalarian ikusi da egokitasun handiena. Jarreraren arteko aldeak estatistikoki esanguratsuak dira Khi-karratu probaren arabera ( $\chi^2 =$  (a.m.: 3) 120,836;  $p = 0,000$ ). Formatuei erreparatu, ama hizkuntza kontuan izan gabe, egokitasun handiagoa ikusi da audio + bideo formatuan. Formatuen artean ikusitako aldeak estatistikoki esanguratsuak dira Khi-karratua probaren arabera ( $\chi^2 =$  (a.m.: 2) 34,479;  $p = 0,000$ ).

	Ondo	Ondo	Txarto	Txarto
	N	%	N	%
Abegikorra	108	% 54	92	% 46
Errietalaria	88	% 73,33	32	% 22,67
Ironikoa	52	% 32,5	108	% 67,5
Neutroa	113	% 94,2	7	% 5,8
Audioa	94	% 47	106	% 53
Bideoa	116	% 58	84	% 42
Audio + Beideoa	151	% 75,5	49	% 24,5

4. taula: Orokorrean lortutako emaitzak

## 5. Ondorioak

Euskaldun berri eta zaharrak alderatzen dituen ikerketa honetatik atera ditugun ondorioak bi zatitan banatuko ditugu. Alde batetik, datu objektiboak islatuko ditugu hipotesietatik eta inkestetatik abiatuz. Bestetik, ikerketa egin ondoren, orokorrean ateratako beste ondorio batzuk. Azkenik, iradokizun batzuk proposatuko ditugu.

Gure lehenengo hipotesiari dagokionez, bideo-audio formatuak emaitza hoberenak izango zituela pentsatzen genuen eta inkestak gauzatu ondoren, horrela izan dela egiaztatu dugu.

Bigarren hipotesiari dagokionez, ahotsa bakarrik dagoen galderetan, euskaldun berriek zaharrek baino zailtasun handiagoak izango zituztela pentsatzen genuen eta ikusi ahal izan dugun moduan, horrela izan da.

Hirugarren hipotesiari erreparatu, esan dezakegu hipotesia bete dela, baina aldea ez dela oso nabarmena izan.

Azkenengo hipotesiari dagokionez, egokitasuna jarrera bakoitzean desberdina izango zela pentsatzen genuen, baina pentsatzen genuen ordena ez da gauzatu. Izan ere, hoberen ulertu duten jarrera neutroa izan da eta guk haserrea eta abegikortasuna izango zirela pentsatzen genuen. Bestalde egokitasun txikiena eduki duena ironia izan da.

Egindako prozesua kontuan harturik beste ikerketa baterako gure iradokizuna aldagaiak aldatzea izango litzateke, adibidez: lagin handiagoa erabili, emakumeak bakarrik erabili beharrean emakumezkoak eta gizonezkoak erabili, adin-tarte desberdinak alderatu, hizkuntza bat baino gehiago alderatu, beste jarrera batzuk ikertu...

Azkenik, ikerketa honekin ondorioztatzen dugu jarrerren pertzepzioa ahozko hizkuntzarekin baino, informazio paralinguistikoarekin duela erlazio handiagoa. Izan ere, euskara edo gaztelania ama-hizkuntza izan, egokitasun handiena keinuen laguntza egon denean gertatu da.

## **6. Erreferentzia bibliografikoak**

- Aurrekoetxea, G.; Etxebarria, A.; Gaminde, I.; Garay, U. eta Romero, A. (2014). *Ahoskera lantzeko argibideak eta jarduerak. Laguntzarako materiala: teoria eta praktika*. Leioa: Euskal Herriko Unibertsitatearen Argitalpen Zerbitzu Nagusia.
- Gaminde, I. (2010). *Bizkaiko Gazteen Prosodiaz*. Mendebalde kultura Alkartea.

## HIZKUNTZA-ALDAKETAREN INGURUKO JARRERA ETA PRESTIGIOAZ<sup>15</sup>

**Ariane Ensunza**  
**Deustuko Unibertsitatea**  
[ariane.ensunza@deusto.es](mailto:ariane.ensunza@deusto.es)

### Laburpena

Ekarpen honetan, alde batetik, euskaldunek hizkuntza-aldaketaren inguruan dituzten usteez arituko gara eta, bestetik, euskara estandarraz, euskalkiez eta, oro har, hizkuntza-aldaketez dituzten ikuspegiak azaleratuko ditugu. Horretarako, Busturialdea osatzen duten 20 herrietako 40 lekukoren testigantzak izan ditugu aztergai; izan ere, lan hau gauzatzeko aurreko ikerketetako emaitzak hartu ditugu kontuan. Horrenbestez, uste linguistikoak jasotzeko galdetegia gauzatzeaz gain, aurreko ikerketetako emaitzak ere ekarri ditugu hitzunez hizkuntzaz egiten duten erabilera erreala azaleratzeko eta, horrela, usteetatik errealitatera dauden jauziak antzeman ditugu. Lan hau gauzatzeko, beraz, 10 hizkuntza-aldagai aztertu ditugu: 5 fonologian, 3 morfologian eta 2 sintaxian.

**Gako-hitzak:** jarrera, prestigioa, hizkuntza-aldaketa, euskalkiak

### 0. Laburpena

Ekarpen honetan, alde batetik, euskaldunek hizkuntza-aldaketaren inguruan dituzten usteez arituko gara eta, bestetik, euskara estandarraz, euskalkiez eta, oro har, hizkuntza-aldaketez dituzten ikuspegiak azaleratuko ditugu. Horretarako, Busturialdea osatzen duten 20 herrietako 40 lekukoren testigantzak izan ditugu aztergai; izan ere, lan hau gauzatzeko aurreko ikerketetako emaitzak hartu ditugu kontuan.

Horrenbestez, uste linguistikoak jasotzeko galdetegia gauzatzeaz gain, aurreko ikerketetako emaitzak ere ekarri ditugu hitzunez hizkuntzaz egiten duten erabilera erreala azaleratzeko eta, horrela, usteetatik errealitatera dauden jauziak antzeman ditugu. Lan hau gauzatzeko, beraz, 10 hizkuntza-aldagai aztertu ditugu: 5 fonologian, 3 morfologian eta 2 sintaxian.

### 1. Sarrera

Badira urteak ikerkuntzaren bidean urratsak ematen hasi ginela. Hasiera batean hirigune bateko (Gernika-Lumo) euskararen aldakortasun diastratikoa izan genuen aztergai eta, beraz, hautatutako hizkuntza-aldaira bakoitzaren erabilera zein gizarte-faktorek baldintzatzen duen aztertu genuen (Ensunza 2014). Ondoren, doktoretzatesirako Busturialdea eskualdeko 20 herrien artean dagoen aldakortasun diatopikoa eta soziolinguistikoa aztertu genuen (Ensunza 2015).

Busturialdea mendebaleko euskalkiaren barruan dago kokatuta eta osatzen duten 20 herrietan euskalki hori egiten da. Hala ere, eremu honetako hizkerak, hizkera gehienek bezala, konbergentzia linguistikoa edo *levelling fenomeno*a jasan dute euskararen etengabeko estandarizazio prozesuaren ondorioz. *Levelling* deritzan fenomenoak hizkeren arteko desberdintasuna berdintzeari dagokio; hau da, hizkeren

<sup>15</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

bereizgarri diren ezaugarriak galtzen dira, normalean hizkera estandarraren eraginagatik. Eta horixe izan da lan honen helburua: Busturialdeko hizkeretan euskara estandarrak edo hizkera indartsuek eragin duten aldaketa aztertzea. Izan ere, gauza jakina da euskara estandarraren sorrerak eta hezkuntzan erabiltzen hasteak eragin handia izan dutela euskaldun guztien berbakeran oso denbora-tarte txikian.

Azken urte hauetan Busturialdeko hizkeran gertatutako aldaketa linguistikoak aztertzeaz gain, hiztunek hizkuntzaren gainean dituzten iritzi eta jarrerak ere bildu ditugu.

## 2. Metodologia

Busturialdeko hiztunek hizkeran gertatutako aldaketa linguistikoen gainean dituzten iritzi eta jarrerak ere biltzeko galdetegi bat egin diegu gure lekukoei. Galdetegi horretan, alde batetik euskara estandarraren erabilgarritasunaz eta balioaz galdetu zaie eta, bestetik, euskalkiaren inguruko galderak ere egin zaizkie. Horrela, bai hizkuntza estandarrak bai euskalkiak duen prestigioa antzeman ahal izan dugu.

Hori dela eta, bi eratako datuak bildu ditugula esan genezake: batetik, hiztunak dituen ikuspegi edo uste linguistikoak eta, bestetik, ezaugarri hauetaz egiten duen erabilera erreala.

## 3. Datuak

Datuak aurkezterako orduan multzo bitan banatu ditugu: alde batetik, hiztunek hizkuntzaren inguruan dituzten iritziak eta, bestetik, hiztunek hizkuntzaz egiten duten erabilera erreala.

### 3.1. Datuak: hiztunen uste linguistikoak

Hiztunei euskara estandarraren erabilgarritasunaz galdetu zaienean, % 92k erantzun du erabilgarri ikusten duela euskaldunok elkar ulertzeko eta lantresna baliagarria dela. Gainontzeko % 8k, ostera, adierazi du ez duela beharrezko ikusten.

Hala ere, lekukoen % 97k adierazi du euskara estandarrak euskalkien “aberastasuna” galtzea ekarri duela, euskalkien ezaugarriak desagertzea eta hitz asko galtzea. Izan ere, ondoren, lekukoei galdetzen zaienean Busturialdeko hizkeretatik zein den desberdinena, % 95ek Bermeokoa dela esan du, han beste berba batzuk erabiltzen direla azpimarratuz. Beraz, batez ere lexiko desberdintasunean oinarritzen dira hiztunen argumentuak eta honi ematen diote pisua. Hala ere, Bermeoko euskarak ez dauka prestigiorik, zazkarra dela zehaztu baitute lekukoek. Hizkera ederra adierazteko eskatu zaienean, ostera, helduak beti Busturialdetik kanpoko herrietara joan dira: Lekeitio edo Arratia esanez. Gazteek, ostera, beraien herriko hizkera balioesten dute, harro daude beraien hizkeraz eta ondoren Lekeitiokoa aipatu dute % 89k.

Bestetik, lekukoek adierazi duten beste ezaugarri bereizgarri bat hitzen amaiera izan da; hau da, leku batzuetan *astoa* gisakoak esaten direla eta besteetan *astue* gisakoak zehaztu dute, beraz, *Artiakue* baina *Gernikekoa* esaten dutela jatorriaren arabera. Badirudi, beraz, banaketa hau ere ondo errotuta dutela Busturialdeko hiztunek.

Hala ere, euskara estandarrak hiztunak ohartzen ez diren hizkuntzaren beste alor

batzuetan ere eragin ditu aldaketak, beraiantzat hain begibistakoak ez direnak. Ikus ditzagun adibide erreal batzuk.

### 3.2. Datuak: ezaugarrien erabilera erreal

Lan hau gauzatzeko 10 hizkuntza-aldagai aztertu ditugu: 5 fonologian, 3 morfologian eta 2 syntaxian. Ezaugarri hauek euskalkietako ezaugarri bereizgarriak dira; hau da, euskara estandarrean lekurik ez daukatenak eta, aldi berean, Busturialdeko hizkeren arteko aldakortasun diatopiko eta diastratikoa islatzen dutenak ere badira. Batzuetan Gaminderen 2007ko lanaren erreferentziak egingo ditut, baina kontuan izan behar da desberdintasuna denbora kontua dela, bere lekukoak eta nireak ez baitira garai berekoak.

#### 3.2.1. Fonologia

**(oa) aldagaia:** *-o* bokalez amaitzen diren hitzei *-a* artikulua gehitzen zaienean, *-oa* batuketak aldaketak jasaten ditu hainbat euskalkitan. Hori dela eta, *-ua*, *-ue* zein *-u* bilaka daiteke, edo *-oa* bere horretan mantendu. Busturialdeko hizkera batzuetan *astue* egiten da eta beste batzuetan, ostera, batuan bezala *astoa*.

*/o+a, e, o/>u* egitea araukoa da sortaldeko azpieuskalkietan, baina sartaldean ez. Busturialdeko helduen hizkeran hedatuago dago arau hau, beraz, sortaldeko azpieuskalkiekin egiten dute bat Gernika-Lumo, Ajangiz, Forua, Errigoiti, Morga, Muxika eta Arratzuk izan ezik. Hauetan *-oa* bere horretan mantentzen da, hitz jakin batzuetan (*pianue*, *karniserue*) izan ezik. Gazteen hizkeran ere antzeko eremua hartzen du aldagai honek, baina Sukarrieta, Murueta, Ibarrangelu eta Mendatan indarra galdu du eta *-oa* egiten hasi dira gazteak, ondoko herrietako hizkeren eraginagatik edota euskara estandarraren influentziagatik.

**(it) aldagaia.** Gauza jakina da hizkera batzuetan [i]bokalak edo haren kide [j]irristariak ondoan dituen kontsonanteak palataliza ditzakeela. (it) aldagaiaren kasuan, aurreko "i" horrek ostean "t" kontsonantea duenean gauzatzen diren aldaerak aztertuko ditugu. (tt) aldaera ([c] herskari sabaikaria), (tx) aldaera ([tʃ] afrikatu sabaiaurrekoa) eta (tz) aldaera ([tʃ] afrikatu hobikari-lepokaria). Bestalde, Busturialdeko hizkeretan palatalizatu gabeko aldaera ere entzun daiteke, (it) aldaera. Gaminderen ikerketaren arabera (2007: 39) aldagai honek kostaldeko eta barrualdeko euskalkiak bereizteko balio du; izan ere, kostaldeko hizkeretan (tx) aldaera erabiltzen dela dio eta (tt) barnealdean. Horrez gain, (tx) aldaera kostaldetik barrurantz zabalduz doala adierazten du.

Gure datuak ere bat datoz Gaminderen datuekin. Hala ere, gure helduen kasuan (tx) aldaera hedatuago ageri da Gaminderen datuetan baino. Izan ere, Murueta, Kortezubi eta Nabarnizen hau da lekukotu dugun aldaera bakarra; Gernika-Lumon (tt % 40) aldaera nagusia eta Foruan eta Ean (tx) aldaerarekin batera. Bestalde, Ibarrangelun ere lekukotu dugu (tt) aldaera (% 40), baina Gaminderen arabera (tx) soilik erabiltzen da.

Gazteen hizkeran, ostera, aldakortasun handia dagoela ikusi dugu eta herri askotan (tz) erabiltzen hasi direla gazteak. Batez ere kostako herrietan eta Gernika-Lumotik hur daudenetan gertatu da aldaketa hau. (tt) aldaera zaharra, ostera, galduz doa; izan ere,

Morga eta Mendatan soilik mantentzen da, eta Nabarnizen (tx) aldaerarekin batera erabiltzen da.

Bestetik, (it) aldagaiaren aldaerek iradokitzen dute afrikatuak neutralizatu egin direla Busturialdeko zenbait hizkeratan; izan ere, helduek (tt), (tx) eta (tz) arteko bereizketa egiten baitute, baina gazteek denak berdin ahoskatzen dituzte, dela (tx), dela (tz). Beraz, helduek [kotʃe], [ajcice] eta [atʃo] desberdintzen badituzte ere, gazteek berdin esaten dituzte guztiak.

**(ind) eta (ild) aldagaiak:** [i]bokalak edo haren kide [j]irristariak ondoan duen "t" kontsonantea palataliza dezakeen bezala, mendebaleko sortaldeko hizkeretan gauza bera gertatzen da "d" kontsonantearekin; izan ere, "ind" edo "ild" taldeetako "n" edo "l" palatalizatzen ditu aurrean duen goiko bokalak eta, honen ondorioz, "d" ere palatalizatu egiten da. Busturialdeko hizkeretan forma palatalizatuak zein palatalizatu gabeak entzun daitezke. Gaminderen ikerketaren arabera (2007: 39) Busturialdeko hizkera guztietan egiten da "y", baina Bermeoko eta Gernikako informatzaile gazte askok "dx" egiten dutela adierazten du.

Gure datuen arabera, helduen hizkeran banaketa nabarmena dago Busturialdea mendebaleko eta ekialdeko hizkeren artean; mendebaleko hizkeretan helduek beti darabiltzate forma palatalizatuak eta ekialdean ez. Nabarnizen arrotzak dira forma palatalizatuak. Gazteen kasuan, ostera, ikusi dugu forma ez palatalizatuak ari direla nagusitzen. Aldaera palatalak arrotzak direnez euskara estandarrean, uste dugu ez dela ausartegia proposatzea hezkuntzan euskara estandarra baliatzeak helduen hizkeratik gazteen hizkerara gertatutako aldaketa hau eragin duela. Gainera, sartaldeko hizkeretan ere arrotzak direnez, honek ere eragina dauka Busturialdeko hizkeretan.

**(i\_a) aldagaia:** Gauza jakina da euskalki askotan *a* bokalaren aurretik *i* eta *u* bokalak nahiz *Vj* eta *Vw* bokal-erdiak daudenean, *a>e*-ra ixtea eragiten dutela, baita bi bokalen artean kontsonantea dagoenean ere. Busturialdeko hizkeretan aldaera biak aurki daitezke. *i* eta *u* bokalek eraginda *a>e* egiteak ala ez egiteak sartaldeko eta sortaldeko azpieuskalkiak bereizteko balio du; izan ere, esan dugunez, sartaldean indartsuagoa da joera hau, sartaldean ezinezkoak diren inguruneetan gertatzen delako. Ezinezko ingurune horien artean daude *-gaz* soziatibo singularra, *bat* zenbatzailea, *izan* aditz laguntzailea klitikoa denean, silaba bakarreko atzizkiak, *-tzat* prolatiboa, *eta* juntagailua, etab.

Busturialdeko hizkeretan arau honek indar handia daukala esan daiteke, testuinguru gehienetan gertatzen baita. Hala ere, ez da sartaldean bezain indartsua; esate baterako, *-gaz* soziatibo singularrean ez da gertatzen, ezta *-tzat* prolatiboarekin ere.

Gazteen hizkeran, ordea, *i* bokala eta *j* bere kide irristariak osteko *a>e* bihurtzeko zuen indarra galduz doa eta ia Busturialdeko herri guztiak hartu ditu, Morga eta Mendata izan ezik. Aldaera hau arrotza izanik euskara estandarrean, helduen hizkeratik gazteen hizkerara gertatutako aldaketa honen arrazoia hauxe dela esan daiteke. Gainera, sartaldeko hizkeretan ere arrotzak denez, honek ere eragina dauka Busturialdeko hizkeretan.

**(au) aldagaia:** Euskalki batzuetan *-au-* diptongoa *-eu-* egiten da: *eurre* 'aurre'; *geur* 'gaur', *geuse* 'gauza'; *euzo* 'auzo', *jeusi* 'jausi [erori]', *euntz* 'ahuntz'. Busturialdeko hizkeretan *-au-* zein *-eu-* aurkitzen dugu.

Gure datuen arabera, 12 herritan bakarrik egiten da beti-beti *au>eu* helduen hizkeran. Mundaka, Gernika-Lumo, Kortezubi, Arratzu eta Nabarnizen ohikoena da, baina sarritan *au* diptongoa mantentzen da (% 30 - % 50 artean); izan ere, pertsona berarengan jaso dugu, adibidez *eurrien* zein *aurrien* aldaera, edota *euzo* zein *auzo*, esate baterako.

Busturialdeko kostako hizkeretan *-au->-eu-* bilakabidea arrotzagoa da barrualderagoko hizkeretan baino; baina, dirudienez, sasoi batean bilakabide hau Busturialdeko herri guztietan gauzatzen zen eta indarra galduz joan da. Honen adierazle da helduen hizkeratik gazteen hizkerara gertatutako aldaketa nabarmena. Hala ere, kontuan izan behar da *-au->-eu* bilakaera arrotza dela euskara estandarrean, beraz, ez da arrisksuegia proposatzea hezkuntzan euskara estandarra baliatzeak helduen hizkeratik gazteen hizkerara gertatutako aldaketa hau eragin izana.

### 3.2.2. Morfologia

#### 1. Izenaren morfologia

**(-kin) aldagaia:** Gauza ezaguna da mendebaleko euskalkian sozietiboko morfema berezia dagoena: *-gaz* singularrean (*lagunegaz*); *-kaz* (*lagunekaz*) pluralean. Hala ere, singularrean *-gaz* darabilten hizkera askotan, pluralerako *-kin* morfema (*lagunekin*) erabiltzen dute, batez ere sortaldeko hizkerak. Beste hizkera batzuetan, oster, bi morfemak erabiltzen dituzte pluralerako; batzuetan *-kaz*, besteetan *-kin*, inolako desberdintasun barik.

Gure datuen arabera, helduen hizkeran Nabarniz, Mendata eta Kortezubiz gain Ibarrangelu eta Ean (-kin) da forma bakarra; Sukarrietan aldaera nagusia; eta Arratzu eta Ereñon (-kaz) da nagusi, baina (-kin) ez da arrotza. Dena dela, Busturialdeko hizkeretan (-kin) sozietiboa galduz doa gazteen hizkeran eta haren lekua (-kaz) sozietiboak hartu du, beraz, sartaldeko azpieuskalkitik hurrago dago gaur egun.

#### 2. Aditzaren morfologia

**(zait) aldagaia:** Gauza jakina da “mendebaleko euskalkian bilakaera fonetiko berezia egon dela NOR-NORI saileko adizkietan [‘zait’ gisakoak]; *y- / dx- / j-* aurkitzen ditugu mendebaletik ekialdera goazela; ez *z-* orokorra. Busturialdean hainbat aldaera aurkitzen ditugu; hala nola, (z) aldaera ([ʒ] frikari sabaiaurreko ahostuna), (dz) aldaera ([d͡ʒ] afrikatu sabaiaurreko ahostuna), (y) aldaera ([j] frikari sabaikari ahostuna) eta (d) aldaera ([d] herskari horzkari ahostuna). Bestalde (y) eta (d) aldaerak erabiltzen dituzten hitzunik, hasierako kontsonante hori ez dute beti ahoskatzen ([ate] ‘zait’, [ako] ‘zaio’, etab.). Beraz, (0) aldaera ere badugu ([ø] kontsonanterik ez ahoskatzea).

Gure datuen arabera, helduen hizkeran nagusi dira (dz) eta (z) aldaera sabaiaurreko ahostunak, beraz, sortaldeko azpieuskalkiarekin egiten du bat zentzu honetan. Gauegiz-Arteagan eta Elantxoben (d) aldaerak dira nagusi. Kortezubin, oster, aldaera guztiak lekukotu dira gure datuetan, bai sabaiaurreko ahostunak, bai (d), bai (y) ere. Gainera, Kortezubi, Gauegiz-Arteaga eta Elantxoben ez eze, Bermeo, Busturia, Murueta eta Ibarrangeluko helduen hizkeran ere lekukotu dugu (d) aldaera, beraz, sartaldeko azpieuskalkiekin datoz bat. Bestalde, helduen kasuan Bermeo, Mundaka,

Forua eta Muxikan (y) aldaera da nagusi gure datuetan, baina ez bakarra, hauetan sabaiurreko ahostunak ere erabiltzen baitira, beraz, hauek ere sartaldeko azpieuskalkiekin datoz bat.

Gazteen hizkeran, ostera, aldaera sabaiurreko ahostunak lekua galduz doaz, Nabarniz, Mendata eta Ean izan ezik; hauek bat datoz sortaldeko azpieuskalkiarekin oraindik. Horren lekua sartaldeko (y) aldaerak hartu du Busturialdea barnealdeko hizkeretan; kostaldean, ostera, (d) aldaera nagusitu da. Beraz, aldaera honek batetik kostaldeko eta barrualdeko hizkerak banatzen ditu gaur egun; eta bestetik, sartaldeko hizkerekin bat datozenak eta sortaldekoekin bat datozenak. Gainera, esango nuke, (d) aldaera euskara estandarren eraginez ere izan daitekeela; hau da, *jausi da* HURA (NOR)> (NOR-NORI) *jeusi da+t* analogiaz sortutako forma izan daiteke, baina oraindik hipotesi hutsa baino ez da.

**(joan-eroan) aldagaia:** *joan* eta *eroan* adizkietan *-i-* bokala tartekatzen da hizkera batzuetan. Busturialdean, *doie*, *daroi* gisako formak daude, baina baita batuarekin bat egiten duten *doa* eta *daroa* gisakoak ere. Horrez gain (due)aldiera ere entzuten da.

Busturialdean banaketa garbia eta antzekoa dauka bai helduen hizkeran, bai gazteen hizkeran; beraz, ez denez aldaketa nabarmenik gertatu helduen hizkeratik gazteen hizkerara, lehengo banaketa bera dago orain ere. Alde batetik, Busturialdeko mendebaleko kostako hizkerak ditugu, *due/darue* formak erabiltzen dituztenak; bestetik, barrualdeko hizkerak ditugu, sartaldeko azpieuskalkian eta euskara estandarrean bezala *doa/daroa* formak erabiltzen dituztenak; eta, azkenik, ekialdeko eta ekialde kostako hizkerak ere bat datoz, *-i-*dun formak erabiltzen baitira, sortaldeko azpieuskalkietan bezala. Beraz, badirudi aldagai hau hizkeraren identitariotzat hartzen dela.

### 3.2.3. Sintaxia

**(bi) aldagaia:** Gauza jakina da mendebaleko hizkeretan bi zenbatzailea izen sintagmaren ondoren jarri ohi dela: *Harek gixonak libru bi ostu deuz* ‘Gizon hark bi liburu ostu ditu’. Hala ere, *bi+izena* da egitura ere entzun daiteke; hau da, *bi* zenbatzailea izen sintagmaren aurretik jartzea.

Helduen kasuan Busturialdeko herri guztietan nagusi eta bakar da (I+bi) aldaera; hau da, *bi* zenbatzailea izenaren ostean jartzen da: *neska bi*, *etxe zuri bi*, gazteen hizkeran, ostera, aldaketa nabarmena gertatu da; izan ere, herri askotan sartu da (bi+I) aldaera eta berau bilakatu da forma bakar. Aldiera zaharrari ondoen eutsi dioten hizkerak Mundaka, Murueta, Ea, Ereño, Nabarniz eta Mendata izan dira. Ajangiz, Sukarrieta eta Gernika-Lumon oraindik erabiltzen da aldaera zaharra, baina berriarekin batera. Hori dela eta, Busturialdeko gazteen hizkera gehienetan *bi neska* eta *bi etxe zuri* gisakoak entzutea ohikoagoa da. Kontuan izan behar da hau dela euskara estandarrekin eskolan erakutsi den aldaera bakarra eta, beraz, ez da harritzekoa aldaketa hau gauzatu izana.

**(hau) aldagaia:** Gauza jakina da, berebat, mendebaleko hizkeretan erakusleak izen sintagmaren aurrean jartzeko joera dagoela: *Onék mutillek guzurre esan deu* ‘Mutil honek gezurra esan du’; *Ori liburue loi dau* ‘Liburu hori lohi dago’. Beste hizkera batzuetan, *erakuslea+izena+erakuslea* egitura erabiltzen da. Gamindek (2007: 95) azaltzen duenez, erakuslearen lekua funtzioaren arabera da eta, beraz, erakuslea



deiktikoa denean izenaren aurrean eta atzean agertzen da, Bedaronan bezala. Bestalde, leku askotan izenaren aurrean edozein erakusle doan arren, izenaren ostean artikulua baino ez dela erabiltzen dio Gamindek. Gure datuetan, ostera, jaso ditugu artikulua ez diren adibideak ere:

- A este chico le han mentido  
*Mutill **oneri** gusurre esan dotzie* (Muxika heldua)
- Mutill **oneri** gusurre esan dotzoe* (Mundaka heldua)
- Mutil **oneri** gusurre esantze* (Nabarniz gaztea)
- Estos chicos están cansados  
*Mutiy **onek** kantzeta daus* (Muxika gaztea)
- Mutiy **auerk** kantsaute dauz* (Nabarniz gaztea)
  
- Aquel hombre ha robado dos libros  
*Gixon **arek** bi liburu ostu deus* (Muxika gaztea)

Busturialdeko herri gehienetan (erak+I) aldaera da nagusi helduen hizkeran; hau da, mendebaleko euskalkietan gaur egunera arte ohikoa izan den bezala, erakuslea izenaren aurretik jartzen da. Hala ere, jada herri batzuetako helduek erakuslea izenaren ostean jartzen dutela ikus daiteke eta (erak+I) aldaerarekin batera (I+erak) aldaera ere erabiltzen dute Mundaka, Busturia, Forua eta Muxikako helduek.

Baina gazteen hizkeran erakuslearen erabileran aldaketa nabarmena gertatu dela ikus daiteke; izan ere, (erak+I) aldaerak indarra galdu du eta honen ordez (I+erak) aldaera lekua hartuz joan da gazteen hizkeran. Gainera, Muxikan zein Gauegiz-Arteagan erabat nagusitu da eremu honetan berria den aldaera.

Aldaera zaharrari ondoen eutsi dioten herriak, berriz, Bermeo, Murueta, Kortezubi, Arratzu, Elantxobe, Ea, Ereño, Mendata eta Morga ditugu. Beste alde batetik, Busturian eta Foruan atzera egin da bilakabidean eta berriro ere (erak+I) forma zaharra bihurtu da nagusi eta bakar.

### 3.2.4. Laburbiltzeko

Aztertutako aldaeren artean, euskara **estandarren** eragina jaso dutenak honako hauek dira: (oa) aldagaia, (it) aldagaia, (ind/ild) aldagaia, (i\_a) aldagaia, (bi) aldagaia eta (hau) aldagaia.

Hala ere, ezaugarri horietako batzuk ondoko **hizkerak** eragindakoak ere izan daitezke. (oa) aldagaiari dagokionez, aldaketa Gernika inguruko eraginagatik gertatu izana ere posible da, izan ere, ondoko herrietako askok bertan topatu dute lana edo bizimodua eta horrek ezinbestean hizkera aldatzea ekarri du.

Bestetik, (it) aldagaiari dagokionez zehaztapenak egitea komeni da, esan baitugu (tx) ahoskatzea kostako hizkeretatik hedatu dela. Hala ere, *aita* edo *badakit* esatea, euskara estandarrek eragin du, zalantzarik gabe.

Bestetik, ikusi dugunez, (joan-eroan) aldagaiak banaketa berarekin jarraitzen du eta, beraz, Busturialdea barruan nahiko **identitarioa** dela esan daiteke.

Soziatiboko *-kaz* morfema ere euskalkiko identitariotzat hartu da, *-kin* izanik hezkuntzan euskara estandarrean erakusten den forma bakarra ez duelako euskalkian errotzerik lortu. Are gehiago, helduen hizkerako *-kin* horiek sarritan estandarreko formatzat hartu izan dituzte lekuko gazteek, helduaroan euskaraz alfabetizatu den pertsonaren ezaugarritzat, jakin gabe Busturialdeko hizkeretan forma hori arrunta zela euskara estandarren aurreko sasoietan.

#### 4. Ondorioak

Lan honek erakusten duenez, konbergentzia linguistiko hau ezaugarri sozialekin lotuta dago; alde batetik, euskara estandarrak esanguratsuki eragin duelako gazteen hizkeran eta, bestetik, eskualdeko beste aldaera batzuk prestigioa hartzen ari direlako.

Hala ere, euskara estandarrak ez du berdin eragin aditz morfologian, fonologian, lexikoan edo syntaxian. Lexikoaren inguruko lanketarik ez dugu egin oraingoan, baina begibistakoa da izan duen eragina. Bestetik, gure datuek erakutsi dutenez, euskara estandarreko formak fonologia eta sintaxi alorrean sartu dira errazen; aditz eta izen morfologian, ordea, ez du aldaketa nabarmenik ekarri, momentuz, behintzat.

Bestetik, hiztunen usteen artean eta errealitatearen artean aldea dagoela ere ikusi dugu; izan ere, haien argumentuak batez ere lexiko desberdintasunean oinarritzen dira, gainontzeko arloetan desberdintasunik ez balego bezala Busturialdeko hizkeren artean.

Baina oraindik ezaugarri asko daude hizkeretan mantentzen direnak, sakon errotutako arrazoi emozionalei lotuta daudelako eta eskualde edo herrietako nortasunaren adierazle direlako. Beraz, lan honek hiztunak dituen ikuspegi edo uste linguistikoaren eta hauen erabilera errealean arteko harremana erakusten du.

#### 5. Erreferentzia bibliografikoak

- Ariztimuño, B. (2010). Tolosako eta Ataungo hizkerak: hizkuntz bariazioa eta konbergentzia-joerak. *Uztaro*, 72, 79-96.
- Aurrekoetxea, G. (2003). Euskalkiak estandarren uholdepean (Arratiako kasua). *Ahokotasuna aztergai*, 167-178.
- Aurrekoetxea, G. (2011). Hizkuntza estandarren eta dialektoen arteko bateratze joerak. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, 40(1-2), 133-159.
- Ensunza, A. (2015). Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan. UPV/EHU
- Ensunza, A. (2014). "Gernika-Lumoko euskararen aldakortasuna: aldagai fonetiko zenbait", BAT Soziolinguistika aldizkaria 90. zenbakia. Andoain: Soziolinguistika Klusterra, 123-149.
- Gaminde, I. (2009). Prosodiaren transmisioaz eta beronen azterketaz. *Mendebalde Kultura Alkartea: Transmisioa eta Erabilerea: XIII. Jardunaldiak. Bilbo*, 73-111.
- González, S. (2001). Barietatea eta estandarra: batuaren eragina ume bizkaitar biren hizkeran. *Enseiukarrean: Deustuko Unibertsitateko aldizkaria*, (15), 89-144.
- Ormaetxea, J. L. (2008). Otxandioko hizkera: adinaren arabeko bariazioa. *FLV*, 108, 249-262.

# KEINUEN ETA BOKALIZAZIOEN ARTEKO HARREMANAK KOMUNIKAZIO-FUNTZIO GOIZTIARRETAN: ELAN SOFTWAREA ERABILIZ<sup>16</sup>

Asier Romero, Aintzane Etxebarria, Irati de Pablo eta Alba Sanz  
Euskal Herriko Unibertsitatea

[a.romero@ehu.eus](mailto:a.romero@ehu.eus), [aintzane.etxebarria](mailto:aintzane.etxebarria), [iratidepablo@gmail.com](mailto:iratidepablo@gmail.com),  
[alba9003@gmail.com](mailto:alba9003@gmail.com)

## Laburpena

Denbora aurrera joan ahala, hizkuntzaren jabekuntzaren gaiaren inguruan ikerkuntza gehiago aurkitzen dira literatura zientifikoan. Aurrerapen teknologikoak metodologia berriak aurrera eramatea ahalbideratu du. Horrek ahozko corpusa lortzeko erabili diren metodologiak erabilgarriak diren ala ez hausnartzera bultzatu du. Hori dela eta, ikerlan honetan hizkuntzaren jabekuntzaren azterketa egiteko erabili daitezkeen metodologia desberdinak deskribatuko dira eta bata edo bestea erabiltzearen ondorioak zehaztuko dira.

**Gako-hitzak:** hizkuntzaren jabekuntza, corpusa, ahozko hizkuntza, metodologia

## Abstract

Insofar as time goes by, in the scientific literature they are more and more research on the subject of language acquisition. Technological progress has also enabled to implement new methodologies. This has led to reflect on whether the methodologies that have been used and those used today are useful to get the data we want. Therefore, in this study, different methodologies that can be used for the analysis of language acquisition contracts and the consequences that use one or the other will be described.

**Keywords:** language acquisition, corpus, oral language, methodology

## 1. Sarrera

XX. mendearen erdialdetik aurrera, hizkuntzalaritzaren orientazio teoriko-metodologikoan aldaketak egon dira, horiek testuinguruan kokaturiko datu errealak ikerkuntzan erabiltzea indartuz. Orientazio hori jarraitzen dutenek, hizkuntzen funtzionamendua ulertzeko posizionamendu idealistak alde batera utzi behar direla adierazten dute, orain arte egin diren hainbat hizkuntza-ikerketak zalantzan jarri dituzte eta posible egin dute diziplina berrien jaiotza, horien artean: Ipar-Amerikako soziolinguista bariasionista (Labov, 2001), britaniar kontextualismoa (Firth, 1957; Halliday & Hasan, 1985), testuaren hizkuntzalaritza (Hymes & Gumperz, 1964), Goffman-en interakzionismo sinbolikoa (1981), antropologia-linguistikoa (Dell Hymes, 1964) edo Garfinkel-en etnometodologia (1967). Diziplina horien guztien baieztapenak behaketan eta erabilera linguistiko errealaren datuen analisisan oinarritzen dira, datuak testuinguru errealetatik hartuak, esanguratsuak eta hiztun errealek emititutakoak.

Ikerlan honen helburua hizkuntzaren jabekuntzaren inguruko ikerketetan ahozko corpusa eratzeko dauden inplikazio metodologiko desberdinak deskribatzea da, euskarri teknologikoan grabatutako datuen tratamendu eta kodifikazioan eta baita laginen erregistrorako tekniken inguruan hartutako erabakien ondorioak baloratzeko xedearekin. Datu hauek kontuan izanik, testuinguruan kokatutako behaketa metodologikoa

---

<sup>16</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

proposatzen da ikus-entzunezko informazioaren tratamendurako hainbat tresnatan oinarrituz.

## **2. Corpusaren linguistikaz: datu errealen beharra eta horien kodifikazioa**

Datu errealak lortzearen beharrak, metodo hauekin eskuratutako hizkuntza-corporak ahozko interakzio naturalaren funtzionamenduan ondorioak erdiesteko balio duen ala ez hausnartzera bultzatu du, batez ere ikertzaileak kontrolatzen ez dituen egoeratan; baita helmen mugatuagoa daukaten ondorio linguistikoak erabilitako metodoen eragina jasan duten ala ez aztertzer ere.

Gaur egun, corpusaren kontzeptua asko aldatu da lehenengo hizkuntzalariek erabiltzen zutenarekin konparatuz. Orain corpusak oso zehaztuak daude eta hurrengo ezaugarriak bete behar dituzte: hizkuntzalariek ataza guztiak automatizatzeko aukera ematen dion tresna elektronikoa izan behar du, datuen egiazkotasuna bermatu behar du, helburuaren arabera irizpideak hautatu behar ditu eta, azkenik, adierazgarria den lagina lortu behar du (Villayandre, 2008).

Bestetik, badaude sailkapen desberdinak corpusen tipologiak ezartzeko (Sinclair, 1996; Llisterri, 1999). Ikerlan honen helburuak aintzat harturik, ahozko corporetan zentratuko gara eta sailka daitezkeen bi tipologietan: ahozko corpusak eta ahozko hizkuntzaren corpora. Lehendabizi, ahozko corporak hizkuntzaren deskribapen fonetiko-fonologikoa egitera bideratuta daude edo hizkuntza-teknologiaren esparruarekin erlazionaturiko sistemetara. Corpus mota hauen baldintzak oso kontrolatuta egon behar dira eta, segmentuetan edo esaldi isolatuetan, irakurritako testuetan edo grabaketetan eta pertsonen arteko elkarrizketen transkribapenetan zentratu behar dira. Orokorrean, diseinu zehatza daukate eta tamaina murriztekoak izan ohi dira, solaskide kopuru handia erabiltzen ez dutelako. Ahozko hizkuntzaren corporan, berriz, grabaketak ingurune naturaletan egiten dira eta lagin espontaneoak, ez planifikatuak bilatzen dira. Kasu honetan, helburu nagusia ez da ezaugarri segmental eta suprasegmentalen analisia egitea, baizik eta ahozko hizkuntzaren transkribapen ortografikoa eta transkribatutako testuaren gaineko analisi linguistiko desberdinak burutzea. Aurrekoarekin konparatuz, diseinu zabalagoa eta tamaina handiagoko laginak erabili ohi dira.

## **3. Datuak biltzeko metodologiak:**

Orokorrean, ahozko hizkuntzaren corpusen formakuntzarako eta sorkuntzarako metodologiak bi talde handitan sailkatzen direla azpimarratu behar da: ez intrusio- eta intrusio-metodoak. Ez dago inolako zalantzarik komunikazio-ekoizpenen dinamikan, ahozko igorpen naturalen lagin hobereenak ikertzailearen eragina gehien murrizten denean lortzen direla. Metodo ez-intrusiboetan ikertzailearen eragin ezak abantaila handiak dakartza: alde batetik, solaskideek ahozko interakzio askea eta rol trukaketa izan dezakete. Horrez gain, audio eta bideo grabagailuei esker, datuen fidagarritasuna bermatzeaz gain, informazioa kontserbatu eta geroago analiza daiteke. Kritiken artean, batez ere, datu pertsonalen inplikazio etikoak kudeatzeko zailtasuna dago. Halaber, analisi linguistiko egokia egiteko beharrezkoa den informazio nahikoa lortzearen zailtasuna ere kritikatzeko da, hots, laginen adierazgarritasun eza. Metodologia intrusiboetan, parte-hartzaileei (familia-kide, tutore edo irakasleari) komunikazioa gidatzeko jarraibideak ematen zaizkie, elkarrizketa natural batean egongo balira bezala jokatzeko. Honen alde dauden ikertzaileak rol baten eta diskurtso naturalaren arteko

parekotasunean oinarritzen dira. Modu horretan, elkarrizketa soziolinguistikoa datuen bilketarako metodotzat hartuta, lagin zabalak biltzeko aukera ematen du. Kritiken artean, artifizialtasuna dago, eta informatzaileak azterturik ez daudenean bezala jokatzeko zailtasuna, “behatzailearen paradoja” bezala ezagutzen dena.

Hizkuntzaren garapenaren inguruko ikerkuntzetan gehienbat zeharkako diseinuak (pertsona-talde bi edo gehiagorekin lan egiten da aldi berean eta elkarren artean konparatu egiten dira), luzetarakoak (pertsona baten edo gehiagoren jarraipen bat egiten da mugatutako denbora batean, bai hainbat astetan, hilabetetan edo urtetan) edo bien arteko konbinazioa, diseinu sekuentziala deitu dena (Baltes, Reese & Nesselroade, 1977).

Diseinua edonolakoa dela ere, metodologia behatzaileaz daiteke; zeinak portaera linguistiko espontaneo aztertzen duen eta ikertzaileak ez duen parte-hartzen, horrenbestez, ez dio solaskideari bere ekoizpen linguistiko eragozten, baina egoerak ematen dionarekin konformatu behar da. Beste metodologia bat esperimental da; kasu honetan, pertsonari portaera jakin bat eginarazten zaio eta konpetentzia linguistikoaren aspektu partikular baten informazio zehatza ematen zaio. Ikerkuntzaren baldintzen kontrol handia dago, zaila da haurren kolaborazioa lortzea eta naturaltasun falta agerikoa da. Eta hirugarren metodologia simulazioa da, horretan ordenagailuaren bitartez giza hizkuntza-portaeraren simulazioa egiten da, informazioa lortuz.

#### **4. Proposamen metodologikoa: testuinguruan kokatutako behaketa**

Metodologia aurkeztea eta azaltzea, zentsu batean, ikerlan osoa aurkeztea eta azaltzea da. Hasteko, haur hizkuntzaren eremuan, ikertuko den objektuaren bereizgarrien inguruan hausnartu behar da, hau da, kasu bakoitzean analisi bide egokiak zehaztu behar dira. Analisi-metodoen hausnarketak halabeharrezko ikuspegi baztertzailerak dakar teknika kualitatiboek baliatzen diren ikerkuntzen eta prozedura kuantitatiboak aurrera eramaten dituzten artean. Oso gaineratik esanda, ikerkuntza kuantitatiboa analisi estatistikoarekin erlazionatu izan da, eta ikerkuntza kualitatiboa, berriz, antropologia eta etnografia esparruekin. Logikoki, bakoitzaren abiapuntua ezberdina da, baina horrek ez du esan nahi euren aplikazioak ere bide desberdina izan behar dutenik. Horrexegatik, esan behar da, gaur egun, ikuspegi integratzaileko metodologia ez-baztertzailerak literatura zientifikoan gehien erabiltzen ari dena dela. Ikertzaile batek analisi modu konkretu bat hautatu dezake oinarritzko erreminta metodologiko bezala, beste metodo batzuen emaitzak alde batera utzi gabe (Ruiz Olabuénaga, 1999). Literatura zientifikoak jada erakutsi du ikuspegi kuantitatibo eta kualitatiboak lan berean erabiltzearen garrantzia, hainbat oztopo gainditzea ahalbidetuz (Bericat, 1988; Bryman, 1992).

Hezkuntza eremuan proposatzen den metodologia “testuinguruan kokatutako behaketa” da eta hurrengo ezaugarriak ditu: metodologia kualitatibo batetik abiatuz, aldagaien arteko erlazioak azter daitezke, geroago metodologia kuantitatibo batekin aldagaien deskribapena egiteko. Pertsona baten edo gehiagoren hizkuntza ikertuko litzateke eta “testuinguruan kokatutakoa” kalifikatzailea gehituko litzaioke, izan ere, behaketa hori ikertzailearen parte-hartze motagatik, ikerketa burutzen den tokiagatik eta ariketa motengatik baldintzatuta dagoelako. Hortaz, testuinguruan kokatutako behaketa hezkuntza testuinguruan erabiltzen den behaketa parte-hartzailea litzateke. Guasch-ek (1997) jada esaten zuen giza taldeen kultura eta bizitza estiloa ulertzeko modu bakarra,

euren eguneroko bizitzan murgilduz izan behar zela eta bertan datuak eskuratuz. Beraz, teknika horretan garrantzitsuena datuen lorpena da eta errealitate linguistikoaren analisia egitea. Testuinguruan kokatutako behaketak izaera induktiboa du eta ikertzailea egunerokotasunean murgilduta dagoenean haren grabaketarako metodologia egoerari egokituz doa bere ikerkuntza helburuei erantzuna emateko xedearekin. Umeen hizkuntza jabekuntza aztertzen den ikerkuntza batean badaude bi elementu kontuan hartzea ezinbestekoak direnak: (1) ikertzailearen parte-hartze maila. Behatzen denean, nahiz eta ikertzailea bigarren maila batean kokatu, ikertzailearen parte-hartzeak eragina izango du solaskideen jokaeran, izan ere, eurek ikertzailearen presentzia jarraituko dute, hori dela eta, efektu hori saihestu beharrean, solaskidearekin lan egin beharko da (Duranti, 2000). Horrexegatik, ikertzaileak haurrekin egokitzapen denboraldi bat egin dezan gomendatzen da grabaketekin hasi baino lehen, konfiantza eta segurtasun giroa bilatuz. (2) Behaketa testuinguru naturalean murgiltzean, oso garrantzitsua da datuen konfidentzialtasuna mantentzea eta Batzorde Etikoaren Aldeko Txostena lortzea. Azkenik, ahaztu behar ez den beste puntuetako bat haurrarekin interakzioa izango duten solaskide oinarrizkoei informatu eta haiekin komunikatu behar dela da, haiek (familia kideak, irakasleak...) izango baitira behaketa testuinguruan ikertzailea barneratzea ahalbideratuko dutenak.

## **5. Datuen tratamendua**

Hizkuntza jabekuntzaren inguruko corpusaren eraikuntzarako funtsezko lau elementuei ipini behar zaie arreta: informatzaileak, grabaketarako irizpideak, transkribapena eta kodifikazioa eta, azkenik, emaitzen analisia.

### *5.1. Informatzaileak*

Informatzaileak aukeratzeko kontuan hartu beharreko parametroak honakoak dira: haurren bizilekua (herri eta ikastetxe mota “publikoa edo itunpekoa”; pasa daitezkeen gertakizunak -adib. haurrak gaixotzea, oporraldiak...-; eta gertaera horietara moldatzea), kopurua (informatzaileei dagokienez, nesken zein mutilen kopurua orekatua izango da eta lagin esanguratsua izan behar da; bestalde, grabaketen kopuruari erreparatuz, maiztasun bat egon behar da batetik bestera aukeratutako etapa kronologikoan ikertu nahi den fenomeno aztertzen ahalbidetuko diguna), adina (ikerkuntzan hizkuntzaren zein aspektu aztertu nahi den arabera, tarte kronologikoa zehaztuko da) eta azkenik, egoera soziala (“maila soziala, maila kulturala, ikuspegi soziokulturala...” eta gaitasun soziokognitikoak). Parametro hauen aplikazioa beharrezkoa da, elementu hauek ikerketaren aldagaiak izan daitezkeelako.

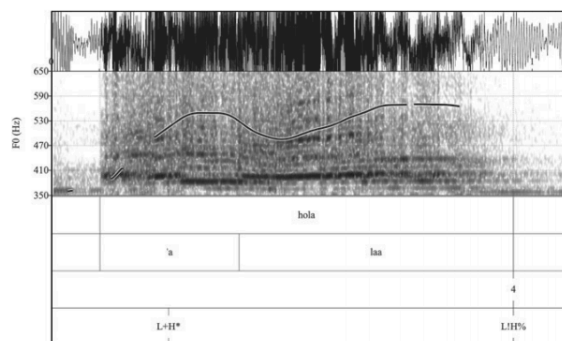
### *5.2. Grabaketarako irizpideak*

Grabaketen irizpidetan, batez ere bost faktore hartu behar dira kontuan: materialaren aukeraketa, erregistraturiko produkzio motak, ariketen hautaketa, solaskideak eta grabaketen iraupena eta maiztasuna. Materialari erreferentzia egiten zaionean, informazioa lortzeko ahalbidetzen diguten teknologien gainean hitz egiten ari gara. Gaur egun arte egon diren berrikuntzek datuen bilketa asko erraztu dute (adibidez, audio kalitate handia erregistratu ahal duten bideokamera asko daude). Erregistraturiko produkzio motei dagokienez, ahozko ekoizpenei egiten zaie erreferentzia. Ariketen kasuan, haurrak gure solaskideak izanik ezinezkoa da saioen planifikazio zorrotza egitea. Ikertzailearen jarrera malgua izan behar da eta interakzio dinamikoa izateko egin

daitezkeen ariketen artean, ipuinak irakurri, haurrentzat ezagunak diren objektuekin jolastu behar dute, beti ere haurrarekin interakzioa izanik. Solaskideak, haurrak izateaz gain, haientzat ezaguna den heldu batekin elkarrekintza mantentzea garrantzitsua da. Azkenik, grabaketen iraupena eta maiztasunari dagokionez, esan behar da haurraren produkzio guztiak erregistratzea ezinezkoa dela, horrexegatik, ezinbestekoa da grabaketen aukeraketa egitea. Grabaketen maiztasunean oreka bat egon behar da (adibidez, astero, hamabostero, hilabetero...) eta iraupena haurrak momentu bakoitzean aurkezten duen kolaboratzeko portaeraren arabera izango da.

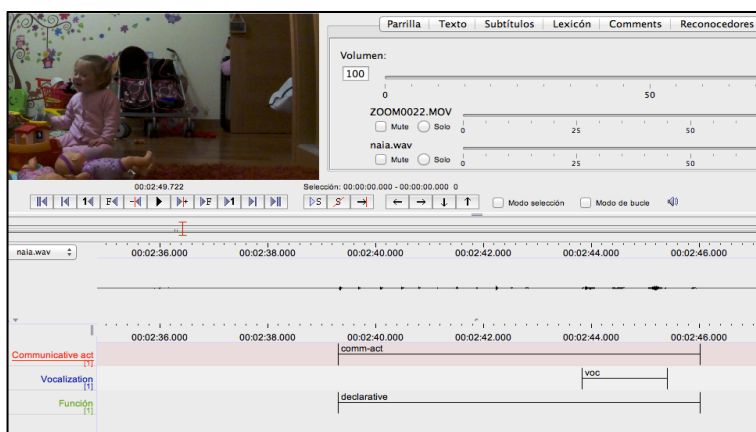
### 5.3. Transkribapena eta kodifikazioa

Gaur egungo aukera teknologikoen ikertzaileei euren lana asko errazten diete eta, horrez gain, haiei esker, ahozko komunikazioaren izaera multimodala azter daiteke (adib. elementu linguistikoak edota keinuak). Ahozko hizkuntzaren ezaugarri guztiak transkribatzea ezinezkoa izango da, horrexegatik, mugak kontuan izanik, kodifikazioaren transkribapena eta etiketatzea presta daiteke, ondoren programa informatikoekin analizatzeko. Transkribapen sistemaren aukeraketa burutu nahi den analisiarekin eta helburuekin zuzenki erlazionatuta egongo da. Datuen transkribapena, kodifikazioa eta analisia burutzeko (hizkuntzaren jabekuntzaren eta garapenaren testuinguruan) ikerkuntzan gehien erabili den proposamenetariko bat CHILDES proiektua izan da (*Child Language Data Exchange System*). *Talkbank* datu-basearekin batera, plataformak CHAT (*Codes for the Human Analysis of Transcripts*) transkribapen formatua eskaintzen du baita hizkuntzaren analisirako programa multzo bat ere, CLAN (*Child Language Analysis Programs*) delakoa. CHILDES plataformak eskaintzen dituen aukerez gain, badaude beste erreminta batzuk ahozko corpusen analisirako: (1) hizkuntzaren analisi akustikoa burutzeko eta ahozko corpusa etiketatzeko erreminten artean PRAAT eta PHON programak daude, eta (2) ahozko hizkuntzaren corpusa zein corpus multimodalaren transkribapena eta idatzoharrak gauzatzeko erreminta ELAN dugu. PRAAT eta PHON programek seinale akustiko-bisualak, sintesi artikulatorioa eta ahozko datuen prozesamendu estatistikoa analizatzen dute, audio seinaleen edizio eta manipulazioa... ahalbidetuz. PRAAT (Boersma & Weenink, 2015) programa ahozko analisi fonetikorako dohaineko erreminta da eta ahozko emisioen parametroen ezaugarriak behatzeko aukera ematen du, bereziki, ahots-tinbrearen ezaugarriak azter daitezke grabatutako ekoizpen akustikoen espektogramaren analisiarekin. Programa horren implementazioa hizkuntzaren jabekuntzaren ikerketetan oso ohikoa da; horrela, ez da soilik software hori erabili duten fonetika edo fonologia azterketa-lanak bilatzen, baizik badira lan askok PRAAT (ikusi 3 irudia) programa erabili dutenak patroï prosodikoen analisia burutzeko ere. PHON (Rose, Hedlund, Byrne, Wareham & MacWhinney, 2007) programa dohainik den eta euskarri teknologikoa duen softwarea da eta, hizkuntzaren atal fonologikoa ikertzeko diseinatu da.



1. irudia: F0ren kontornoaren eta etikatatze prosodikoaren espektogramaren adibidea “hola” hitzarako. (Prieto et al. 2011)

Programa honek hainbat funtzio desberdin ditu, L1ren jabekuntzaren, denboraldian zein hizkuntzaren garapena aztertzea ahalbidetzen dutenak. Bestetik, transkribapen multimodean, ELAN (Lausberg & Sloetjes, 2009) softwarea *freeware* lizentzia duen erreminta profesionala da, audio eta bideo baliabideen inguruko idatzohar konplexuak garatzeko. ELAN hizkuntza, zeinu hizkuntza eta keinuak espresuki analizatzeko diseinatu dago. ELAN programarekin ikertzaileak, aztertutako ikus-entzunezko ezaugarrietatik, idatzohar (hitzak, esaldiak, komentarioak,...) mugagabeak erants ditzake. Oharrak tira desberdinetan kodifika daitezke eta hierarkikoki elkarlotuak egon daitezke (bokalizazioak, keinuak, funtzio pragmatikoak...). Programa honek, hizkuntzaren jabekuntzaren ikerkuntzan, eskaintzen dituen abantailak asko dira: .wav artxiboen uhina ikusteko aukera ematen du, potentzia handiko bilaketa sistema, CHAT eta PRAAT (ikusi 4. irudia) programetatik artxihoak inportatzeko eta esportatzeko aukera.



2. irudia: ELANen argazkia anotazio tirekin

#### 5.4. Emaitzen analisi estatistikoa

Azken hamarkadan haur hizkuntzaren inguruko ikerketen analisiak aldaketa handia izan du datuen analisi estatistikoak direla medio. Zenbat eta baliabide gehiago erabili, orduan eta fidagarriagoak dira emaitzak. Datuen interpretaziorako erabilitako prozedura estatistikoak ikerkuntzaren marko metodologikoarekin erlazionatuta daude. Horrela, datu berdinek interpretazio ezberdinak izan ditzakete marko teorikoaren edota ikuspegi teorikoaren arabera. Baina interpretazio desberdinak ere izango dituzte erabiltzen den prozedura estatistikoaren arabera. Hizkuntzaren jabekuntzaren



ikerkuntzan, literatura zientifikoak prozedura estatistikoa erabiltzen du emaitzen interpretaziorako. Ohikoenak hurrengokoak dira: (1) egindako kodifikazio eta transkribapenaren fidagarritasuna kalkulatzeko epaileen arteko akordioa erabil daiteke, kanpoko ebaluatzaileez eta Kappa indizearen (Cohen, 1960) kalkuluaz baliatuz, eta % 15-20 bitartekoa (2) bi aldagai (edo gehiagoren) artean erlazio esanguratsua dagoen zehazteko proba estatistiko desberdinak erabiltzen dira, nabarmentzen diren artean: Khi-karratua eta Pearson korrelazio koefizientea. Bi lagin edo gehiagoren batezbestekoen konparaketa egiteko, horien arteko erlazio edo independentzia maila kontuan izanik, askotan ere erabili izan dira Bariantza Analisia, ANOVA edo T Student proba; eta, azkenik, (3) zeharkako izaera duten ikerlanetan bereziki Eredu Lineal Mistoa (LMM; West, Welch & Galecki, 2007) erabiltzen da. Zalantzarik gabe, azken eredu hori izaera longitudinaleko datuak aztertzeke egokiena da, izan ere, datuak analizatzeko aukera ematen dute; batetik, emaitzen garapena aztertuz denboran zehar eta koaldagaien efektu zehatzak analizatuz (batez besteko egitura), bestetik, errepikatuta dauden neurrien aldakortasuna aztertzeke aukera ematen du, barruan eta subjektuen artean errorearen eta zorizko efektua analizatuz (kobariantza egitura).

## 6. Ondorioak

Artikuluaren hasieran aipatzen zen bezala, XX. mendearen erdialdetik aurrera, haurren ahozko corpusa erregistratzeko gaur egun erabiltzen diren metodologiek tokiaren berezko errealitate linguistikoa islatzen duten ala ez hausnartzeko beharra sortu da. Metodoak eragiten dituen egoeren datuak biltzeko balio duten ala ez egiaztatu nahi da, hortaz, egoera artifizialak aztertu nahi dira. Autore askok erabilitako metodologiaren ondorioak analizatu dituzte eta, horien artean, Chomsky (1957), Abercrombie (1965) eta Beebe eta Cummings (1996) autoreek hainbat arrazoinamendu eman dituzte orain arte erabili diren metodologiek gabeziak azaltzen dituztela erakutsiz. Horrela, diziplina berri guztiak aipatzen dute, besteak beste, Ipar Amerikako soziolinguista bariasionista (Labov, 2001), britaniar kontestualismoa (Firth 1957; Halliday & Hasan, 1985), testuaren linguistika (Hymes & Gumperz, 1964), Goffman-en interakzionismo sinbolikoa (1981), antropologia linguistikoa (Dell Hymes, 1964) eta Garfinkel-en etnometodologia (1967), beharrezkoa dela behaketan eta erabilera linguistikoko erreal eta esanguratsuetan oinarritzea. Horrexegatik artikulu honetan haurren ahozko corpus naturala euskarri teknologikoan erregistratzeko metodologia ezberdinei erreferentzia egin zaie eta bata edo bestea erabiltzearen ondorioak islatu dira.

Hasteko, kontuan izan behar dugu corporek hurrengo ezaugarriak bete behar dituztela: hizkuntzalariak ataza guztiak automatizatzeko aukera ematen dion tresna elektronikoa izan behar du, datuen egiazkotasuna bermatu behar du, helburuaren arabera irizpideak hautatu behar ditu eta, azkenik, adierazgarria den lagina lortu behar du (Villayandre, 2008).

Bestalde, Sinclair (1996) eta Llisterri (1999) autoreek azaltzen duten bezala, corpusaren tipologia sailkatzeko era desberdinak daude eta, ikerlan honen helburuak kontuan izanda, Llisterri (1996) egiten duen ahozko hizkuntzaren eta ahozko corpusen arteko desberdintasunean oinarritu gara. Ahozko hizkuntzaren corpusa erregistratzeko metodologiei dagokienez, metodo intrusibo eta ez-intrusiboetan sailkatzen dira, azken horiek ahozko hizkuntzaren laginak biltzeko egokienak izanik. Horrez gain, ikerkuntzak zeharkakoak, luzetarakoak edo bien arteko konbinazioa izan daitezkeela gogoratu behar da. Diseinu bat edo bestea hautatu arren, metodologia behatzailea, esperimental edo

simulaziokoa izan daiteke. Halaber, ikerketa kuantitatiboa, kualitatiboa edo bien arteko ikuspegi integratzailea jarrai dezake.

Aurrekoa kontutan hartuz, testuinguruan kokatutako behaketa proposatzen da haurren ahozko corpus naturala erregistratzeko metodologia egokiena bezala. Horretan, pertsona bat edo gehiago aztertzen dira gelan, behaketa parte-hartzailearen bidez eta eragina duten hainbat faktore (tokia, ariketak...) kontuan hartuz. Guasch-ek (1997) azpimarratzen duenez, beharrezkoa da giza taldeen kultura eta bizitzeko modua ulertzeko eta euren eguneroko bizitzaren datuak biltzeko, testuinguru horretan murgiltzea eta barneratzea. Azken hamarkadan izan den aurrerakuntza teknologikoari esker, behaketa zuzena asko hobetu da, gaur egun datuak grabagailu ikus-entzunezkoen bidez kontserbatu daitezke gero analizatu ahal izateko (Golato, 2005). Aitzitik, behatzeko momentuan egintzan eragiten da, solaskideak ikertzailearen presentziaz jakitun direlako, horregatik, efektu hori saihestu beharrean, integratu behar da (Duranti, 2000). Hortaz, haurrak dauden leku batean murgilketa egiten dugunean, ikertzaileak grabaketak hasi baino lehen egokitzapen-aldi bat eman behar du haurrek bera elementu ezezaguntzat hartzeari utzi eta beste elementu baten parekotzat har dezaten; hala, konfiantza eta segurtasun giroa lortuko da. Era berean, funtsezko bitartekarien laguntzarekin (familia, irakasleak...) ikertzaileak testuinguruan murgiltzea lortuko du, eta hauek ere solaskide moduan jardungo dute, beraz, Warren-ek (2006) esaten zuen bezala, elkarriketak betetzen duten papera oso garrantzitsua da; horrexegatik, etengabeko kontaktua mantendu behar da haiekin. Horrez gain, datuen konfidentzialtasuna bermatu behar da eta Batzorde Etikoaren Aldeko Txostena lortu.

Artikulu honen azken partean, datuen tratamenduari erreferentzia egin zaio, nagusiki lau elementuetan arreta jarritz: informatzaileak, grabaketarako irizpideak, transkribapena eta bere kodifikazioa eta, emaitzen analisia. Gainera, egoera bakoitzak eragiten dituen aldagaiak eta horien ezaugarriak deskribatu dira.

Azkenik, inplikazio metodologiko ezberdinen garrantzia nabarmendu nahi da, bai ahozko corpusen bilketarako hizkuntzaren jabeakuntzaren ikerketan, bai datu errealak lortzeko beharrean tokiko egoera linguistikoa islatzeko.

## 7. Erreferentzia bibliografikoak

- Abercrombie, D. (1965). *Studies in Phonetics and Linguistics*. London: Oxford University Press.
- Baltes, P.B., Reese, H.W. & Nesselrode, J.R. (1977). *Life-span developmental psychology*. Monterrey, CA: Brooks/Cole.
- Beebe, L.M. & Cummings, M.C. (1996). Natural speech act data versus written questionnaire data: How data collection method affects speech act performance. In S.M. Gass & J. Neu (Eds.), *Speech Acts Across Cultures: Challenges to Communication in a Second Language* (pp. 65-86). Berlin: Mouton de Gruyter.
- Bericat, E. (1988). *La integración de los métodos cuantitativo y cualitativo en la investigación social. Significado y medida*. Barcelona: Ariel.
- Boersma, P. & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.22. <http://www.fon.hum.uva.nl/praat/>
- Bryman, A. (1992). "Integrating quantitative and qualitative research: how is it done?". *Qualitative Research*, 6(1), 97-113.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Duranti, A. (2000). *Antropología lingüística*. Cambridge: Cambridge University Press.
- Fernández Pérez, M. (1986): *La investigación lingüística desde la Filosofía de la ciencia (A propósito de la lingüística chomskiana)*, Anexo 28 de Verba. Santiago de Compostela, Universidad de Santiago de Compostela.
- Firth, J.R. (1957). *Papers in Linguistics*. Londres: Oxford University Press.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Golato, A. (2005): *Compliments and Compliment responses: Grammatical Structure and Secuencial Organization*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Guasch, O. (1997). *Observación Participante*. Cuadernos metodológicos, 20. Madrid: CIS.
- Halliday, M.A.K. & Hasan, R. (1985). *Language, context and text: a social semiotic perspective*. Victoria: Deakin University Press.
- Hymes, D.H. (1964). *Language in Culture and Society: A Reader in Linguistics and Anthropology*. New York: Harper & Row.
- Hymes, D. & Gumpertz, J.J. (1964). The Ethnography of Communication. *American Anthropologist*, 66(6).
- Labov, W. (2001). The anatomy of style-shifting. In P. Eckert y J.R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp.85-108). Cambridge: Cambridge University Press.
- Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41 (3), 841-849.
- Llisterri, J. (1996). Preliminary recommendations on spoken texts. EAGLES Documents EAG-TCWG-STP/P. May 1996. Disponible en <http://www.ilc.cnr.it/EAGLES96/spokentx/spokentx.html>
- Llisterri, J. & Torruella, J. (1999). Diseño de corpus textuales y orales. In J.M. Blecua, G. Clavería, C. Sánchez & J. Torruella (Eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos* (pp. 45-77). Universidad Autónoma de Barcelona: Milenio.
- Rose, Y., Hedlund, G., Byrne, R., Wareham, T. & MacWhinney, B. (2007). Phon 1.2: A computational basis for phonological database elaboration and model testing. In Buttery, P., A. Villavicencio y A. Korhonen (Eds.) *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, 45th Annual Meeting of the Association for Computational Linguistics* (pp. 17-24). Stroudsburg, PA: ACL.
- Ruiz Olabuénaga, J.I. (1999). *Metodología de la investigación cualitativa*. Bilbao: Universidad de Deusto.
- Sinclair (1996). *Preliminary recommendations on corpus typology*. EAGLES Document EAG-TCWG-CTYP/P.
- Villayandre Llamazares, M. (2008). Lingüística de corpus (I). *E.H. Filología*, 30, 329-349.
- Warren, M. (2006): *Features of Naturalness in Conversation*. Amsterdam: John Benjamins.
- West, B., Welch, K.B. & Galecki, A.T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. New York: Chapman & Hall/CRC.

## HIZKUNTZA ALDAKORTASUNA LARRABETZUKO ADITZ MORFOLOGIAN<sup>17</sup>

Aintzane Etxebarria, Iñaki Gaminde, Ander Olalde eta Ursua Gaminde  
Euskal Herriko Unibertsitatea

[aintzane.etxebarria@ehu.eus](mailto:aintzane.etxebarria@ehu.eus), [inaki.gaminde@ehu.eus](mailto:inaki.gaminde@ehu.eus), [ander.olalde@ehu.eus](mailto:ander.olalde@ehu.eus),  
[ursua.gaminde@gmail.com](mailto:ursua.gaminde@gmail.com)

### Laburpena

Orain arteko hainbat lanek jaso dituzte belaunaldi zaharrek erabili dituzten aditz formak, beste zenbaitek adizkien formak erkatu dituzte, baina gutxik ikertu dute aldagai sozialek aditz morfologian izan duten eragina. Lan honek, hutsune hori gainditu nahian, jasotako datuen azterketa linguistikoen eta kuantitatiboen bidez, Larrabetzuko hiztun komunitateak erabiltzen dituen adizkietan gertatu diren aldaketa mota nagusiak eta aldagai sozialen eragina aztertzeaz gain, informatzaileen sailkapena egin du aditzaren erabilera kontuan hartuta. Horretarako, eta lan-esparrua mugatu nahian, honako aditz sail hauek hartu dira kontuan: nor-nori, nork-nor, nork-nori-nor, baldintzazkoak, aditz trinkoak eta adizki alokutiboak.

**Gako-hitzak:** hizkuntz aldakortasuna, morfologia, Larrabetzu

### 1. Sarrera

Euskararen barietateen ikerkuntzan aditza tradizio luzeko esparrua dela ezin ukatuzkoa da, Bonaparte printzearen sasoirik gaur egun egiten diren herrietako monografiataraino aditz paradigmak nonahi aurki ditzakegu. Hainbat dira azken urteotan, Bizkaiko euskararen esparrura mugatuta, aditzak batu dituzten monografiak (Aranbarri, 1996; Arretxe, 1994; Bilbao, 2002; Errazti, 1994; Etxebarria, 1991; Gaminde, 1983, 1985, 1989a, 1989b, 1990, 1991, 1992, 2005; Gaminde eta Burgete, 1991; Hualde eta lank., 1994; Iglesias, 2014; Legarra, 2011; Markaida eta lank., 1993; Mugarza, 2006, Roman, 2001; Salazar, 2001, Uriarte, 1995; besteak beste). Monografia horietan agertzen diren paradigmekin ez da barietate handirik islatzen eta barietateak agertzen denean ez da adierazten zergatik gertatzen den. Beste ezaugarri komun bat izaten da gehienetan belaunaldi zaharrenean adizkiak jasotzea.

Aditzak herrietako monografietatik harago batu direnean, erkaketa diatopikoak egitekotan eta adizkien formak erkatzeke egin dira (Gaminde, 1984; Yrizar, 1991). Oso kasu gutxitan egin da adizkien azterketa morfologikoa itxurako formatik harago joateko (Aurrekoetxea, 1995; Gaminde, 2007, Gaminde eta lank. 2012; Léonard, 2015).

Aldagai sozialen eragina kontuan hartzen dituzten aditzen gaineko lanak oso urriak izaten dira (Laka eta lank., 2008; Gaminde eta lank. 2012).

Gure lan honen helburua Larrabetzuko hiztun komunitatearen adizki batzuen azterketa linguistikoa oinarrituta, adina eta generoa aldagai sozialen arabera aldakortasuna aztertzea da.

Helburu hau betetzeko aditzaren gainean jaso ditugun datuak era bitara aztertzea proposatzen dugu; batetik adizkien azterketa linguistikoa eta bestetik, horretan oinarrituta, azterketa kuantitatiboa. Datuak konpara daitezke osotasunean, hau da, adizki

<sup>17</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

bakoitzak daukan formari erreparatuta edo azterketa eginda, bere osagaiak erkatuta. Jo dezagun Larrabetzuko "joan" aditzaren adizki batzuk erkatu nahi ditugula Gatikakoekin (1. taula). Itxuraz adizki guztiak desberdinak dira, herri bietako adizkiak % 100ean desberdinak dira. Hala ere, azterketa linguistikoa eginez gero, ikusten dugu pertsonen morfemak eta denborarena bat direla % 75ean (n-, d-, g-, s-, s- p., d- p.) eta desberdinak % 25ean (g- alok, -ie), era berean pluralgilea bat da herri bietan "-s"; azkenik, aditz erroa ez dator bat herri bietan. Aldeak aztertuta, ikusten dugu ezen, morfemei dagozkienak aldaketa morfologikoak direla eta erroari dagokiona fonetikoak.

	<b>Larrabetzu</b>	<b>Gatika</b>
<b>ni</b>	noa	no
<b>hi</b>	goa	o
<b>hura</b>	doa	do
<b>gu</b>	goas	gos
<b>zu</b>	soas	sos
<b>zuek</b>	soasie	soase
<b>haiek</b>	doas	dos

1. taula: "joan" aditzaren indikatibo orain aldia Larrabetzun eta Gatican.

Beraz, adibide honetan erakutsi bezala, azterketa linguistikoa itxurazko desberdintasunetatik harago doa eta agirian jartzen ditu sakoneko aldeak eta kidesunak. Behin alde horiek tasun terminoetan adierazita azterketa kuantitatiboari ekingo litzaioke. Azterketa kuantitatibo horretan tasunen antolaketa ikertzeko aldagai sozial bi erabiliko ditugu, informatzaileen adina eta generoa.

Adizki baten aldakien artean egon daitezkeen aldaketak era bitakoak izan daitezke; batzuk morfologikoak eta beste batzuk fonetikoak. Alde morfologikoetan morfema desberdinak erabiltzea, adizki batzuk ordezkatzeta eta morfemen kokagunea aldatzea garrantzitsuenak dira. Alde fonetikoetan, ostera, kontsonanteen eta bokalen aldaketak, gehitzeak eta kenketak aipa daitezke.

Lana bost ataletan banatuta aurkezten dugu. Sarrera honen osteko bigarren atala datuak jasotzeko eta prozesatzeko metodologiari dagokio, hirugarren atalean datuen azterketa linguistikoa aurkeztuko da, laugarren atalean datu guztien azterketa kuantitatiboa egingo da, azkenik, bosgarren atalean ondorio nagusiak laburbilduko ditugu.

## 2. Metodologia

Atal honetan lan honen helburuak bete ahal izateko behar izan ditugun datuak jasotzeko eta prozesatzeko xehetasun metodologikoak biltzen ditugu.

Datuak jasotzeko beren beregi prestatutako galdekizun bat egin da 24 galderarekin. Azentu arloko galderak 18 izan dira eta intonazio arlokoak 6.

Gure azterketa burutu ahal izateko hogeita lau informatzaile izanditugu. Informatzaileetako hamabik nagusien taldea osatzen dute eta beste hamabik gazteen taldea. Nagusiek 1947-1966 urte artean jaiotakoak dira eta gazteak 1977-1994 urte artean. Adin talde bakoitzeko sei andrazkoak eta beste sei gizonezkoak izan dira. Bigarren adin taldeko informatzaileek euren ikasketak euskaraz egin dituzte. 2. taulan

informatzaile bakoitzeko adin taldea (1 edo 2), generoa (G gizonezkoa edo A andrazkoa), informatzaile zenbakia (batetik seira) eta jaiotze urtearen azken zenbaki biak adierazten dira.

Adina	Gizonezkoak	Andrazkoak
1. Taldea	1G1_49; 1G2_49; 1G3_52; 1G4_53; 1G5_61; 1G6_66	1A1_47; 1A2_48; 1A3_49; 1A4_55; 1A5_61; 1A6_64
2. Taldea	2G1_77; 2G2_79; 2G3_85; 2G4_88; 2G5_92; 2G6_94	2A1_77; 2A2_78; 2A3_85; 2A4_88; 2A5_91; 2A6_94

2. taula: Informatzaileak adin talde bien eta generoaren arabera sailkatuta.

Galdekizunak Marantz PMD610 izeneko grabagailuarekin eta kanpo mikrofono batekin grabatu dira. Grabaketak egindakoan Praat (Boersma eta Weenink, 2015) izeneko programaren bidez transkribatu eta etiketatu egin dira datu basea osatzeko.

Datuen azterketa linguistikoa egitean galdera bakoitzaren tasunak definitu dira geroagoko azterketa kuantitatiboa egin ahal izateko. Azkenik, azterketa kuantitatiboan, behin tasunen hedadura finkatuz gero, SPSS programaren bidez estatistiko batzuk erabili ahal izan dira. Datuak gutxi direnez gero, datuetan ager daitezkeen balizko aldeak estatistikoki esanguratsuak diren ala ez erabakitzeke Mann-Whitney U proba ez parametrikoa eta khi karratua erabili dira eta informatzaileen artean egon daitezkeen multzoak agirian uzteko Multzokatze hierarkikoa deritzon estatistikoa erabili da.

### 3. Datuen azterketa linguistikoa

Atal honetan jaso ditugun adizkien azterketa linguistikoa aurkeztuko dugu. Gaurko belaunaldi bien datuekin batera jasota eduki ditugun gerra aurreko informatzaile biren datuak emateari ere interesgarri iritzi diogu egon litezkeen aldeak kontuan hartzeko.

Azterketa errazago ulertzeko eta aurkezteko aztertu ditugun adizkiak dagozkien aditz sailen arabera aurkeztuko ditugu; horrela bada, atal honen lehen azpi-atalean Nor-Nori saileko adizkiak ikusiko ditugu, bigarren azpi-atalean Nork-Nor sailekoak, hirugarrenean Nork-Nori-Nor sailekoak, laugarrenean baldintzen sailekoak, bosgarrenean aditz trinkoen sailekoak eta, azkenik, seigarrenean adizki alokutiboan erabilerari dagokiona. Adizkiak izendatzeko euskara estandarrenak erabiliko ditugu etiketa gisa.

#### 3.1. Nor-Nori

Sail honetan lau adizki aztertu ditugu; hiru orain aldian “zaio”, “zait” eta “zaizkigu”, eta bat iragan aldian “zitzaizkion”.

“Zaio” adizkiak aldaki bi dauzka: “yako” (10, % 41,67) eta “dako” (14, % 58,33); orain aldian aztertu ditugun hiru adizkien emaitzak aztertuta, ikusten dugu informatzaile bik “da”dun zein “ya”dun adizkiak erabili dituztela. Gerra aurreko informatzaileen artean “dako” eta “yako” biak agertu dira. Tasun hau adinaren arabera adin talde bietan berdin hedatzen da, generoaren arabera, gizonen artean gehiago erabiltzen den arren, aldea ez da estatistikoki esanguratsua. Datuen azterketa kuantitatiboa egitean, adizki honi dagokionez, tasun bi erabiliko dira datuen matrizean 101da eta 102ya.

“Zait” adizkiari dagozkion erantzunak lau izan dira: “yat” (6, % 16,67), “yate” (6, % 25), “date” (11, % 45,83) eta “daste” (3, % 12,5). Gerra aurreko informatzaileek “yat” eta “date” erabili dituzte. Adizkietan agertzen denez, lehen pertsonari dagokion morfemak hiru alomorfo dauzka: “-t” (4, % 16,67), “-te” (17, % 70,83) eta “-ste” (3, % 12,5). “-t” morfema gizonetzko bik eta andrazko bik erabiltzen dute; adinari dagokionez, lautik hiru lehen adin taldekoak dira eta bakar bat bigarren adin taldekoa. “-te” morfemaren hedadurari dagokionez, nagusiek bederatzi aldiz erabiltzen dute eta gazteek zortzi; generoaren arabera gizonetakoek zortzi aldiz erabili dute eta andrazkoek bederatzi aldiz. Azkenik, “-ste” morfema gazteek baino ez dute erabiltzen eta generoaren arabera ez dago alde handirik, izan ere, gizonetzko bik erabili dute eta andrazko batek. Datuen matrizean erabiliko ditugun tasunak morfema bakoitzeko bat izango dira: 103t, 104te eta 105ste.

“Zaizkigu” adizkiaren gaineko erantzunak hiru izan dira: “yakus” (12, % 50), “dakus” (8, % 33,33) eta “daskus” (4, % 16,67). Gerra aurreko informatzaileetan “yakus” baino ez dugu jaso. Pertsonari dagokion morfema “ku” (20, % 83,33) izan daiteke edo “sku” (4, % 16,67). “sku” morfema gazteen erabileretan baino ez da agertzen; generoaren arabera, ostera, ez dago alderik, andrazko bik eta gizonetzko bik erabili dute eta. Datuen matrizean tasun bakarra erabiliko dugu: 106ku.

“Zitzaizkion” adizkiak bost erantzun dauzka: “yakosan” (12, % 50), “yakosen” (3, % 12,5), “dakosen” (7, % 29,17), “akosan” (1, % 4,17) eta “akosen” (1, % 4,17). Gerra aurreko informatzaileek “akosan” eta “yakosan” erabili dituzte. Adizkietan hiru hasiera erabiltzen dira “ya” (15, % 62,5), “da” (7, % 29,17) eta “a” (2, % 8,34). “ya” adinaren arabera, nagusiek zortzi aldiz erabili dute eta gazteek zazpi, eta generoaren arabera, gizonetakoek zazpi aldiz erabili dute eta andrazkoek zortzi; honek erabileran ez dauka alde handirik. “da” morfema, berriz, gazteen artean gehiagotan erabiltzen da, bost aldiz, nagusien artean baino, bi aldiz; antzera gertatzen da generoaren araberrako banaketan, izan ere, gizonetakoek gehiagotan erabiltzen dute, bost aldiz, andrazkoek baino, bi aldiz. Azkenik “a” andrazko nagusiek baino ez dute erabili (bi aldiz).

Pluralaren morfemaren eta azken “n”aren artean erabiltzen den bokala “a” da gehienetan (13, % 54,17), gainerako kasuetan “e” da (11, % 45,83). “a” nagusiek (bederatzi aldiz) gehiago erabiltzen dute gazteek (lau aldiz) baino; generoaren arabera ez dago alderik, gizonetakoek sei aldiz eta andrazkoek zazpi aldiz. “e”ren kasuan, adinaren arabera, kontrara gertatzen da nagusiek (hiru aldiz) gazteek (hamaika aldiz) baino gutxiagotan erabiltzen dute eta, lehen bezala, generoaren arabera ez dago alderik, gizonetakoek sei aldiz eta andrazkoek bost aldiz. Adinaren araberrako aldea estatistikoki esanguratsua da (Khi = 4,196 (a.m. 1)  $p < 0,042$ ).

Datu matrizean adizki honi dagozkion tasunak lau izango dira; hiru adizkien hasieraren arabera “ya” 107ya, “da” 108da eta “a” 109a eta bat pluralaren morfemaren eta amaierako “n”aren arteko bokalaren araberrakoak, “a” denean 110a.

### 3.2. *Nork-Nor*

Sail honetan hamar adizki aztertu ditugu; sei orain aldikoak: “naute”, “gaituzte”, “dute”, “ditut”, “ditu” eta “dituzte” eta lau iragan aldikoak “ninduen”, “nituen”, “zituen” eta “genituen”.

“Naute” adizkiak bost erataro agertzen da; hauetako 13 (“neude” 1, % 4,17; “neudie” 4, % 16,67; “neure” 1, % 4,17 eta “neurie” 7, % 29,17) Nork-Nor sailekoak dira (% 58,33) eta 11 (“dostie”) Nork-Nori-Nor sailekoak (% 45,83); nagusiek gehiagotan erabiltzen dituzte Nork-Nori-Nor sailekoak (sei aldiz) gazteek baino (bost aldiz), aldea oso txikia den arren. Generoari dagokionez, aldea ere txikia da, gizonezkoek sei aldiz eta andrazkoek bost aldiz. Gerra aurreko informatzaileek “neudie” eta “neure” erabili dituzte. Nork-Nor saileko adizkien artean aldaketa nagusi bi gertatzen dira “haiek” morfemaren “d” eta “r” txandaketa eta “e” eta “ie”rena. “d” bost kasutan agertzen da eta “r” zortzi kasutan; “d” gazteen artean gehiagotan erabiltzen da (lau aldiz) zaharren artean baino (behin) eta “r” zaharren artean gehiagotan (bost aldiz) gazteen artean baino (hiru aldiz), hala ere, aldea ez da estatistikoki esanguratsua. Generoaren arabera dagoen aldea oso txikia da eta ez da estatistikoki esanguratsua. “ie”ren erabilerari dagokionez, ez dago alderik ez adinaren arabera ez eta generoaren arabera ere; “e” kasu bitan baino ez zaigu agertu eta biak gazteak eta andrazkoak izan dira. Datu matrizean adizki honi dagozkion tasunak hiru dira: Nork-Nor saileko adizkia izatea edo ez 201nn; haiek morfeman “d” erabiltzea 202d eta “ie” erabiltzea 203ie.

“Dute” adizkiari dagokionez, bostaldaki jaso ditugu: “deure” (13, % 54, 17), “deurie” (8, % 33,33), deudie (1, % 4,17), daurie (1, % 4,17) eta “dere” (1, % 4,17). Gerra aurreko informatzaileek “deure”, “deurie” eta “deudie” erabili dituzte. Hiru eratako aldaketak agertzen dira, erroarekin hasierako bokalak osatzen duen diptongoari dagokiona, hau da, “eu” (22, % 91,67), “au” (1, % 4,17) eta “e” (1, % 4,17) izatea eta “haiek” morfemarekin gertatzen diren aldaketa biak, hots, “d” eta “r” txandaketa eta “ie” “e” txandaketa. Lehenari dagokionez “d” kasu bakar batean agertzen da (% 4,17); bigarrenari dagokionez, “ie” hamar kasutan agertzen da (% 41,67) eta “e” hamalau kasutan (% 58,33). “ie” eta “e” aldakien arteko aldeari dagokionez, esan behar da generoak ez daukala eraginik eta adinari dagokionez, eragina estatistikoki esanguratsua ez den arren, badagoela nolabaiteko joera esan daiteke, izan ere, nagusiek “ie” zazpi aldiz erabiltzen dutela eta “e” bost aldiz eta gazteek, berriz, “ie” hiru aldiz eta “e” bederatzi aldiz. Adizki honi dagozkion tasunak horrenbestez honako hauek dira: 204eu, 205au, 206e, 207d eta 208ie.

“Gaituzte” adizkiak hiru aldaki dauzka; horietako 14 (“gaitues” 1, % 4,17 eta “gaitue” 13, % 54,17) Nork-Nor sailekoak dira (% 58,33) eta 10 Nork-Nori-Nor sailekoak (% 41,67), beronen erabileran ez dago alderik adinaren arabera eta generoaren arabera oso txikia da (gizonezkoek sei aldiz, andrazkoek lau aldiz). Nork-Nor saileko adizkien artean aldaketa nagusi bat gertatzen da; informatzaile gehienek (hamairuk) pluralgile bakarra “it” darabilte eta informatzaile batek bi “it” eta “s”. Gerra aurreko informatzaileetan “gaitues” baino ez zaigu agertu. Adizki honi dagozkion tasunak hiru dira: Nork-Nor sailekoak erabiltzea 209nn, “it” pluralgilea erabiltzea 210it eta “s” pluralgilea erabiltzea 211s.

“Ditut” adizkiak hiru aldaki dauzka: “doas” (16, % 66,67), “dotas” (4, % 16,67) eta “dotes” (4, % 16,67). Gerra aurreko informatzaileetan “doas” baino ez zaigu agertu. Aldakiotan ikusten den legez, aldagaiak multzo bitan bil daitezke; batetik “t” eta “ø” dauzkagu (212t) eta bestetik pluralaren morfemaren aurrean “a” edo “e” txertatzea (213a). “t” morfema gehiago erabiltzen da gazteen artean (sei aldiz) zaharren artean baino (bi aldiz), generoari begira ez dago alde handirik. “a” eta “e” txandaketari



dagokionez, esan behar da ez dagoela alde handirik ez adinaren arabera ez eta generoaren arabera ere.

“Ditu” adizkiari dagokionez zazpi aldaki agertu zaizkigu: “daus” (2, % 8,33), “deus” (9, % 37,5), deitu (3, % 12,5), deitus (3, % 12,5), ditu (2, % 8,33), ditus (3, % 12,5) eta “tus” (2, % 8,33). Gerra aurreko informatzaileek “deitus” eta “ditus” erabili dituzte. Gertatzen diren aldagaiak multzo batzuetan bil daitezke; batetik pluralaren morfemari dagokionez “it” eta “s” soilak erabiltzen dira eta biak batera ere bai, beroni dagozkion tasunak 214it eta 215s dira. Bestetik “s” pluralgile soila erabiltzen denean “au” eta “eu” agertzen dira: 216eu. “it” erabiltzen denean “ei” ager daiteke eta “i” ere bai: 217ei. Azkenik, informatzaile bik egiten duten moduan, hasierako silaba gal daiteke: 218σ>ø.

“Dituzte” adizkiak 8 aldaki dauzka: deitue (1, % 4,17), deitues (5, % 20,83), ditues (4, % 16,67), “dotues” (1, % 4,17), “tues” (1, % 4,17), deuries (7, % 29,17), “deures” (4, % 16,67) eta “deres” (1, % 4,17). Gerra aurreko informatzaileek “deitues” baino ez dute erabili. Aurreko adizkiarekin ikusi dugun bezala, kasu honetan ere pluralaren morfemak “it” eta “s” izan daitezke edo biak batera: 219it eta 220s. “s” erabiltzen denean, “eu” agertzen da lehen silaban eta “e” ere bai: 221eu. “it” erabiltzen denean hasierako silaban “ei”, “i” eta “o” ager dakizkiguke: 222ei, 223i eta 224o. Azkenik, hasierako silaba osoa gal daiteke 225 σ>ø.

“Ninduen” adizkiak sei aldaki dauzka: horietako 3 Nork-Nor sailekoak (“nendun” 2, % 8,33; “nindun” 2, % 8,33 eta “neuen” 5, % 20,83) eta beste 3 Nork-Nori-Nor sailekoak (“dosten” 8, % 33,33; “osten” 2, % 8,33 eta “yosten” 5, % 20,83). Orain aldiko adizkiekin ikusi dugun moduan, erabiliko dugun tasun bat Nork-Nor saileko adizkiak erabiltzea izango da: 226nn. Nork-Nor saileko adizkiekin informatzaile batzuek iragana lortzen dute orain aldiari “-en” gehituta: neu> neuen. Orduan beronen arabera iragana erabiltzea tasun bat izango da: 227iragana. Bestalde, “nend” zein “nind” morfema biak erabiltzen direnez gero, beroni dagokion tasuna 228e izango da. Gerra aurreko informatzaileetan agertu zaigun bakarra “nendun” izan da. Nork-Nori-Nor saileko adizkietan hasieran gertatzen da aldaketa, bertan “y”, “d” eta “ø” erabiltzen dira: 229y, 230d eta 231ø.

“Nituen” adizkiak 4 aldaki dauzka: “neitusen” (8, % 33,33), “nitusen” (3, % 12,5), “notusen” (3, % 12,5) eta “neusen” (10, % 41,67). Gerra aurreko informatzaileek “nitusen” eta “neusen” erabili dituzte. Aldakiak aztertuta ikus dezakegun moduan, pluralgilea “s” soila izan daiteke edo “it” eta “s” batera: 232it eta 233s. Gainerako aldaketak “ei” silabari dagozkio, berau “ei”, “i” eta “o” izan daiteke: 234ei, 235i eta 236o.

“Zituen” adizkiak 6 aldaki dauzka: “eusen” (2, % 8,33), “seusen” (7, % 29,17), “seitusen” (9, % 37,5), “situen” (1, % 4,17), “situsen” (3, % 12,5) eta “tusen” (2, % 8,33). Gerra aurreko informatzaileetan “situsen” baino ez zaigu agertu. Adizkiaren pluralgileari bagagozkio, berau “s” eta “it” izan daiteke eta baita biak batera ere: 237it eta 238s. Hasierako morfema “s” izan daiteke edo “ø”: 239s eta 240ø. “ei” silabaren diptongoa desegin daiteke “i” eginez: 241ei. Azkenik, informatzaile bik egiten duten moduan, hasierako silaba osoa gal daiteke 242 σ>ø.

“Genituen” adizkiak aldaki bi dauzka: “gendusen” (23, % 95,83) eta “dusen” (1, % 4,17). “dusen” orainari “n” gehituta sortzen da; besteetan subjektu aldaketa gertatzen da<sup>18</sup> eta morfema aldaketa: 243subj; morfema aldaketari dagokion tasunak ez du ezer ere gehitzen, izan ere, aurrekoarekin batera gertatzen da. Gerra aurreko informatzaileek “gendusen” eta “genusen” erabili dituzte.

### 3.3. *Nork-Nori-Nor*

Sail honetan bost adizki aztertu ditugu: “dizkiot”, “zidan”, “genion”, “zenion” eta “nizkion”.

“Dizkiot” adizkiak 5 aldaki dauzka: “dotzedas” (5, % 20,83), “dotzeas” (11, % 45,83), “dotzias” (1, % 4,17), “dotzetas” (3, % 12,5) eta “dotzetes” (4, % 16,67); gerra aurreko belaunaldian “dotzeas” baino ez dugu jaso. Gertatzen diren aldaketa fonetikoak hiru multzotan bil daitezke; batetik “t”, “d” eta “ø” agertzen dira: 301t, 302d eta 303ø. Bestetik, pluralaren morfemaren aurrean “a” eta “e” ager daitezke 304a. Azkenik “tze” morfemaren “e”ren ordez “i” ere ager daiteke 305i.

“Zidan” adizkiak 4 aldaki dauzka: “osten” (6, % 25), “eusten” (1, % 4,17), “yosten” (8, % 33,33) eta “dosten” (9, % 37,5). Gerra aurreko informatzaileek “osten” baino ez dute erabili. Erroan kasu gehienetan “o” agertzen da eta kasu bakar batean “eu”, beroni dagokion tasuna 306o izango da. Hasieran hiru aukera agertzen dira “y”, “d” eta “ø”: 307y, 308d eta 309ø.

“Genion” adizkiak 9 aldaki dauzka: “geuntzen” (5, % 20,83), “geuntzon” (1, % 4,17), “gontzen” (3, % 12,5), “dotzegun” (5, % 20,83), “dotzun” (1, % 4,17), “otzegun” (2, % 8,33), “yotzegun” (5, % 20,83), “yotzogun” (1, % 4,17) eta “tzun” (1, % 4,17). Gerra aurreko informatzaileek “geuntzen” eta “gontzen” erabili dituzte. Adizkietan ikusten den moduan, batzuk orain aldiari “n” gehituta eta beste aldaketa batzuk eginda sortzen dira (1). Beste batzuek, ostera, tradiziozko bidetik egiten direnek, subjektuaren leku aldaketa eta morfemaren aldaketa behar dituzte (2).

(1) 
$$\begin{array}{c} d\_o\_tze\_gu\_+n \\ | \\ \text{Subj.} \end{array}$$

(2) 
$$\begin{array}{c} g\_o\_n\_tze\_n \\ | \quad | \\ \text{Subj} \end{array}$$

Tasun hau ondoko era honetara adieraziko dugu: 310subj. Horrezaz gainera erroa ere aldatzen da, “o” eta “eu” agertzen direla; 311o, 312eu. Nori-ren morfema kasu gehienetan “tze” da, baina “tzo” ere ager daiteke; 313tze. Subjektu aldaketa gertatzen ez den kasuetan hasieran hiru aukera daukagu “y”, “d” eta “ø”: 314y, 315d eta 316ø. Azkenik kasu baten, “tzun” hasierako silaba osoa galdu da 317σ>ø.

“Zenion” adizkiak 7 aldaki dauzka: “seuntzen” (5, % 20,83), “seuntzon” (1, % 4,17), “sontzen” (3, % 12,5), “dotzesun” (5, % 20,83), “otzesun” (5, % 20,83),

<sup>18</sup> Ikusi beherago “genion” adizkiaren azalpena.

“yotzesun” (3, % 12,5) eta “otzesun” (5, % 20,83). Gerra aurreko informatzaileetan “seuntzen” baino ez zaigu agertu. Aurreko adizkiarekin bezala; kasu batzuetan subjektu eta morfema aldaketa gertatzen da: 318subjal. Erroan molde bi ager daitezke “o” eta “eu”: 319o eta 320eu. Nori morfema gehienetan “tze” den arren “tzo” ere ager daiteke: 321tze. Azkenik subjektu aldaketa egiten ez den kasuetan hiru hasiera dauzkagu “y”, “d” eta “ø”: 322y, 323d eta 324ø.

“Nizkion” adizkiak 9 aldaki dauzka: “nentzesan” (1, % 4,17), “neuntzesan” (2, % 8,33), “neutzesan” (1, % 4,17), “nontzesan” (7, % 29,17), “notzesan” (8, % 33,33), “notzesen” (1, % 4,17), “yotzesan” (1, % 4,17), “dotzetasen” (1, % 4,17) eta “dotzetesen” (2, % 8,33). Gerra aurreko informatzaileek “nontzesan” eta “neutzesan” erabili dituzte. Aurreko adizki batzuetan gertatzen den moduan, hemen ere subjektu aldaketa eta morfema aldaketa gerta daitezke ala ez: 325subjal. Subjektu aldaketa gertatzen den kasuetan -n- artizkia gal daiteke kasu batzuetan “neutzesan”; beraz, tasun bat erabiliko dugu -n- artizkia atxikitzen dela adierazteko 326n. Erroan gertatzen diren aldaketak “o”, “eu” eta “e” dira: 327o, 328eu eta 329e. Subjektu aldaketa gertatzen ez den kasuetan adizkiaren hasieran “d” eta “y” ager daitezke: 330d eta 331y. Azkenik pluralaren morfema eta amaierako “n”ren artean txertatzen den bokala “a” izaten da gehienetan, baina “e” ere izan daiteke, beroni dagokion tasuna 332a izango da.

### 3.4. Baldintzak

Baldintzen sailean aztertzeko aukeratu ditugun adizkiak hiru izan dira: “nintzateke”, “litzateke” eta “nuke”.

“Nintzateke” adizkiak 5 aldaki dauzka: “nitzeke” (1, % 4,17), “nitzike” (1, % 4,17), “nintike” (5, % 20,83), “nitike” (13, % 54,17) eta “nintzen” (4, % 16,67). Gerra aurreko informatzaileetan “neitike” eta “nintike” jaso ditugu. Lau informatzailek baldintzaren ordeaz iragana erabiltzen dute, beraz lehen tasuna baldintza erabiltzea edo ez izango da: 401baldintza. Adizkietan agertzen den bezala informatzaile bik “tz” daukaten morfemak darabiltzate (“tze” eta “tzi”) eta gainerakoek “ti” daukatenak; azken morfema hau ahalezkoa egiteko erabiltzen den bera da; honen arabera 402tz eta 403ti tasunak erabiliko ditugu. Gertatzen diren gainerako aldaketa fonetikoak biltzeko “tze” eta “tzi” txandaketaren kasuan 404tze tasuna erabiliko dugu eta “nin” eta “ni” txandaketaren kasuan 405n erabiliko dugu.

“Litzateke” adizkiak 6 aldaki dauzka: “litzeke” (1, % 4,17), “litzike” (1, % 4,17), “leitike” (2, % 8,33), “letike” (1, % 4,17), “litike” (6, % 25) eta “san” (13, % 54,17). Gerra aurreko belaunaldian “leitike” eta “litike” jaso ditugu. Adizki honen kasuan baldintzaren ordeaz iragana hamairu informatzailek erabili dute; orduan, aurreko adizkiarekin bezala tasun bat 406baldintza izango da. Informatzaile bik “tz”dun adizkiak erabiltzen dituzte, 407tz, eta gainerakoek “ti”dunak, 408ti. “tze”-“tzi” txandaketaren kasuan 409tze erabiliko dugu eta “ti”dunetan gertatzen diren “ei”, “e” eta “i”ren kasuetan 410ei, 411e eta 412i tasunak erabiliko dira.

“Nuke” adizkiak hiru aldaki dauzka: “neuke” (18, % 75), “neke” (5, % 20,83) eta “neuen” (1, % 4,17). Gerra aurrekoen artean jaso dugun bakararra “neuke” izan da. Gainerako baldintzetan gertatu den moduan, hemen ere informatzaile batek baldintzaren ordeaz iragana erabili du, hortaz 413baldintza tasuna ere erabiliko dugu. Gainerako

kasuetan erroan baino ez dira aldaketak gertatzen; informatzaile batzuek “eu” darabilte eta beste batzuek “e”; beraz erailiko dugu tasuna 414eu izango da.

### 3.5. Adizki trinkoak

Sail honetan hiru adizki trinko baino ez ditugu aztertu: “dauzkat”, “dakarte” eta “daroate”.

“Dauzkat” adizkiak 4 aldaki dauzka: “dekoas” (16, % 66,67), “deketas” (1, % 4,17), “dekotes” (3, % 12,5) eta “dekotas” (4, % 16,67). Gerra aurreko belaunaldian “dekoas” baino ez dugu jaso. Adizkietan ikusten den bezala hiru aldagai nagusi dauzkagu. Batetik, “ko”ren ordeaz “ke” erabiltzen da kasu batean 501ko; “t” desagertzen da pluralaren morfema gehitzean hamasei kasutan 502t eta, azkenik, pluralaren morfemaren aurrean “a” ager daiteke, hogeita bat kasutan, edo “e” 503a.

“Dakarte” adizkiak 3 aldaki dauzka: “dakardie” (20, % 83,33), “dakarrie” (2, % 8,33) eta “perifrastikoa” (2, % 8,33). Gerra aurreko belaunaldietan “dakardie” da jaso dugun bakarra. Informatzaile bik adizki perifrastikoa erabili dute trinkoaren ordeaz, beraz, tasun bat trinkoa erabiltzea edo ez izango da, 504trinkoa. Nork-en morfemari dagokionez aukera bi daude “die” eta “ie”, horrenbestez, tasun bat erabiliko da 505die.

“Daroate” adizkiak 6 aldaki dauzka: “daroadie” (13, % 54,17), “daroardie” (5, % 20,83), “daroarie” (2, % 8,33), “daroudie” (2, % 8,33), “doorie” (1, % 4,17) eta “perifrastikoa” (1, % 4,17). Gerra aurreko informatzaileetan “daroardie” baino ez dugu jaso. Informatzaile batek ez du adizki trinkoa erabili, berorren ordeaz perifrastikoa erabili du, orduan tasun bat 506trinkoa izango da. Bestalde, adizkietan ikusten dugun moduan, erroa aldatzen da eta Nork morfemari dagokiona ere bai. Erroan hiru mota agertu dira “roa”, “rou” eta “oo” 507roa, 508rou eta 509oo hurrenez hurren. Morfeman aldaketa bi agertzen dira “d”-“r” txandaketa 510d eta “r”ren gehitzea 511r.

### 3.6. Alokutiboa

Azpi-atal honetan alokutiboaren erabileraz baino ez dugu jardungo. Gure gerra aurreko informatzaileen datuen arabera andrazkoen alokutiboa galtzen hasita egon bazen ere, batzeko aukera izan genuen; eduki genituen informatzaile andrazko biek erabiltzen zuten bai andrazkoen artean bai eta gizonezkoekin ere. 3. taulan “etorri” aditzaren alokutiboak ematen ditugu adibidetzat.

	(m)	(n)
<b>NI</b>	nayatok	nayatona
<b>HURA</b>	yatok	yatona
<b>GU</b>	gayatosak	gayatosena
<b>HAIEK</b>	yatosak	yatosena

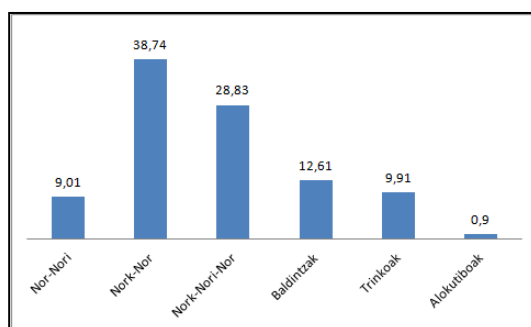
3. taula: “etorri” aditzaren alokutiboak.

Gure oraingo informatzaileei galdera bat egin zitzaien alokutiboen erabileraz eta beste batzuk adizkien formaz; azken hauek lan honetatik kanpo geratuko badira ere, esan dezagun bide batez, informatzaile batzuek “nayatok” eta “gayatosak” adizkien ordeaz “ñatok” eta “gatoxak” erabiltzen dituztela.

Erabileraren gaineko galderaren erantzunak aztertuta esan behar da ezen, andrazkoen alokutiboak guztiz galdu direla eta alokutiboa gaur egun nagusien belaunaldiko gizonezkoek baino ez dutela erabiltzen eta euron artean baino ez. Beraz, azterketa kuantitatiborako honi dagokien tasuna 601hika izango da.

#### 4. Datuen azterketa kuantitatiboa

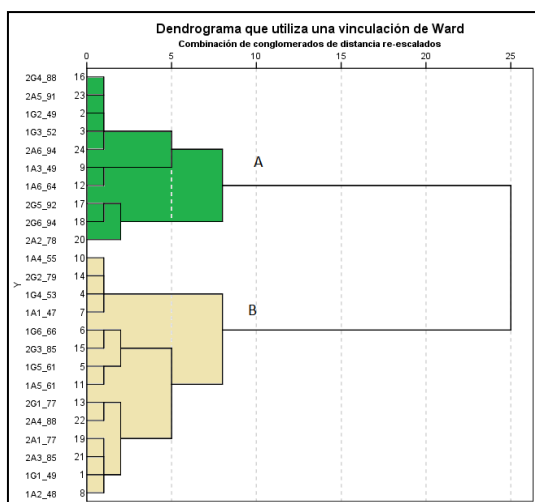
Aurreko atalean egin dugun datuen azterketa linguistikoaren ondorioz informatzaileei egin zitzaizkien 26 galderen bidez 111 tasun bildu ditugu; hauetarik 10 (% 9,01) Nor-Nori sailari dagozkio, 43 (% 38,74) Nork-Nor sailari, 32 (% 28,83) Nork-Nori-Nor sailari, 14 (% 12,61) baldintzen sailari, 11 (% 9,91) aditz trinkoei eta 1 (% 0,9) alokutiboen erabilerari (1. irudiko grafikoa).



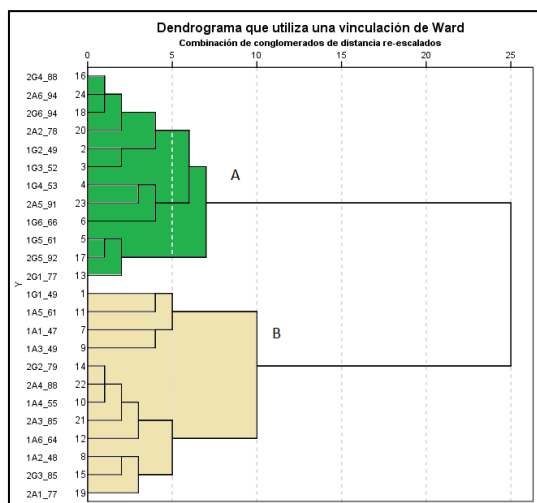
1. irudia: Aztertutako sail bakoitzaren tasunen portzentajeak.

Informatzaileak Nor-Nori saileko hamar tasunen arabera multzokatzen baditugu (2. irudiko dendograma), A multzoan hamar informatzaile kokatzen dira eta B multzoan hamalau. Adinaren arabera nagusien % 33,33 A multzoan kokatzen dira eta % 66,67 B multzoan; gazteen kasuan % 50 kokatzen da multzo bakoitzeko. Mann-Whitney U probaren arabera, adinaren alde estatistikoki esanguratsua da,  $z = -2,067$ ,  $p = 0,039$ ; nagusien batez besteko ibiltartea 15,29 da eta gazteen batez besteko ibiltartea 9,71. Generoaren arabera, gizonezkoen zein andrazkoen %41,67 A multzoan kokatzen dira eta % 58,33 B multzoan.

Nork-Nor saileko 43 tasunen arabera multzokatzen baditugu (3. irudiko dendograma), A multzoan hamabi informatzaile kokatzen dira eta B multzoan beste hamabi. Adinaren arabera nagusien %41,67 A multzoan kokatzen dira eta % 58,33 B multzoan; gazteen kasuan alderantziz gertatzen dira portzentajeak. Aldeak ez dira estatistikoki esanguratsuak. Generoaren arabera, gizonezkoen % 25 A multzoan kokatzen dira eta % 75 B multzoan; andrazkoen kasuan alderantziz gertatzen da; alde hau ez da estatistikoki esanguratsua.



2. irudia: Informatzaileen multzokatzea Nor-Nori saileko adizkien tasunen arabera.

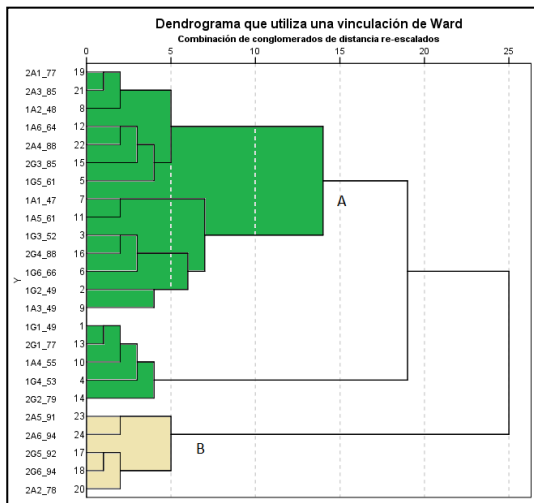


3. irudia: Informatzaileen multzokatzea Nork-Nor saileko adizkien tasunen arabera.

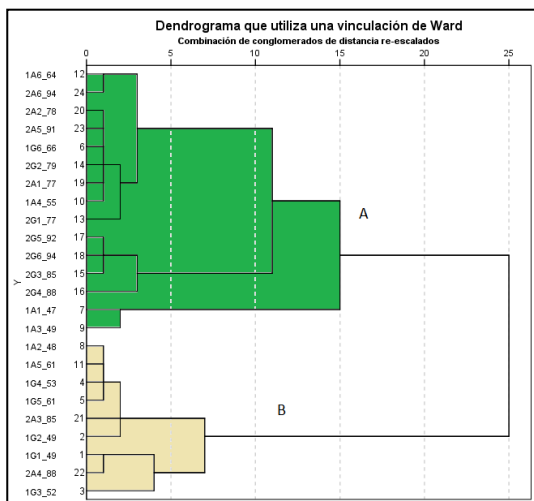
Nork-Nori-Nor saileko 32 tasunen arabera multzokatzen baditugu (4. irudiko dendograma), A multzoan hamazazpi informatzaile kokatzen dira eta B multzoan 5. Adinaren arabera nagusi guztiak A multzoan kokatzen dira; gazteen kasuan % 58,33 A multzoan kokatzen dira eta % 41,67 B multzoan. Aldea ez da estatistikoki esanguratsua. Generoaren arabera, gizonezkoen % 83,33 A multzoan kokatzen dira eta % 16,67 B multzoan; andrazkoen % 75 A multzoan kokatzen dira eta % 25 B multzoan. Aldea ez da estatistikoki esanguratsua.

Baldintzen saileko 14 tasunen arabera multzokatzen baditugu (5. irudiko dendograma), A multzoan hamabost informatzaile kokatzen dira eta B multzoan 9. Adinaren arabera nagusien % 41,67 A multzoan kokatzen dira eta % 58,33 B multzoan; gazteen kasuan % 83,33 A multzoan kokatzen dira eta % 16,67 B multzoan. Mann-Whitney U probaren arabera, adinaren aldea estatistikoki esanguratsua da,  $z = -2,420$ ,  $p = 0,016$ ; nagusien batez besteko ibiltartea 15,92 da eta gazteen batez besteko ibiltartea

9,08. Generoaren arabera, gizonzkoen % 58,33 A multzoan kokatzen dira eta % 41,67 B multzoan; andrazkoen % 66,67 A multzoan kokatzen dira eta %33,33 B multzoan. Aldea ez da estatistikoki esanguratsua.

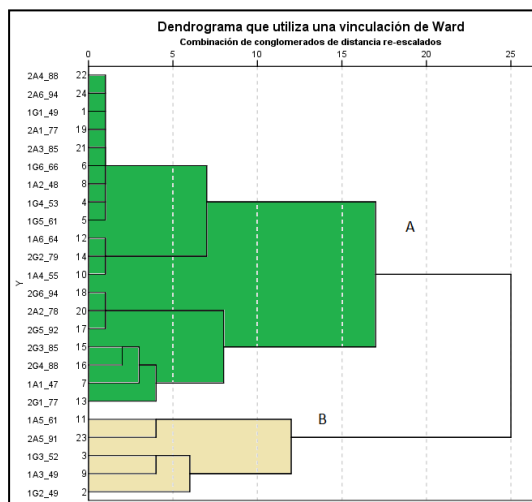


4. irudia: Informatzaileen multzokatzea Nork-Nori-Nor saileko adizkien tasunen arabera.



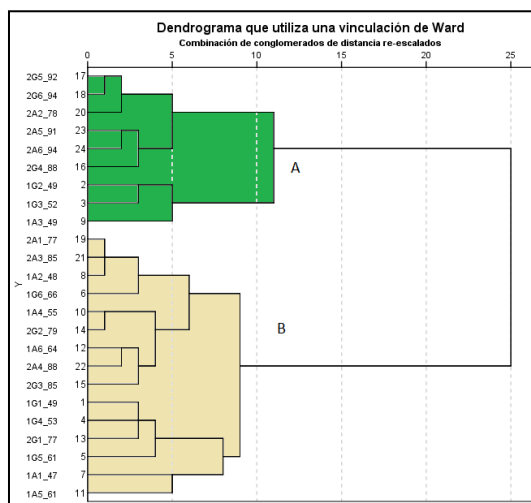
5. irudia: Informatzaileen multzokatzea baldintzen saileko adizkien tasunen arabera.

Aditz trinkoen saileko 11 tasunen arabera multzokatzen baditugu (6. irudiko dendograma), A multzoan hemeretzi informatzaile kokatzen dira eta B multzoan 5. Adinaren arabera nagusien % 66,67 A multzoan kokatzen dira eta % 33,33 B multzoan; gazteen kasuan % 97, 67 A multzoan kokatzen dira eta % 8,33 B multzoan. Aldeak ez dira estatistikoki esanguratsuak. Generoaren arabera, gizonzkoen % 83,33 A multzoan kokatzen dira eta % 16,67 B multzoan; andrazkoen % 75 A multzoan kokatzen dira eta % 25 B multzoan; alde hau ez da estatistikoki esanguratsua.



6. irudia: Informatzaileen multzokatzea adizki trinkoen saileko adizkien tasunen arabera.

Aditz arloko 111 tasunen arabera informatzaileak multzokatuz gero (7. irudiko dendograma), A multzoan bederatzi informatzaile kokatzen dira eta B multzoan 15. Adinaren arabera, nagusien % 25 A multzoan kokatzen dira eta % 75 B multzoan; gazteen kasuan % 50 kokatzen dira multzo bakoitzeko; aldea ez da estatistikoki esanguratsua. Generoaren arabera, gizonezkoen % 41,67 A multzoan kokatzen dira eta % 58,33 B multzoan, andrazkoen kasuan % 33,33 A multzoan kokatzen dira eta % 66,67 B multzoan; aldea ez da estatistikoki esanguratsua.



7. irudia: Informatzaileen multzokatzea sail guztien tasunen arabera.

Multzoen osaketa xehekiago aztertzen badugu (4. taula), ikusten dugu gunea osatzen duten informatzaileak hamabost direla; adinaren arabera hamar nagusi eta bost gazte eta generoaren arabera zazpi gizonezko eta zortzi andrazko. Honek joera argia erakusten du, hau da, nagusi gehienak multzo batean dauden artean gazte gehienak bestean kokatzen dira; beraz estatistikoki esanguratsua ez den arren, uste dugu joera argia dela.



		Nagusiak	Gazteak	Oro	Gizonezkoak	Andrazkoak	Oro
A (9)	A1 (6)	0	6	6	3	3	6
	A2 (3)	3	0	3	2	1	3
	Oro	3	6	9	5	4	9
B (15)	B1 (9)	5	4	9	3	6	9
	B2 (6)	5	1	6	4	2	6
	Oro	10	5	15	7	8	15
Orotara		13	11	24	12	12	24

4. taula: Informatzaileen multzokatzea euron adinaren eta generoaren arabera sailkatuta.

## 5. Ondorioak

Azken atal honetan aurreko bietan lortu ditugun ondoriorik nabarmenenak laburbiltzen ahaleginduko gara. Ondorioak hiru multzo nagusitan bilduko ditugu; aldaketa mota nagusiak, aldagai sozialen eragina eta informatzaileen sailkapena.

Aztertu ditugun aldaketa morfologikorik garrantzitsuenak analogiaz egiten direnak dira. Ikusi dugun bezala, joera handia dago iragan aldiko adizki tradizionalen ordeztan orain aldikoen gainean egiteko "-n" morfema gehituta. Hau Nor-Nori, Nork-Nor eta Nork-Nori-Nor sailetan ikusi dugu. Nork-Nor sailean kasu gutxi batzuetan baino ez zaigu agertu, baina Nork-Nori-Nor sailekoa nahikoa azkarra da. Sail honetan erabilerak bi ikusi ditugu; batean "y-" gehitzen da "d-"ren ordeztan eta, bestean, "d-" eusten zaio iraganera pasatzean, azken hau gazteen artean gehien zabaltzen ari dena da.

Nork-Nor saileko adizkien ordeztan Nork-Nori-Nor sailekoak joera handia dago adin talde bietan. Bestalde, sail honetan "it" pluralgilearen ordeztan "s" orokortzeko joera nabarmendu dugu; horrela bada "deitus > deus", "neitusen > neusen" eta abar gertatzen dira.

Morfemen aldakiei dagokienez, Nor-Nori sailean "-t/-te" txandaketa agertu zaigu eta Nork-Nor sailean "-ie/-e" txandaketa pluralgiletan. Nork-Nori-Nor saileko iragan aldiko adizkietan "ø", "y-" eta "d-" txandakatzen dira.

Baldintzazko adizkien ordeztan "-ko + iragana" erabiltzeko joera handia dago gazteen artean. Sail honetako baldintzak erabiltzen direnean, "ti" dunak agertzen dira gehienetan "tz" dunak oso kasu bakaretan agertu zaizkigu.

Aldaketa fonetikoeta diptongoen gauzapenei dagozkienak ditugu: "eu" > "au"; "ei" > "i" > "o". "-t" morfemaren galerari dagokiona bokalen artean, ikusi dugun bezala, gazteen artean joera dago eusteko Nork-Nor (dotas), Nork-Nori-Nor (dotzetas) eta aditz trinkoetan (dekotas). Nork-nor sailean "haiek" morfemaren "d" eta "r" txandakatzen dira (neudie/neurie; deudie/deurie). Iragan aldiko "-n" morfemaren aurrean "a" eta "e" ere txandakatzen dira. "Eroan" aditzaren erroak aldaketa asko dauzka, adizki honi dagokionez "die" morfemaren aurrean "r" bat analogiaz txertatzeko joera ere badago. Azkenik, Nor-Nori saileko adizkietan kontsonantearen aurrean "s" txertatzeko joera agertu da gazteen artean "daste", "daskus".

Azkenik, alokutiboen erabileraren galera nabarmena eta azkarra azpimarratu behar da, izan ere, andrazkoek eta belaunaldi gaztearen gizonezkoek ez dituzte erabiltzen.

Aldagai sozialen eraginari dagokionez, adina estatistikoki esanguratsua da Nor-Nori sailari, baldintza sailari eta alokutiboaren erabilerari dagokionez; azken honetan generoaren eragina ere estatistikoki esanguratsua da. Edozelan ere, eta denak batera aztertuta ikusi dugu ez adina ez generoa ez direla erabateko erabakigarriak taldeak egitean; honek argi adierazten du nahikoa nahaste handia dagoela.

Gauza bera agertu zaigu informatzaileak multzokatzean, izan ere, hamabost informatzailek osatzen dute gune nagusia eta beste bederatzi periferian egongo lirateke. Multzo nagusi horretako nagusiak gehiengoa dira. Horrenbestez esan dezakegu, joera nagusia dela gazteak aldentzeko.

## 6. Bibliografia

- Aranbarri, F. (1996). *Ermua eta Eitzako euskara*. Ermua: Ermuko Udala.
- Arretxe, J. (1994). *Basauriko euskara*. Basauri: Basauriko Udala.
- Aurrekoetxea, G. (1995). *Bizkaieraren egituraketa geolinguistikoa*. Bilbao: EHU.
- Bilbao, B. (2002). *Bermeoko euskera kresaltsua: aditza eta fonetika*. Bermeo: Bermeoko Udala.
- Boersma, P. eta Weenink, D. (2015). *Praat: doing phonetics by computer*, Version 5.4. [Computer program]. <<http://www.praat.org>>
- Errazti, A. (1994). *Iurreta elizateko euskara eta toponimia*. Iurreta: Iurretako Udala.
- Etxebarria, J.M. (1991). *Zeberio haraneko euskararen azterketa etno-linguistikoa*. Euba: Ibaizabal.
- Gaminde, I. (1983). Orozkoko aditzak. *Fontes Linguae Vasconum: Studia et Documenta*, 42/42, 37-97.
- Gaminde, I. (1984). *Aditza Bizkaieraz I, II eta III*. Iruña: Udako Euskal Unibertsitatea.
- Gaminde, I. (1985). Gaminiz inguruko aditzak. *Plentziako Azter Lanak*, 1, 315-462.
- Gaminde, I. (1989). *Foruko euskararen morfoxintaxiaz*. Forua: Foruko Udala.
- Gaminde, I. (1989b). *Leioako Euskararen Gramatikaz*. Bilbao: Leioako Udala.
- Gaminde, I. (1990). *Bilbo Berbaz Berba*. Bilbao: Bilbo Berreuskalduntzeko Koordinakundea.
- Gaminde, I. (1991). *Abadiñoko Euskaraz*. Bilba: Abadiñoko Udala.
- Gaminde, I. (1992). *Urduliz eta Gatikako Herri hizkeren Azterketa Linguistikoa*. Doktorego Tesia, Deustuko Unibertsitatea.
- Gaminde, I. (2005). *Lezama Berbarik Berba*. Lezama: Zorritzketan eta Lezamako Udala.
- Gaminde, I. (2007). *Bizkaian Zehar: Euskararen Ikuspegi Orokorra*. Bilbao: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I. eta Burgete, X. (1991). *Otxandioko Euskaraz*. Otxandio: Otxandioko Udala.
- Gaminde, I., Romero, A. eta Legarra, H. (2012). *Gramatika eta Hizkuntz Bariazioa Bermeon*. Bermeo: Bermeoko Udala eta Campos Hegaluzea.
- Hualde, J.I., Elordieta, G. eta Elordieta, A. (1994). *The Basque Dialect of Lekeitio*. Anejos del Anuario del Seminario de Filología Vasca Julio de Urquijo - Julio Urkixo Euskal Filologi Mintegiaren Urtekariaren Gehigarriak, 34, UPV/EHU.
- Iglesias, A. (2014). *Igorreko Hizkeraren Azterketa Dialektologikoa*. Doktorego Tesia, EHU/UPV.
- Laka, E., Olondo L., eta Gaminde, I. (2008). *Bermeoko gazteen euskararen aditz morfologiaz*. *Euskalingua*, 13, 27-36

- Legarra, H. (2011). *Muxikako euskararen azterketa linguistikoa*. Doktorego Tesia, Deustuko Unibertsitatea.
- Léonard, J. L. (2015). Modeling Regional Variation from EAS: Complexity and communal aggregates. In Aurrekoetxea, G., Romero, A. eta Etxebarria, A. (Eds.), *Linguistic Variation in the Basque and Education-I*, (pp. 143-169). Bilbao: UPV/EHU.
- Markaida, E., Gaminde, I. eta Markaida, B. (1993). *Sopelako Euskaraz*. Sopela: Urdulizko Euskara Taldea.
- Mugarza, P. (2006). *Mallabiko euskara*. Bilbo: Mallabiko Udala, Bizkaiko Foru Aldundia eta Eusko Jaurlaritzia.
- Roman, I. (2001). *Berrizko euskara*. Berriz: Gerediaga Elkarte.
- Salazar, B. (2001). *Ugaoko Euskara*. Bilbao: Ediciones Beta.
- Uriarte, J.A. (1995). *Fruizko Hizkera: Azterketa Linguistikoa*. Doktorego Tesia, Deustuko Unibertsitatea.
- Yrizar, P. (1991). *Morfología del verbo auxiliar vizcaino*. Liburuki 2, Bilbo: Euskaltzaindia.

## BIZKAIKO ADITZ LAGUNTZAILEEN BILAKAERAREN AZTERKETAZ<sup>19</sup>

Iñaki Gaminde, Asier Romero, Aintzane Etxebarria eta Naia Eguskiza  
Euskal Herriko Unibertsitatea

[inaki.gaminde@ehu.eus](mailto:inaki.gaminde@ehu.eus), [a.romero@ehu.eus](mailto:a.romero@ehu.eus), [aintzane.etxebarria@ehu.eus](mailto:aintzane.etxebarria@ehu.eus),  
[eguskiza.n@gmail.com](mailto:eguskiza.n@gmail.com)

### Laburpena

Hemen aurkezten dugun lan hau iazko jardunaldietan Léonard (2015) irakasleak hizkuntzen aldakortasuna sistema konplexu dinamiko moldagarritzat jotzen duen teoriarekin lotzeko ebakitako ildoan kokatzen da. Lan honen helburua da aurkeztea ze metodologia erabili den Bizkaian XX. mendearen amaieran jaiotako belaunaldiak erabiltzen dituen aditz laguntzaileen bilakaera aztertzeko. Aditz arloko aldakortasuna aztertzeko 4 adizkik osatutako galdekizun bat erabili da eta 550 informatzaileekin jaso da Bizkaiko euskararen eremu tradizionalako 57 herritan. Belaunaldi horren datuen azterketa burutzeko adizkien analisi bi egin dira; alde batetik, adizkien analisi morfologikoa egin da eta, bestetik, morfema bakoitzaren frekuentzien arabera azterketa matematikoa egin da eroldaren datuak kontuan hartuta, populazioan morfema bakoitzak eduki dezakeen balizko pisua agerian uzteko. Azkenik, belaunaldi gaztearen emaitzak gerra aurrean jaiotako belaunaldiaren datuekin erkatu dira, aldaketan joeren nondik norakoak aurreikusteko.

**Gako-hitzak:** Aditz morfologia, soziolinguistika, hizkuntzaren bilakaera

### 1. Sarrera

Jardunaldi hauen zuzuan aurkezten dugun lan honen helburu nagusia Bizkaian erabiltzen diren aditz laguntzaile batzuen bilakaeraren azterketa egiteko metodo berriak aurkeztea da. Metodo horiek lagundu diezagukete une honetako errealitatearen argazkia egiten eta baita etorkizunerako joerak aurreikusten ere. Bariazioaren azterketan dialektologia tradizionalaren metodoek eta berrikiago soziolinguistikarenek egin dituzten ekarpenak inolaz ere gutxietsi gabe, uste dugu unea heldu dela beste bide batzuk azterkatzeko.

Bariazioaren gainean egiten diren lan batzuetan ez da kontuan hartzen Euskal Herrian gertatu diren eta gertatzen ari diren aldaketa handiak, hala nola, betiko euskararen transmisioaren etena, populazio euskaldunaren nahastea, populazioaren mugimenduak, gazte euskaldun kopuruen murriztapena, euskara batuaren eragina, hiritartze prozesu azkarra, populazioaren kontzentrazioa, eta abar. XX. mendearen amaieran eta XXI. mende honen hasieran gertatu diren horiek guztiak ezin bazter utzi daitezkeelako gogorra edukita ekin diogu hemen dakarguntxo honi.

Hori guztia gomutan, uste dugu datuak jasotzerakoan informatzaile bat baino gehiago behar direla inkesta puntu bakoitzeko. Bestalde, populazioen mugimenduen eta kontzentrazioaren ondorioz jasotzen den molde bakoitzari bere “pisua” eman beharko litzaioke jakiteko zein den molde bakoitzak daukan agertzeko probabilitatea. Hemen gure azterketa metodologia aplikatzeko hartu dugun populazioaren segmentua 1986-2000 urte bitartean jaiotako gazteek osatzen dute; hurrengo atalean adieraziko diren baldintza zehatzak betetzen dituzte eta nahi izan dugu lagin adierazgarria jaso gure entsegu honetarako. Gorago aipatu dugun populazioaren kontzentrazioaren adibidetzat lan honetarako aukeratu ditugun herrien egoera aurkez dezakegu (herrien zerrenda 2.

---

<sup>19</sup> Lan hau GIU 13/23 proiektuaren barnean egin da.

atalean ikus daiteke).1986-2000 urte bitartean herri horietan jaiotako gazte kopurua 27.789koa da; EUSTATEk ematen dituen herri bakoitzeko ama hizkuntzaren arabera ehunekoak kopuru horri aplikatzen badiogu, ikusten dugu denetara 11.608 direla euskara ama hizkuntzatzat jaso duten gazteak. Gure hemengo azterketarako interesatzen zaigun gazte euskaldun populazio osoaren % 65,92 hamar herritan kontzentratzen da: Bermeo, (% 10,37), Gernika-Lumo (% 9,65), Durango (% 9,6), Mungia (% 7,32), Zornotza (% 7,12), Ondarroa (% 6,9), Lekeitio (% 5,35), Markina-Xemein (% 3,67), Elorrio (% 3,4) eta Abadiño (% 2,53).

Datuak erkatzerakoan galdera bi egin ditugu: (a) forma hutsak erkatu behar dira ala formen azterketa morfofonologikoa eginda ezaugarrien erkaketa egin behar da? (b) forma guztiek balio bera daukate ala bakoitzari daukan agertzeko probabilitatea eman behar zaio?

Lehenengo galdera azterketa linguistikoari dagokio eta guk bigarren bidea aukeratu dugu; adizkien aldakien azterketa morfofonologikoan oinarrituko gara ezaugarriak agerian uzteko. Bigarren galderaren kasuan bigarren bidea ere aukeratu dugu, azterketa morfofonologikoa agerian uzten dituen ezaugarrien probabilitateak kalkulatu ditugu. Horretarako zuhaitz diagramak egin ditugu lagin-espazioak zedarrizteko eta adizkien aldaki bakoitzaren probabilitatea kalkulatu ahal izateko.

Beraz, gure helburu orokorra gorago aipatu dugun hori izanagatik ere, helburu zehatz batzuk ere baditugu:

- (a) Gazteen adizkiak morfologikoki eta fonologikoki aztertzea
- (b) Ezaugarrien probabilitateen kalkuluak egitea
- (c) Belaunaldien arteko aldeak agerian uztea

Hizkuntza gizarte fenomeno konplexutzat jo izan da eta bariazioa, era berean, hizkuntza bizien ezaugarritzat, kontuan hartu behar da ez dagoela bariazioa hiztunik ez badago. Bariazioari hizkuntzaren barne faktoreek zein gizarte kanpo faktoreek eragiten diotenez gero, sistema konplexu dinamiko moldagarrien azterketarako erabiltzen diren hatsarre berberak erabil daitezke (gure lan hau kokatu daiteke guren artean honetan aitzindaria izan diren Léonard eta lankideen ildoan, 2015). Guretik kanpo hainbat izan dira hizkuntza eta bariazioa sistema konplexuen teorien aztertu dutenak (Beckner eta lankideak, 2011; Ellis, 2011; López Rivera, 2013; Moreno Cabrera, 2008; besteak beste). Hizkuntza sistema konplexu dinamiko moldagarrien ezaugarriak honela ematen dituzte Beckner-ek eta lankideek (2011: 2):

“Antes bien, la lengua, como sistema adaptativo complejo de uso dinámico, y su experiencia encierran las siguientes características principales: (1) el sistema incluye agentes múltiples (los hablantes de una comunidad de habla) que actúan recíprocamente los unos sobre los otros; (2) el sistema es adaptativo; es decir, el comportamiento de los hablantes se basa en sus últimas interacciones, así como en las interacciones presentes y futuras, que conjuntamente determinan su comportamiento futuro; (3) el comportamiento de un hablante se debe a la concurrencia de factores muy diversos, que van desde los mecanismos perceptivos a las motivaciones sociales; (4) las estructuras de la lengua emergen de los patrones interrelacionados de la experiencia, la interacción social y los procesos cognoscitivos”

Zilegi izan bekigu sistema hauen gainerako zehaztapen batzuk egitea. Gell-Mann-en lanari jarraikirik (1994), tasun garrantzitsua da gorago aipatu duguna, hau da, barne eta kanpo faktoreen arteko eragina:

“Lo que tienen en común todos estos procesos es la existencia de un sistema complejo adaptativo que adquiere información acerca tanto de su entorno como de la interacción entre el propio sistema y dicho entorno, identificando regularidades, condensándolas en una especie de «esquema» o modelo y actuando en el mundo real sobre la base de dicho esquema. En cada caso hay diversos esquemas en competencia, y los resultados de la acción en el mundo real influyen de modo retroactivo en dicha competencia.”

Sistema konplexuak aztertzean, eta ondorioz hizkuntza haietakotzat jota aztertzean, agirian ez dauden egiturak azaleratu egiten direla nabarmentzen da bereziki (Léonard eta lank. 2015). Gure hemengo aukera metodologikoaren xedea ez da besterik azpiko egitura horiek azaleratzea baino. Hala ere, kontuan hartu behar da probabilitateez jarduten dugunean beti geratzen dela indeterminazio marjina bat, izan ere, ezin dugu zehazki neurtu hitzunek egiten duten guztia. Gell-Mann-ek (1994) honela adierazten du neurgarritasun eza:

“Dado que nada puede medirse con una precisión absoluta, el caos da origen a una indeterminación efectiva en el nivel clásico que se superpone a la indeterminación cuántica.”

Gure lan hau lau ataletan banatuta aurkezten dugu; sarrera honen osteko atalean corpusa eta metodologiak jardungo dugu, hirugarren atalean gazteen datuen azterketa aurkeztuko da, laugarren atalean belaunaldi bien datuak erkatuko ditugu, azkenik, bosgarren atalean ondorioak laburbilduko ditugu eta iradokizun batzuk aurkeztuko ditugu hurrengo lanetarako.

## 2. Corpusa eta metodologia

Atal honetan corpusa osatzeko erabili dugun metodologia azalduko dugu batetik eta, bestetik, datuak prozesatzeko erabili duguna.

Gure corpusean 19 adizkiren gaineko informazioa jaso badugu ere, hemen lau baino ez ditugu landuko; euskara estandarren adizkiak “*diet*”, “*zidan*”, “*genion*” eta “*zenion*” dira. Denetara 2230 erantzun jaso ditugu. Informatzaileak XX. mendean 1986 eta 2000 urte bitartean jaiotako neska-mutilak izan dira. Denek euskara ama hizkuntzatzat daukate eta Bizkaian egiten den barietate tradizionalen bat jaso dute familia transmisioaren bidez. Datu horietako batzuk grabatuta jaso ditugun arren (240) gehienak idatziz jaso ditugu, interneten bidez edo esku-orrien bidez. Inkesta batzuk institutuetan edo ikastetxeetan burutu dira (Mungia, Bermeo, Gernika, Markina, Igorre eta Durango) eta beste batzuk ikasleen eta ikasle ohien bidez egin dira<sup>20</sup>. Idatziz jaso diren datuek badituzte arrisku batzuk, izan ere, informatzaileek ez dakite idazten euren barietateetan eta sarritan kontsultak egin ditugun arren, gauza batzuk ilun geratu dira. Arazorik nabarmenena, hemengo lanari dagokionez, afrikatuen bereizketa ezarena da; azkenean “tz” idaztea ebatzi dugu kasu guztietan afrikatu bat erabiltzen dela adierazteko, baina badakigu kasu askotan “tx” erabiltzen dela. Beste lan batzuetan

<sup>20</sup> Bihoazkie gure eskerrik beroenak lagundu diguten guztiei.

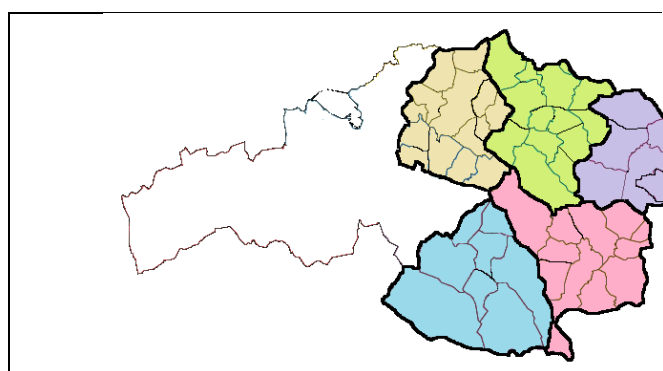
aipatu den bezala, Bizkaiko barietate askotan genero markatzailatzat erabiltze da frikari bien erabilera (Gaminde, 2010, Gaminde eta lank., 2012, Ensunza, 2015).

Inkestak herri gehiagotan jaso ditugun arren, azkenean 57 herritako datuak baino ez ditugu hemen erabiliko, izan ere, baztertu ditugun herri batzuetan inkesta bakarra jaso dugu (Urduliz, Ispaster, Mendexa, Murueta, Zaratamo, e.a.). Herri guztiak Bizkaiko euskararen eremu tradizionalan kokatuta agertzen dira; esan dezagun bidenabar, eremu tradizional hori hemen murriztuta agertzen dela azken urteotan hainbat herritan utzi zaiolako bertako euskara transmititzeari. Honek ez du esan nahi hemen aurkezten dugun eremutik kanpo bertako euskarari eusten dion euskaldun gazterik ez dagoenik, guk geuk batu ditugu banaka batzuk Urdulizen, Galdakaon eta Zaratamon, arazoa da hemen jasotzen dugun eremutik kanpo oso zaila izatea datuak lortzea.

Jarraian herriak bost eskualdetan banatuta aurkezten ditugu: (1) Ipar mendebaldea, (2) Busturialdea, (3) Lea-Artibai, (4) Hego mendebaldea eta (5) Durangaldea. 1. taulan eskualde bakoitzeko (Esk) zein herritan batu ditugun datuak (HK), zenbat informatzaile izan dugun eskualdeko (In.) eta herriak zeintzuk diren eta bakoitzeko zenbat informatzaile izan dugun adierazten da (1. irudiko mapa).

sk	K	Inf	Herriak
1)	4	89	Arrieta (2), Bakio (6), Fruiz (3), Gamiz-Fika (3), Gatika (3), Jatabe (2), Larrabetzu (12), Laukiz (3), Lezama (9), Loiu (2), Meñaka (2), Mungia (31), Sondika (3) eta Zamudio (8).
2)	6	164	Arratzu (2), Ajangiz (4), Busturia (3), Arteaga (13), Bermeo (31), Ea (12) Elantxobe (2), Forua (5), Gernika (65), Ibarangelu (4), Kortezubi (2), Mendata (5), Morga (2), Mundaka (4), Muxika (8) eta Nabarniz (2).
3)		63	Alesti (5), Berriatua (5), Etxebarria (2), Lekeitio (11) Markina (13), Munitibar (13) eta Ondarroa (14)
4)		69	Areatza (6), Bedia (2), Dima (13), Igorre (20), Lemoa (9), Orozko (4), Zeanuri (5) eta Zeberio (10)
5)	2	191	Abadiño (16), Atxondo (2), Berriz (20), Durango (83), Elorrio (29), Ermua (8), Iurreta (6), Mallabia (3), Mañaria (2), Otxandio (4), Zaldibar (8) eta Zornotza (10)

1. taula: Herri bakoitzeko eduki ditugun informatzaile kopuruak eskualdeka.



1. irudia: Eskualde bakoitzeko azterketarako aukeratu diren herriak.

Gerra aurrean jaiotako informatzaileek osatzen dute zaharren belaunaldia, euren corpora osatzeko Gamindek egindako lana (1985) hartu da oinarritzat, Arratzuko datuak falta izan direnez gero, Yrizarren (1992) datuekin osatu dira.

Datuak prozesatzeko metodologiari dagokionez, adizki estandarren aldakien azterketa morfofonologikoa izan da gure lanaren abiapuntua, hau egiteko, adizkien morfemen hurrenkeraren arabera sortzen diren txantiloiak (Artiagoitia, 2000; Gaminde eta lank. 2012) oinarritzat hartu ditugu eta gero adizkiak eraikitzeke erabili behar diren eragiketak hierarkikoki antolatu ditugu. Eragiketa horiek ordenatuta burutu behar dira; batzuek zuhaitzaren urkuluetan bide paraleloak sorrarazten dituzte eta beste batzuek aukera ematen dute ezaugarri bakoitzaren aldaki bat edo beste txertatzeko euron probabilitateen arabera.

Zuhaitz-diagramen bidez adizkirik adizki gertatzen diren ezaugarriak antolatzen dira eta lagin-espazioak zedarrizten dira. Zuhaitz-diagrametan urkulu bakoitzaren adarren probabilitateak adierazten dira. Informazio honekin, lan honetatik kanpo geratzen den arren, aldaki bakoitzaren probabilitateak kalkula daitezke. Esate baterako “dotziet” adizkiaren probabilitatea ondoko formularen arabera kalkulatu litzateke:

$$P (“1MH” \cap “do” \cap “tzie”)$$

“1MH” lehen morfema hurrenkera erabiltzeari dagokion ezaugarria da, “do” aditz oina da eta “tzie” “nori” morfemari dagokio. Kalkulua ondoko era honetara egiten da.

$$P (“1MH”) * P (“do”) * P (“tzie”)$$

Termino bakoitzaren probabilitateak 3.1. azpi-atalean ikus daitezke; behin kalkulua eginda “dotziet” adizkiaren agerpen probabilitatea 0,564koa da.

Belaunaldien arteko datuak erkatzean, azterketa morfofonologikoan oinarrituko gara eta erkaketa hiru mailatan egingo da. Batetik, ezaugarri berberak belaunaldi bietan agertzen diren ikusiko dugu, azterketa honen bidez jakin ahal izango dugu zer galdu den eta ze berrikuntza dagoen. Beste alde batetik, ezaugarrien agerpen probabilitateen bidez erreskada hierarkikoak egingo ditugu erkatzeko, hau da, lau ezaugarri hartzen baditugu {a, b, c, d} eta bakoitzaren probabilitateak honako hauek badira: a (0,10), b (0,40), c (0,20) eta d (0,30), probabilitate horien arabera ordena ditzakegu erreskada hierarkikoa eginez<sup>21</sup>:

$$b \gg d \gg c \gg a$$

Azkenik, belaunaldi biek ezaugarri berberak eta erreskada hierarkiko berbera daukatenean ezaugarri bakoitzaren probabilitateen alde edo alde eza emango digute zein den belaunaldien arteko alde; aurreko adibidearekin jarraituta honako hau eduki genezake:

	b>>d >>c >>a			
<b>Zaharrak</b>	0,40	0,30	0,20	0,10
<b>Gazteak</b>	0,35	0,28	0,27	0,10

### 3. Gazteen adizkien azterketa

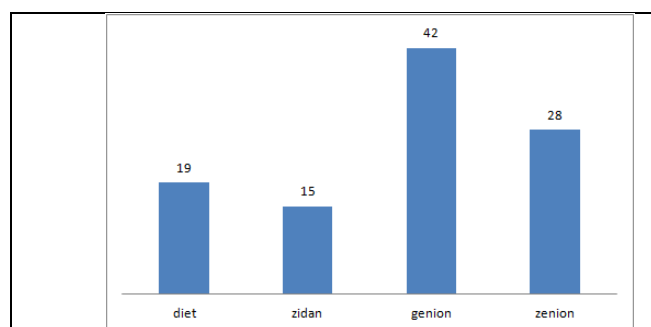
<sup>21</sup>>> ikurraren bidez osagai bat bestearen gainean dagoela adierazi nahi da, hau da, batak besteak baino portzentaje handiagoa dauka; // erabiltzen dugunean osagai biak maila berekoak direla adierazi nahi da.



Gure azterketa egin ahal izateko inkestan jaso ditugun Nork-Nori-Nor saileko lau adizkiren emaitzak aurkeztuko ditugu. Sail honetan jaso ditugun adizkiak euskara estandarrean “*diet*”, “*zidan*”, “*genion*” eta “*zenion*” izan dira. 2. taulan bakoitzeko jasotako adizki kopuruak eta aldakiak erakusten dira. Bertan ikusten den moduan, aldaki kopuru handiak agertzen dira kasu guztietan, “*genion*” eta “*zenion*” adizkiak aldaki kopururik handienak dauzkatenak dira (2. irudiko grafikoa).

Adizkia	Kopuruak	Aldakiak
<b>diet</b>	550	19
<b>zidan</b>	554	15
<b>genion</b>	563	42
<b>zenion</b>	563	28
<b>Guztira</b>	2230	104

2. taula: Adizki bakoitzeko datu eta aldaki kopuruak.



2. irudia: Aldaki kopuruak adizki bakoitzeko.

Aldakien azterketa morfofonologikoa eta matematikoa hobeto aurkezteko adizkiak banan-banan emango ditugu hurrengo azpi-ataletan.

### 3.1. Diet

Adizki honekin 550 erantzun jaso ditugu inkestetan; denetara 19 aldaki agertu zaizkigu. 3. taulan aldaki bakoitzarekin batera euron kopuruak eta ehunekoak ematen ditugu.

Adizkia	Kopurua	%	Adizkia	Kopurua	%	Adizkia	Kopurua	%
dotziet	310	56,36	dotzeat	11	2,00	dotziat	2	0,36
tzatie	70	12,73	deutziet	9	1,64	deutzet	2	0,36
tziet	55	10,00	tzate	8	1,45	dotzete	1	0,18
dotzatie	27	4,91	tzetie	3	0,55	deutzete	1	0,18
dotzet	19	3,45	otzet	3	0,55	tzetia	1	0,18
tzet	13	2,36	tzete	3	0,55	dutziet	1	0,18
dotzate	11	2,00						

3. taula: “*diet*” adizki estandarren aldakien kopuruak eta portzentajeak.

3. taulan ikus daitekeen bezala aldaki batzuek oso ehuneko txikiak dauzkate, %5etik gora dauden aldakiak “*dotziet*”, “*tzatie*” eta “*tziet*” baino ez dira.

Corpuseko adizkien ezaugarri morfofonologikoak aztertzerakoan adizki batzuk 4. taulakomorfema hurrenkeraren arabera egiten direla ikusten dugu.

d	O	tzie	t
oraina	erroa	Nori	Nork
4. taula: Morfemen hurrenkera (1. txantiloia).			

Kasu batzuetan, ostera, “nori” morfema leku bitan agertzen da, 1. taulakomorfema hurrenkera adierazten den legez.

d	O	tza	t	ie
oraina	erroa	Nori <sub>1</sub>	Nork	Nori <sub>2</sub>
5. taula: Morfemen hurrenkera (2. txantiloia).				

4. taulako morfema hurrenkera arabera eraikitzen izan diren adizkiak 425 izan dira (p: 0,773) eta 5. taulako morfema hurrenkera arabera eraikitzen izan direnak 125 (p: 0,227). Diagrama zuhaitz batean morfema hurrenkera motak lehen urkuluaren adar biak osatuko lituzke.

1. morfema hurrenkera bidea aukeratuz gero, lehen eragiketa “nork” morfema txertatzea izan liteke, berau kasu guztietan “t” denez gero, ez dago bariaziorik eta bere probabilitatea beti da bat. Hurrengo eragiketa “nori” morfema txertatzeari dagokio, beronek hiru aldaki dauzka: “tzie” (p: 0.882), “tze” (p: 0,087) eta “tzea/tzia” (p: 0,031). Azken eragiketa adizkien oinari dagokio, lau aukera agertzen zaizkigu denetara “do” (0,812), “deu” (0,026), “du” (0,002) eta “Ø” (0,160).

1. morfema hurrenkera bidea aukeratutakoan, hiru eragiketa eginda nahikoa da adizki guztiak eraikitzeko, 6. taulan erakusten den bezala.

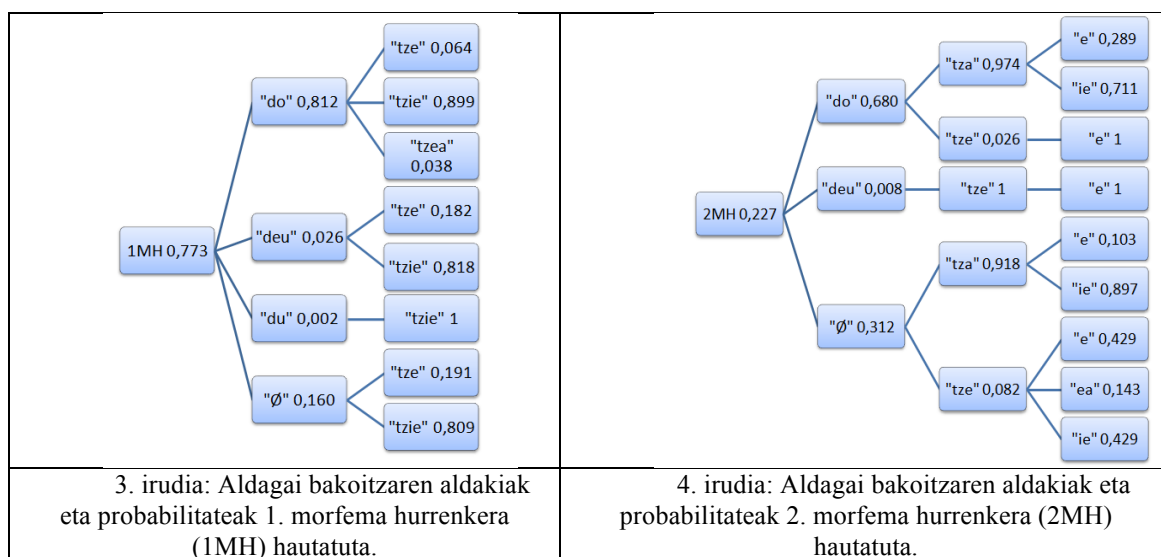
Oina	Nori	Nork
do	“tzie”	t
deu	“tze”	
du	“tzea/tzia”	
Ø		
6. taula: Eragiketa bakoitzeko agertzen diren morfemen aldakiak, 1 morfema hurrenkera arabera.		

2. morfema hurrenkera erabiltzen denean, “nori” morfema gunen bitan txertatzen da; lehen txertaketan aukera bi daude, “tza” (p: 0,928) eta “tze” (p: 0,072) eta bigarren txertaketan hiru, “ie” (p: 0,800), “e” (p: 0,192) eta “ea/ia” (p: 0,008). Hurrengo eragiketa adizkien oinari dagokiona da eta hemen hiru aukera dauzkagu, “do” (p: 0,312), “deu” (p: 0,008) eta “Ø” (p: 0,680).

2. morfema hurrenkera aukeratutakoan, lau eragiketa eginda nahikoa da adizki guztiak eraikitzeko, 7. taulan erakusten den bezala.

Oina	Nori <sub>1</sub>	Nork	Nori <sub>2</sub>
do	“tza”	t	“ie”
deu	“tze”		“e”
Ø			“ea/ia”
7. taula: Eragiketa bakoitzeko agertzen diren morfemen aldakiak, 2 morfema hurrenkera arabera.			

3. eta 4. irudietan morfema hurrenkera bakoitzaren osteko aldagaien arabera lagin-espazio biak ematen ditugu euren probabilitateekin batera.



### 3.2. Zidan

Adizki honekin 554 erantzun jaso ditugu; denetara 15 aldaki agertu zaizkigu. 8. taulan aldaki bakoitzarekin batera euren kopuruak eta ehunekoak agertzen dira.

Adizkia	Kopurua	%	Adizkia	Kopurua	%	Adizkia	Kopurua	%
dosten	168	32,24	sostan	35	4,61	eusten	8	1,47
osten	77	10,77	yosten	33	3,57	naben	3	2,21
sosten	75	12,64	dostan	23	4,64	deusten	1	0,42
sten	66	9,32	eustan	13	5,10	sausten	1	0,16
ostan	40	7,64	stan	10	5,10	usten	1	0,11

8. taula: “zidan” adizki estandarren aldakien kopuruak eta portzentajeak.

Adizkiak aztertzerakoan ikusten dugun lehenbiziko arazoa da “naben” adizkiaren kasuan egiten den paradigma aldaketa, hots, Nork-Nori-Nor sailekoen ordez Nork-Nor sailekoa erabiltzen da. Algoritmoaren lehen eragiketa berau da eta honen arabera ± PA (paradigma aldaketa) dugu; +PAren probabilitatea 0,005koa da eta -PAren probabilitatea 0,995koa. Corpuseko gainerako adizkien ezaugarri morfologikoak aztertzerakoan aldakiak 9. taulako morfema hurrenkerearen arabera egiten direla ikusten dugu.

Ø	EU	sta	n
iragana	erroa	Nori	iragana

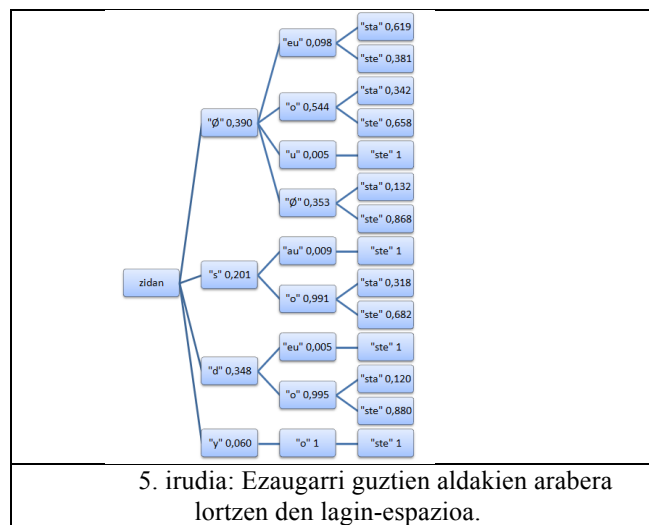
9. taula: Morfemen hurrenkera.

Paradigma aldaketa gertatzen ez den kasuetan hurrengo eragiketa iragan morfema txertatzea da, dauden aukerak lau dira: “Ø” (p: 0,390), “s” (p: 0,201), “d” (p: 0,348) eta

“y” (p: 0,060). Erroari dagokionez, beste bost aukera daukagu: “Ø” (p: 0,138), “au” (p: 0,002), “eu” (p: 0,040), “o” (p: 0,819) eta “u” (p: 0,002). “Nori” morfema, azken “-n”ren aurrean dagoenean, era bitara ager daiteke, “sta” (p: 0,220) eta “ste” (p: 0,780). Ezaugarri hau txertatzean eragiketa guztiak amaitzen dira eta lagin-espazioa guztiz zedarritzen da. Morfema hurrenkeran oinarrituta, 10. taulan agertzen diren aukerak daude (lagin-espazioa eta adarren probabilitateak 5. irudian agertzen dira).

iragana	erroa	Nori	iragana
“Ø”	“Ø”	“sta”	n
“s”	“au”	“ste”	
“d”	“eu”		
“y”	“o”		
	“u”		

10. taula: Morfema mota bakoitzaren aldakiak.



### 3.3. Genion

Adizki honekin 562 erantzun jaso ditugu; denetara 41 aldaki agertu zaizkigu. 11. taulan aldaki bakoitzaren kopuruak eta portzentajeak ematen ditugu. Taulan ikusten den moduan, aldaki guztietatik %5etik gora zortzi baino ez dira agertzen.

Adizkia	Kopurua	%	Adizkia	Kopurua	%	Adizkia	Kopurua	%
tzagun	99	17,58	geuntzon	9	1,60	giñutzan	1	0,18
dotzegun	62	11,01	gauntzen	8	1,42	gontzan	1	0,18
otzegun	55	9,77	geuntzan	7	1,24	gontzon	1	0,18
sotzegun	54	9,59	tzeuen	7	1,24	sautzegun	1	0,18
dotzagun	39	6,93	gontzen	6	1,07	sotzekun	1	0,18
geuntzen	38	6,75	genotzan	3	0,53	sotzogun	1	0,18
tzegun	34	6,04	gentzen	3	0,53	sotzoun	1	0,18
yotzegun	29	5,15	tzauen	3	0,53	tzakun	1	0,18
sotzagun	24	4,26	otzaun	2	0,36	tzeun	1	0,18
otzagun	17	3,02	dotzeun	1	0,18	tzogun	1	0,18
gentzan	14	2,49	gautzen	1	0,18	utzeun	1	0,18
eutzaun	11	1,95	geneutzan	1	0,18	yotzogun	1	0,18
gentzon	11	1,95	genotzen	1	0,18	sotzaun	1	0,18
genotzon	9	1,60	gintzon	1	0,18			

11. taula: “genion” adizki estandarren aldakien kopuruak eta ehunekoak.

Aldaki horiek guztiak eragiten dituzten aldagaiak aztertzen hasteko molde tradizional baten morfema hurrenkerarekin hasiko gara. Horretarako “geuntzen” adizkiaren morfemen hurrenkera aurkezten dugu 12. taulan abiapuntutzat.

g	EU	n	tze	n
Nork	erroa	Nork	Nori	iragana
12. taula: Morfemen hurrenkera subjektuaren leku aldaketa barik.				

11. taulako hainbat adizkitan ikus daitekeen bezala, subjektuaren morfema hasieran egon beharrean amaieran agertzen da iraganaren “-n”ren aurrean (13. taula).

Ø	O	tze	gu	n
iragana	erroa	Nori	Nork	iragana
13. taula: Morfemen hurrenkera subjektuaren leku aldaketarekin.				

Algoritmoaren lehen eragiketan subjektuaren leku aldaketa gertatzen den ala ez zehaztu behar da; subjektuaren leku aldaketaren probabilitatea 0,795ekoa da eta hain bakoarena 0,205ekoa. Lehen urkulketa honetatik aurrera bide paralelo bi segitu behar dira; bata, molde tradizionalerako legokiekeena, hau da, subjektuaren leku aldaketa ez dagoenean eta, bestea, subjektuaren leku aldaketa daukaten moldeei legokiekeena.

Subjektuaren leku aldaketa ez daukaten moldeetan aztertuko dugun lehen aldagaia “nork” morfema izango da; datuetan ikusten den bezala aditz erroarekin amalgamatuta ager daiteke “gEUn” modukoetan legez (p: 0,943) edo aditz erroaren aurretik “genEU” (p: 0,057). Bigarren aukera hau baliteke euskara estandarren eraginez sortu izana; hala ere, Elorrioren zaharrek ere “giñutzan” erabiltzen izan dute (Gaminde, 1985; Agirrebeña eta Elexpuru, 2014). Amalgamatuta agertzen ez direnean, “nork” morfemari dagokionez, hiru aukera daude “g\_\_” (p: 0,063), “gen\_\_” (p: 0,875) eta “gin\_\_” (p: 0,063). Hurrengo eragiketa aditz erroaren txertaketari dagokio, hemen sei aukera daude: “eu” (p: 0,478), “au” (p: 0,078), “o” (p: 0,183), “e” (p: 0,243), “i” (p: 0,009) eta “u” (0,009). Azken eragiketa “nori” morfemari dagokiona da; hiru mota agertzen dira, “tza” (p: 0,235), “tze” (p: 0,496) eta “tzo” (p: 0,270).

Lehen morfema hurrenkeraren araberrako aukerak 14. taulan erakusten direnak dira. Lehen eragiketa ±Amal. izango litzateke, hau da, “nork” morfema eta erroa amalgamatzen direnez zehaztea.

+Amal		-Amal			
Nork	erroa	Nork	erroa	Nori	iragana
g_n	“au” “eu” “e” “i” “o”	g_ gen_ gin_	“au” “eu” “o” “u”	“tza” “tze” “tzo”	n
14. taula: Subjektuaren leku aldaketa egin barik eragiketetan erabiltzen diren morfema bakoitzaren aldakiak.					

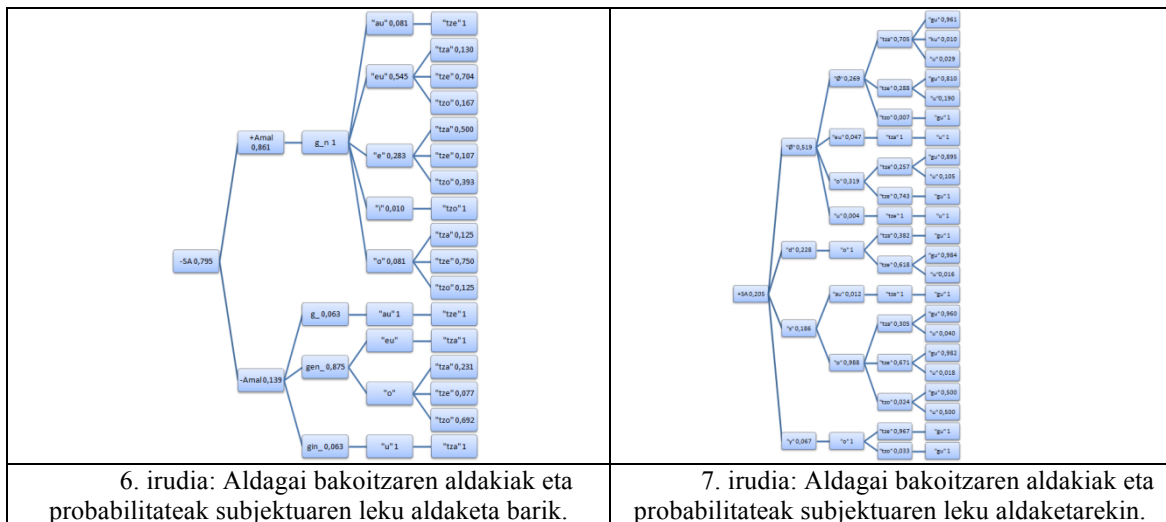
Subjektuaren leku aldaketa gertatzen den kasuetan (p: 0,795), lehen eragiketa iragan morfemaren txertaketari dagokio. Beroni dagokionez lau aukera daude: “Ø” (p:

0,519), “d” (p: 0,228), “s” (p: 0,186) eta “y” (p: 0,067). Hurrengo eragiketa erroari dagokio, hemen bost aukera daude “Ø” (p: 0,327), “au” (p: 0,002), “eu” (p: 0,025), “o” (p: 0,644) eta “u” (p: 0,002). Hurrengo eragiketa “nori” morfema txertatzeari dagokio, hemen hiru aukera daude: “tza” (p: 0,441), “tze” (p: 0,550) eta “tzo” (p: 0,009). Azkenik, azken eragiketa “nork” morfemaren txertaketari dagokio, honetan hiru aukera daude: “gu” (p: 0,933), “ku” (p: “0,004) eta “u” (p: 0,063). Bigarren morfema hurrenkeraren arabera aukerak dira 15. taulan erakusten direnak.

iragana	erroa	Nori	Nork	iragana
“Ø”	“eu”	“tza”	“gu”	n
“d”	“o”	“tze”	“ku”	
“s”	“u”	“tzo”	“u”	
“y”	“au”			
	“Ø”			

15. taula: Subjektuaren leku aldaketa eginda eragiketetan erabiltzen diren morfema bakoitzaren aldakiak.

6. eta 7. irudietan subjektuaren leku aldaketarekin eta aldaketa barik gainerako aldagaien aldakien arabera sortzen diren lagin-espazioak eta adar bakoitzaren probabilitateak erakusten dira.



Adizkia	Kopurua	Ehunekoak	Adizkia	Kopurua	Ehunekoak	Adizkia	Kopurua	Ehunekoak
tzasun	109	19,36	eutzasun	13	2,31	siñutzan	2	0,36
dotzesun	69	12,26	sentzan	9	1,60	eutzesun	1	0,18
sotzesun	65	11,55	sauntzen	7	1,24	sautzen	1	0,18
otzesun	59	10,48	sentzon	5	0,89	sautzesun	1	0,18
tzesun	54	9,59	senotzon	4	0,71	seneutzan	1	0,18
yotzesun	34	6,04	seuntzan	4	0,71	senotzan	1	0,18
dotzasun	33	5,86	sontzen	4	0,71	setzan	1	0,18
otzasun	30	5,33	sentzen	2	0,36	seutzasun	1	0,18
sotzasun	28	4,97	seuntzon	2	0,36	utzesun	1	0,18
seuntzen	22	3,91						

16. taula: “zenion” adizki estandararen aldakien kopuruak eta ehunekoak.

### 3.4. Zenion

Adizki honen kasuan 563 erantzun jaso ditugu; denetara 28 aldaki agertu zaizkigu. 16. taulan aldaki bakoitzaren kopuruak eta portzentajeak ematen ditugu. Taulan ikusten den moduan, aldaki guztietatik % 5etik gora zortzi baino ez dira agertzen.

“genion” adizkiarekin egin dugun bezala, lehenbiziko eragiketa subjektuaren leku aldaketari dagokiona da. Beronen bidez lehen urkulketa egingo dugu eta morfema hurrenkera bi egitendira abiapuntutzat hartzeko; bata, (17. taula) subjektuaren leku aldaketa bakoa eta, bestea, (18. taula) subjektuaren leku aldaketarekin.

s	EU	n	tze	n
Nork	erroa	Nork	Nori	iragana
17. taula: Morfemen hurrenkera subjektuaren leku aldaketa barik.				

Ø	EU	tze	su	n
iragana	erroa	Nori	Nork	iragana
18. taula: Morfemen hurrenkera subjektuaren leku aldaketarekin.				

Beraz, algoritmoaren lehen eragiketa subjektuaren leku aldaketari dagokiona da ( $\pm SA$ ). Subjektuaren leku aldaketaren probabilitatea 0,885ekoa da eta subjektu aldaketa bakoarena 0,115ekoa. “genion” adizkiarekin bezala, lehen urkulketa honetatik aurrera bide paralelo bi segitu behar dira; bata, molde tradizionalei dagokiena, hau da, subjektuaren leku aldaketa ez dagoenean eta, bestea, subjektuaren leku aldaketa daukaten moldeei dagokiena.

Subjektuaren leku aldaketa ez daukaten moldeetan hurrengo eragiketa “nork” morfemaren eta erroaren arteko amalgamatzeari edo ez amalgamatzeari ( $\pm Amal$ ) dagokio. “nork” morfema eta erroa amalgamatuta ager daitezke “sEU<sub>n</sub>” kasuan bezala (p: 0,846) edo amalgamatu barik (p: 0,154). Amalgamatuta agertzen ez direnean “nork” morfemari dagokionez, hiru aukera daude “s\_” (p: 0,2), “sen\_” (p: 0,6) eta “sin\_” (p: 0,2). Hurrengo eragiketa aditz erroaren txertaketari dagokio; hemen bost aukera agertzen dira “eu” (p: 0,446), “au” (p: 0,123), “o” (p: 0,138), “e” (p: 0,262) eta “u” (p: 0,031). Azken eragiketa “nori” morfemari dagokiona da; hiru mota agertzen dira, “tza” (p: 0,277), “tze” (p: 0,554) eta “tzo” (p: 0,169).

Lehen morfema hurrenkerearen araberako aukerak dira 19. taulan erakusten direnak. Lehen eragiketa  $\pm Amal$ . izango litzateke, hau da, “nork” morfema eta erroa amalgamatzen direnez zehaztea.

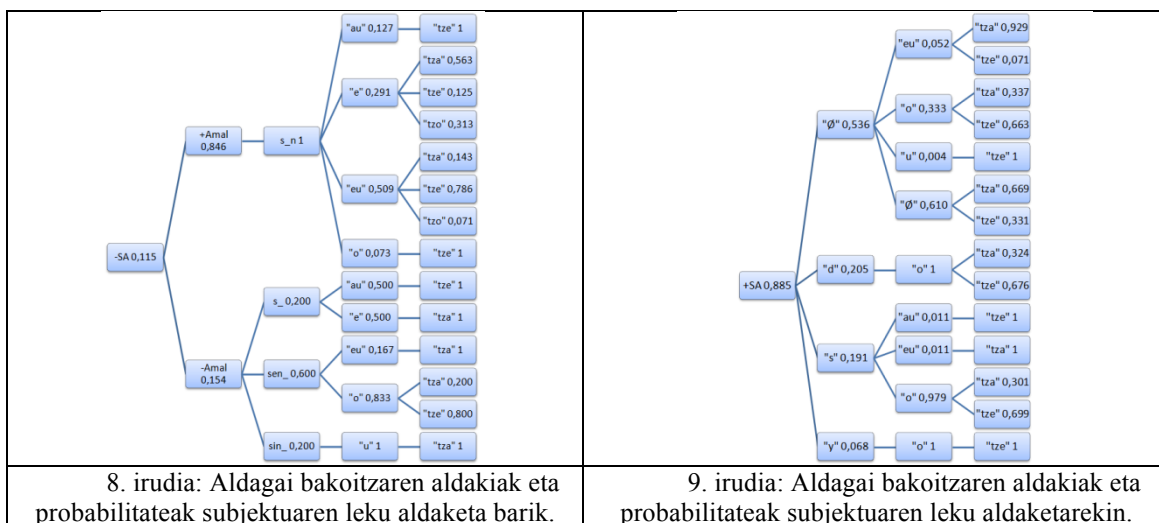
+Amal		-Amal			
Nork	erroa	Nork	erroa	Nori	iragana
s_n	“au” “eu” “e” “o”	s_ sen_ sin_	“eu” “au” “e” “o” “u”	“tza” “tze” “tzo”	n
19. taula: Subjektuaren leku aldaketa egin barik eragiketetan erabiltzen diren morfema bakoitzaren aldakiak.					

Subjektuaren leku aldaketa gertatzen den kasuetan (p: 0,885), lehen eragiketa iragan morfemaren txertaketari dagokio. Honetan lau aukera daude: “Ø” (p: 0,536), “d” (p: 0,205), “s” (p: 0,191) eta “y” (p: 0,068). Hurrengo eragiketa erroari dagokio, hemen bost aukera daude “Ø” (p: 0,327), “au” (p: 0,002), “eu” (p: 0,030), “o” (p: 0,639) eta “u” (p: 0,002). Azken eragiketa “nori” morfema txertatzeari dagokio, hemen aukera bi daude: “tza” (p: 0,43) eta “tze” (p: 0,57). Bigarren morfema hurrenkeraren arabera aukerak dira 20. taulan erakusten direnak.

iragana	erroa	Nori	Nork	iragana
“Ø”	“eu”	“tza”	su	n
“d”	“o”	“tze”		
“s”	“u”			
“y”	“au”			
	“Ø”			

20. taula: Subjektuaren leku aldaketa eginda eragiketetan erabiltzen diren morfema bakoitzaren aldakiak.

8. eta 9. irudietan subjektuaren leku aldaketarekin eta aldaketa barik gainerako aldagaien aldakien arabera sortzen diren lagin-espazioak eta adar bakoitzaren probabilitateak erakusten dira.



#### 4. Belaunaldien arteko aldeak

Azpi-atal honetan belaunaldien arteko datuen erkaketa egingo dugu. Lehenengo eta behin gerra aurrean jaiotako informatzaileen datuak aurreko azpi-atalean egin dugun sistema berberaren arabera aztertu ditugu erkaketak burutu ahal izateko. 21. taulan adizki estandarren datu kopuruak (kopuruak), aldaki kopuruak (aldakiak) eta aldakien ehunekoak (%) ematen ditugu.



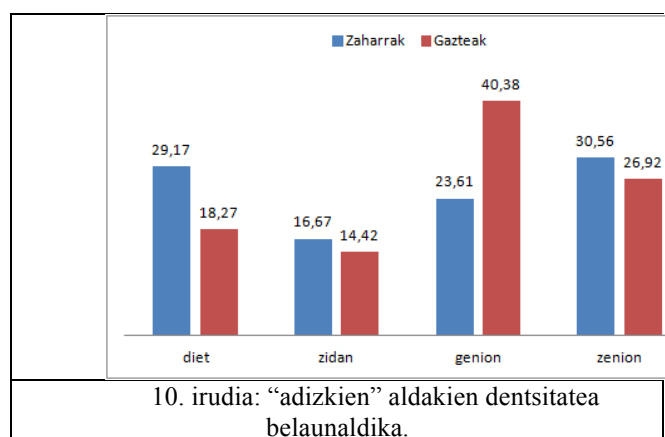
	Zaharrak			Gazteak		
	Kopuruak	Aldakiak	%	Kopuruak	Aldakiak	%
<b>diet</b>	67	21	29,17	550	19	18,27
<b>zidan</b>	70	12	16,67	554	15	14,42
<b>genion</b>	66	17	23,61	563	42	40,38
<b>zenion</b>	73	22	30,56	563	28	26,92
<b>Guztira</b>	276	72	100	2230	104	100

21. taula: Adizki estandar bakoitzeko datu eta aldaki kopuruak euren ehunekoekin belaunaldien arabera.

21. taulan ikusten den bezala, adizki estandar bakoitzaren aldakien dentsitatea ez dator bat belaunaldi bien datuetan; jarraian, hierarkikoki adierazita, zaharren eta gazteen erreskadak erakusten dira (10. irudiko grafikoa).

**Zaharrak**zenion >> diet >> genion >> zidan  
**Gazteak** genion >> zenion >> diet >> zidan

Zaharrek gazteek baino aldaki gehiago erabili dituzte kasu gehienetan “genion” adizki estandarren kasuan izan ezik.



Belaunaldien arteko ezaugarrien erkaketa errazago egiteko adizki estandarren arabera aurkeztuko ditugu banan.

#### 4.1. diet

Adizki honen algoritmoaren lehen eragiketa morfemen hurrenkera (MH) bien artean erabakitzeari dagokio. Zaharren belaunaldian 1MHren probabilitatea 0,672koa da eta 2MHrena 0,328koa; gazteen artean, berriz, 1MHren probabilitatea 0,773koa da eta 2MHrena 0,227koa. Aldeak ez dira oso handiak.

1MH erabiliz gero, gainerako eragiketen aldakiak eta euren probabilitateak 22. taulan agertzen dira. Bertan agertzen den moduan, oinaren aldakiei dagokienez, ikusten dugu “dau” zaharrek baino ez dutela erabiltzen; gainerakoen probabilitateak ere ez dator bat belaunaldi bietan:

**Zaharrak** “do” >> “Ø” >> “du” >> “deu” >> “dau”  
**Gazteak** “do” >> “Ø” >> “deu” >> “du”

“Nori” morfemaren aldakiei erreparatuta, ikusten dugu “tzei” galdu dela gazteen belaunaldian eta gainerakoak berberak diren arren, bakoitzaren probabilitateak oso aldatuta daudela.

1 Morfemen hurrenkera					
oina	“do”	“deu”	“du”	“dau”	“Ø”
Zaharrek	0,622	0,044	0,111	0,022	0,200
Gazteek	0,812	0,026	0,002	0	0,160
Nori	“tzie”	“tze”	“tzea/tzia”	“tzei”	
Zaharrek	0,533	0,356	0,089	0,022	
Gazteek	0,882	0,087	0,031	0	

22. taula: 1. morfema hurrenkera erabilia, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

2MH erabilia, gainerako eragiketen aldakiak eta euren probabilitateak 23. taulan agertzen dira. Oinaren aldakiei dagokienez, ikusten dugu “de” zaharrek baino ez dutela erabiltzen eta “deu”, berriz, gazteek; gainerako bien probabilitateak ere oso desberdinak dira belaunaldi bakoitzean. “Nori<sub>1</sub>” morfemaren aldakiak bat dator belaunaldi bietan baina probabilitateak oso desberdinak dira bakoitzean; gazteen artean “tzie” aldakiaren probabilitatea hazi da. Azkenik, “Nori<sub>2</sub>” morfemaren aldakiei dagokienez, “ea/ia” gazteen belaunaldian baino ez da agertzen eta beste bien probabilitateak ere oso aldatuta agertzen dira; gazteen artean “ie” haziz etorri da.

2 Morfemen hurrenkera				
oina	“do”	“deu”	“de”	“Ø”
Zaharrek	0,500	0	0,045	0,455
Gazteek	0,312	0,008	0	0,680
“Nori <sub>1</sub> ”	“tzie”	“tze”		
Zaharrek	0,773	0,227		
Gazteek	0,928	0,072		
“Nori <sub>2</sub> ”	“ie”	“e”	“ea/ia”	
Zaharrek	0,409	0,591	0	
Gazteek	0,800	0,192	0,008	

23. taula: 2. morfema hurrenkera erabilia, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

#### 4.2. zidan

Algoritmoaren lehen eragiketa paradigma aldaketa egiteari dagokio, hau da, Nork-Nori-Nor saileko adizkien ordezkari Nork-Nor sailekoak erabiltzea. Paradigma aldaketa hau gazte bakan batzuek baino ez dute egiten. Paradigma aldaketarik ez badago gainerako eragiketak, iragan morfemen, erroen eta “nori” morfemen aukeraketak egitea dira (24. taula).

Iragan morfemaren aldakiak bat datoz belaunaldi bietan ez, ordea, bakoitzaren probabilitateak:

<b>Zaharrek</b>	“Ø” >> “d” >> “s” // “y”
<b>Gazteak</b>	“Ø” >> “d” >> “s” >> “y”

Erroaren aldakiak direla eta, zaharren ez dute “au” erabiltzen, kasu honetan ere, aldaki bakoitzaren probabilitateak nahikoa desberdinak dira belaunaldi bakoitzean:

<b>Zaharrek</b>	“o” >> “eu” >> “u” >> “Ø”
<b>Gazteak</b>	“o” >> “Ø” >> “eu” >> “au” // “u”

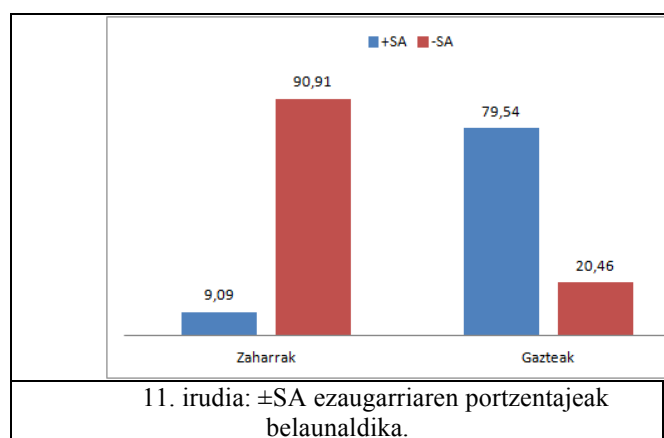
Azkenik, “nori” morfemaren aldakiak berberak dira belaunaldi bietan eta euron probabilitateak ere antzekoak.

- Paradigma aldaketa					
Iragana	“Ø”	“s”	“d”	“y”	
Zaharrek	0,714	0,029	0,229	0,029	
Gazteak	0,390	0,201	0,348	0,060	
Erroa	“Ø”	“au”	“eu”	“o”	“u”
Zaharrek	0,057	0	0,200	0,671	0,071
Gazteak	0,138	0,002	0,040	0,819	0,002
Nori	“sta”	“ste”			
Zaharrek	0,200	0,800			
Gazteak	0,220	0,780			

24. taula: Eragiketa bakoitzaren ezaugarrien aldakien probabilitateak, paradigma aldaketa egiten ez denean.

### 4.3. genion

Adizki honen algoritmoaren lehen eragiketa subjektuaren leku aldaketari dagokio ( $\pm SA$ ), beronen arabera morfema hurrenkera bi sortzen dira. Zaharren kasuan  $+SA$ ren probabilitatea 0,091koa da eta  $-SA$ rena 0,909koa; gazteen kasuan  $+SA$ rena 0,797koa da eta  $-SA$ rena 0,205koa. Aldea honetan oso handia da (11. irudiko grafikoa).



Lehen bidea segitzen badugu, hau da, subjektuaren leku aldaketa gabekoa ( $-SA$ ), gainerako eragiketak “nork” morfema erroarekin amalgamatzea edo ez, “nork” morfema, erroa, eta “nori” morfema dira (25. taula). Lehen eragiketari dagokionez, ikusten dugu belaunaldien artean ez dagoela alde handirik, nahiz gazteen belaunaldian probabilitate handiagoa egon “nork” morfema erroarekin ez amalgamatzeko. “Nork” morfemaren kasuan zaharren belaunaldia “gen\_” eta “g\_” morfemak ez dira agertzen.

Erroen aldakien kasuan ikusten dugu belaunaldi biek berberak erabiltzen dituztela; bakoitzaren probabilitateak desberdinak direnez gero, probabilitateen araberrako erreskada hierarkikoak ez datoz bat.

<b>Zaharrek</b>	“eu” >> “o” >> “e” >> “u” >> “au” // “i”
<b>Gazteak</b>	“eu” >> “e” >> “o” >> “au” >> “u” // “i”

Azkenik, “nori” morfemaren aldakiak bat datoz belaunaldi bietan, bakoitzaren probabilitateen araberako erreskada hierarkikoak bat etorri ez arren:

**Zaharrek** “tze” >> “tza” >> “tzo”  
**Gazteak** “tze” >> “tzo” >> “tza”

- Subjektuaren leku aldaketa						
	+Amal	-Amal				
Zaharrek	0,983	0,017				
Gazteek	0,861	0,139				
“nork”	g_n	g_	gen_	gin_		
Zaharrek	0,983	0	0	0,017		
Gazteek	0,861	0,009	0,122	0,009		
erroa	“eu”	“au”	“o”	“e”	“i”	“u”
Zaharrek	0,517	0,033	0,183	0,150	0,033	0,083
Gazteek	0,478	0,078	0,183	0,243	0,009	0,009
“nori”	“tza”	“tze”	“tzo”			
Zaharrek	0,217	0,667	0,117			
Gazteek	0,235	0,496	0,270			

25. taula: Subjektuaren leku aldaketa egin barik, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

Subjektuaren leku aldaketak eragindako bidea segituta, eragiketak iragan morfemaren, erroaren, “erroaren” eta “nori” morfemaren txertaketei dagozkie (26. taula). Bertan ikusten den moduan, iragan morfemari dagokionez, zaharrek ez dute “y” erabiltzen eta gazteek bai; bestalde, morfema bakoitzaren probabilitateen erreskada hierarkikoak ez datoz bat belaunaldi bietan:

**Zaharrek** “d” >> “s” >> “Ø”  
**Gazteak** “Ø” >> “d” >> “s” >> “y”

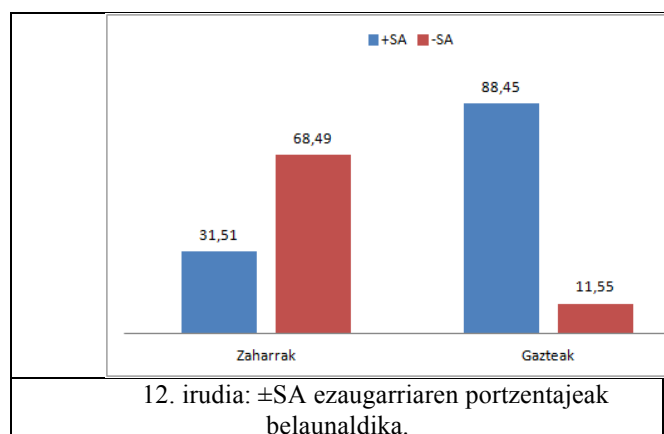
Erroaren aldakiei dagokienez gazteen belaunaldian “eu”, “u” eta “au” agertzen dira probabilitate txikiekin eta zaharren belaunaldian ez; gainerakoei dagokienez “o” da probabilitaterik altuena daukana eta gero “Ø”, probabilitateen erreskada hierarkikoa bat da belaunaldi bietan. “Nori” morfemaren aldakiei dagokienez, zaharren belaunaldian aldaki bat (“tze”) baino ez da agertzen. Azkenik, “nork” morfemaren “gu” aldakia belaunaldi bietan probabilitaterik handiena daukana da, gazteen belaunaldian “u” aldakia ere erabiltzen da, beronen probabilitatea txikia den arren.

+ Subjektuaren leku aldaketa					
Iragan	“Ø”	“d”	“s”	“y”	
Zaharrek	0,167	0,500	0,333	0	
Gazteek	0,519	0,228	0,186	0,067	
Erroa	“eu”	“o”	“u”	“au”	“Ø”
Zaharrek	0	0,833	0	0	0,167
Gazteek	0,025	0,644	0,002	0,002	0,327
“Nori”	“tza”	“tze”	“tzo”		
Zaharrek	0	1	0		
Gazteek	0,441	0,550	0,009		
“Nork”	“gu”	“ku”	“u”		
Zaharrek	0,833	0,167	0		
Gazteek	0,933	0,004	0,063		

26. taula: Subjektuaren leku aldaketa eginda, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

## 4.4. zenion

Aurreko adizkiarekin egin dugun bezala, beronen algoritmoaren lehen eragiketa subjektuaren leku aldaketari dagokio ( $\pm SA$ ), beronen arabera morfema hurrenkera bi sortzen dira. Zaharren kasuan +SAren probabilitatea 0,315koa da eta -SArena 0,685koa; gazteen kasuan +SArena 0,885koa da eta -SArena 0,115koa. Aldea honetan oso handia da (12.irudiko grafikoa).



Lehen bidea segitzen badugu, hau da, subjektuaren leku aldaketa gabekoa (-SA), gainerako eragiketak “nork” morfema erroarekin amalgamatzea edo ez, “nork” morfema, erroa, eta “nori” morfema dira (27. taula). Lehen eragiketari dagokionez, ikusten dugu belaualdien artean ez dagoela alde handirik, nahiz gazteen belaualdian probabilitate handiagoa egon “nork” morfema erroarekin ez amalgamatzeko, hau gerta daiteke euskara estandarren eraginagatik. “Nork” morfemaren kasuan zaharren belaualdia “sen” morfema ez da agertzen. Probabilitateen erreskada hierarkikoak ondoko era honetara agertzen dira:

**Zaharrak**       $s\_n \gg s\_//sin\_$   
**Gazteak**       $s\_n \gg sen\_ \gg s\_ \gg sin\_$

Erroen aldakien kasuan ikusten dugu zaharrek erabiltzen dituzten “i” eta “au” erroak ez direla agertzen gazteen belaualdian, bakoitzaren probabilitateen erreskada hierarkikoak ere ez datoz bat

**Zaharrak**      “eu” >> “e” >> “o” // “u” >> “au” // “i” // “ai”  
**Gazteak**      “eu” >> “e” >> “o” >> “au” >> “u”

Azkenik, “nori” morfemaren aldakiak eta euron erreskada hierarkikoak bat datoz belaualdi bietan, bakoitzaren probabilitateen aldeak handi samarrak diren arren.

- Subjektuaren leku aldaketa							
	+Amal	-Amal					
Zaharrek	0,960	0,040					
Gazteek	0,846	0,154					
“Nork”	s n	s	sen	sin			
Zaharrek	0,960	0,020	0	0,020			
Gazteek	0,859	0,031	0,094	0,016			
erroa	“eu”	“au”	“o”	“e”	“i”	“u”	“ai”
Zaharrek	0,620	0,020	0,100	0,120	0,020	0,100	0,020
Gazteek	0,446	0,123	0,138	0,262	0	0,031	0
“nori”	“tza”	“tze”	“tzo”				
Zaharrek	0,140	0,740	0,120				
Gazteek	0,277	0,554	0,169				

27. taula: Subjektuaren leku aldaketa egin barik, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

Subjektuaren leku aldaketak eragindako bidea segituta, eragiketak iragan morfemaren, erroaren, “erroaren” eta “nori” morfemaren txertaketei dagozkie (28. taula). Bertan ikusten den moduan, iragan morfemari dagokionez, zaharrek ez dute “y” erabiltzen eta gazteek bai; bestalde, morfema bakoitzaren probabilitateen erreskada hierarkikoak bat datoz belaunaldi bietan. Erroaren aldakiei dagokionez, gazteen belaunaldian “u” eta “au” agertzen dira probabilitate txikiakin, gainerakoei dagokionez “Ø” da probabilitaterik altuena daukana eta gero “o” eta “eu”, probabilitateen erreskada hierarkikoa bat da belaunaldi bietan. “Nori” morfemaren aldakiei dagokionez, belaunaldi bakoitzaren probabilitateen aldeak nabarmenak dira, zaharren artean “tza” aldakiak dauka probabilitaterik altuena eta gazteen artean, berriz, “tze” aldakiak.

+ Subjektuaren leku aldaketa					
Iragan	“Ø”	“d”	“s”	“y”	
Zaharrek	0,783	0,130	0,087	0	
Gazteek	0,536	0,205	0,191	0,068	
Erroa	“eu”	“o”	“u”	“au”	“Ø”
Zaharrek	0,043	0,435	0	0	0,522
Gazteek	0,030	0,639	0,002	0,002	0,327
“Nori”	“tza”	“tze”			
Zaharrek	0,565	0,435			
Gazteek	0,430	0,570			

28. taula: Subjektuaren leku aldaketa egin da, eragiketa bakoitzaren ezaugarrien aldakien probabilitateak.

## 5. Ondorioak eta hurrengo lanak

Azken atal honetan azterketaren bidez lortu ditugun ondorioak nabarmenenak laburbilduko ditugu. Ikusi dugun bezala, oso gutxi dira belaunaldi batetik bestera pasatzean galdu diren aldakiak, aldakiok gainera zaharren belaunaldian probabilitate txikiak eduki dituztenak izan dira. Era berean, sortzen direnak edo berrikuntzat har daitezkeenak oso gutxi dira eta probabilitate txikiak edukitzen dituzte.

Galerarik ohikoenak aditz oinaren eta erroaren aldakiei dagokie; horrela bada “dier” adizkiaren “dau” eta “de” oinak galdu dira; “zidan” adizkiaren kasuan “au”, “genion” adizkiaren “eu”, “u” eta “au” eta “zenion” adizkiaren “i”, “ai”, “u” eta “au”. Zaharrek erabiltzen ez duten bat agertu da gazteen belaunaldian; “dier” adizkiarekin

“deu”. “*diet*” adizkiaren “tzei” morfema ere galdu da gazteen belaunaldian. Berrikuntzei bagagozkie, “ea/ia” “nori<sub>2</sub>” morfemaren agerpena dugu “*diet*” adizkiarekin. “*zidan*” adizkiarekin Nork-Nori-Nor sailekoen orde z Nork-Nor sailekoak edo paradigma aldaketa egitea berrikuntza garrantzitsua da. Era berean, “*genion*” eta “*zenion*” adizkiekin “nork” morfema eta erroa amalgamatuta ez erabiltzea aldaketa esanguratsua da, izan ere, honetan euskara estandarren eragina islatzen da. “*genion*” eta “*zenion*” adizkiekin subjektu aldaketa dagoenean “y” iragan morfema gazteen belaunaldian baino ez da agertzen.

Gainerakoan, ezaugarri berberak erabiltzen dira belaunaldi bietan eta aditzak eraikitzeke eragiketa berberak egin behar dira, aldatzen dena bakoitzaren agerpen probabilitatea da belaunaldi bakoitzean. Probabilitate aldaketa horiek eragin dezakete aldakien probabilitateen arabera egin daitezkeen erreskada hierarkikoak bat ez etortzea eta, bat badatoz ere, euren probabilitateen aldeak handiak izatea. Honek esan nahi du galerak eta berrikuntzak gutxi izanda, aldatzen dena erabilera bakoitzaren probabilitatea dela eta eurok islatzen dituztela zeintzuk diren joerak.

Beraz, hemen aztertu ditugun adizkien bilakaeran gertatzen diren aldaketak lau multzotan laburbil daitezke:

- (a) Ezaugarrien galerak
- (b) Ezaugarrien berrikuntzak
- (d) Ezaugarrien aldakien probabilitateen arabeko erreskada hierarkikoen aldeak
- (e) Ezaugarrien probabilitateen aldeak

Hurrengo lanei begira uste dugu hemen aurkeztu dugun metodologia arakaten segitu beharko genukeela. Aurreikusten ditugun azterkizun batzuk honako hauek izango lirateke besteak beste:

- (a) Herrien populazio euskaldunaren arabera adizkien aldakien probabilitateak kalkulatzeko
- (b) Herriak sailkatzea bertan erabiltzen diren adizkien tasunen probabilitateen arabera
- (c) Eragin gune geografikoak zeintzuk diren zehazki zehaztea
- (d) Sistema konplexu dinamiko moldagarrien ikuspuntutik sakonean dautzan egituraketa geografikoak azalerratzeko.

## 6. Bibliografia

- Agirrebeña, A. eta Elexpuru, J. M. (2014). *Elorrixoko Berbetia*. Elorrioko Udala. <http://elotop.net/euskara/ElorrixokoBerbetia.pdf>
- Arteagaitia, X. (2000). *Hatsarreak eta parametroak lantzen*. Gasteiz: AFA.
- Beckner, C.; Blythe, R.; Bybee, J.; Christiansen, M. H.; Croft, W.; Ellis, N. C.; Holland, J.; Ke, J.; Larsen-Freeman, D. eta Schoenemann, T. (2011). La lengua es un sistema adaptativo complejo Lingüística en la Red. [http://www.linred.es/articulos\\_pdf/LR\\_articulo\\_04092011.pdf](http://www.linred.es/articulos_pdf/LR_articulo_04092011.pdf)
- Ellis, N. C. (2011). The Emergence of Language as a Complex Adaptive System. In James Simpson (Eds.), *Routledge Handbook of Applied Linguistics* (1-29 orr.). London: Routledge/Taylor Francis.

- Ensunza, A. (2015). *Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan*. Doktorego tesia.
- Gaminde, I. (1985). *Aditza Bizkaieraz III*. Iruña:UEU.
- Gaminde, I.(2010).*Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I.; Romero, A. eta Legarra, H. (2012).*Gramatika eta Hizkuntz Bariazioa Bermeon*. Bermeo: Bermeoko Udala eta Campos Hegaluzea.
- Gell-Mann, M. (1994).*El quark y el jaguar: Aventuras en lo simple y lo complejo*. San Francisco: EpubLibre
- Léonard, J. L., Heinsalu, E., Patriarca, M. eta Darlu, P. (2015). Modeling regional variation from eas: complexity and communal aggregates.In Aurrekoetxea, G.; Romero, A. eta Etxebarria, A. *Linguistic Variation in the Basque an Education-I/Euskararen Bariazioa eta Bariazioaren Irakaskuntza-I* (50-58or.) Bilbo:UPV/EHU
- López Rivera, J. J. (2013). Aplicación del marco de los sistemas complejos adaptativos a un modelo de variación lingüística *Moenia* 19, 5-24.
- Moreno Cabrera, J. C. (2008).*El nacionalismo lingüístico: Una ideología destructiva*.Bartzelona:Ediciones Península.
- Yrizar, P. (1992).*Morfología del verbo auxiliar vizcaíno*. Bilbo:Euskaltzaindia.