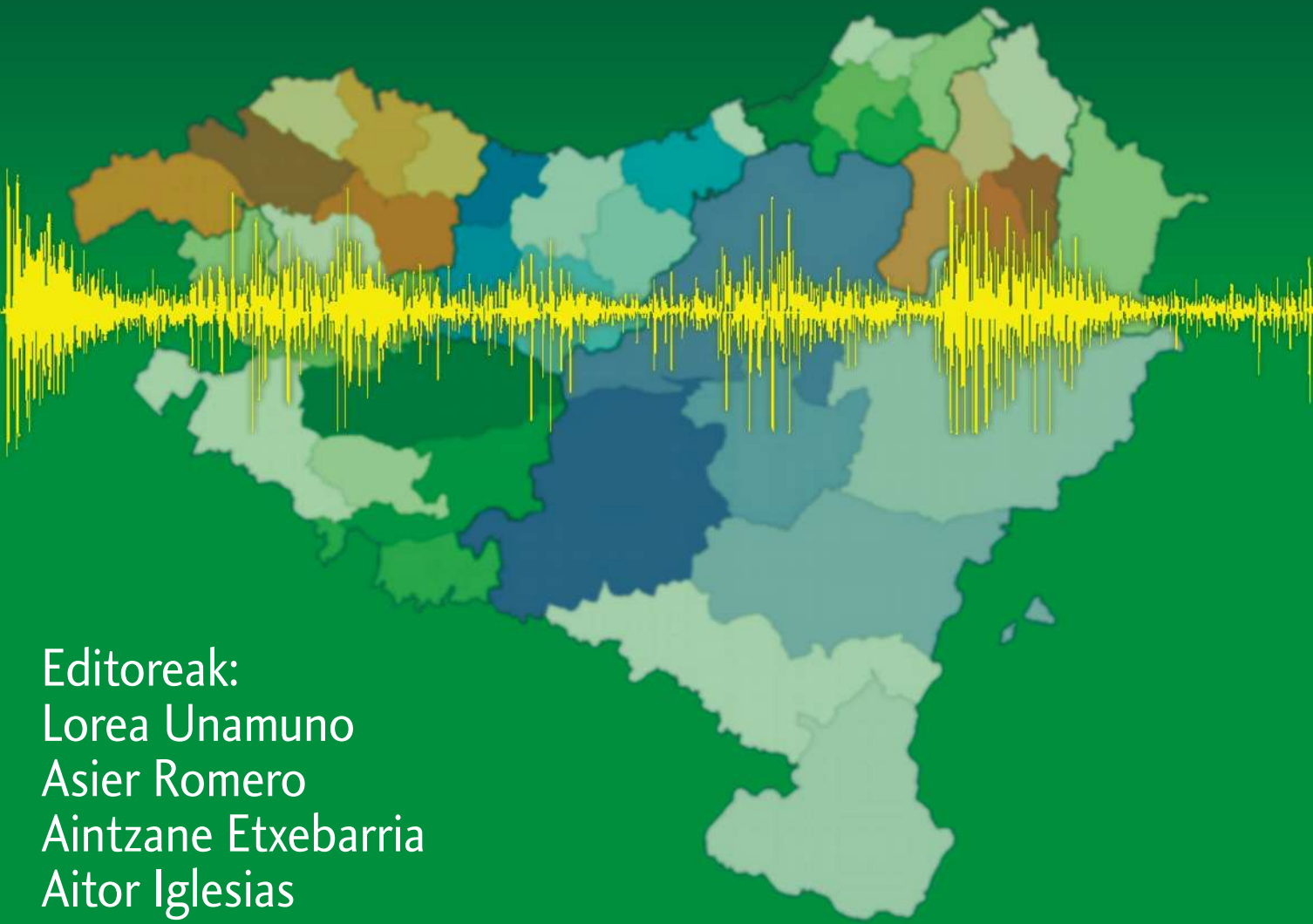


Linguistic Variation in the Basque and Education-III

Euskararen bariazioa eta bariazioaren irakaskuntza-III



Editoreak:

Lorea Unamuno

Asier Romero

Aintzane Etxebarria

Aitor Iglesias



Linguistic variation in the basque language and education - III [Recurso electrónico] = Euskararen bariazioa eta bariazioaren irakaskuntza -III /editoreak, Lorea Unamuno ...[et al.]. – Datos. - Bilbao : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, D.L. 2017.

1 disco compacto CD-Rom (PDF, 156 p.)

Requisitos del sistema: Adobe Acrobat Reader.

Textos en euskara, inglés y español.

D.L.: BI-2108-2017. - ISBN: 978-84-9082-798-7.

1. Euskara (Lengua) - Dialectos. 2. Lenguaje y lenguas - Variación. 3. Prosodia (Lingüística). I. Unamuno, Lorea, coed. II. Título: Euskararen bariazioa eta bariazioaren irakaskuntza -III.

(0.034)809.169-087



UPV-EHUko EUDIA Ikerketa-taldea

Euskal Herriko Unibertsitateko ikerketa taldea (GIU 22/16)

Euskal Unibertsitate Sistemako ikerketa taldea (IT 1028/16)

AURKIBIDEA

	pág.
Sarrera <i>Lorea, U.; Romero, A.; Etxebarria, A.; Iglesias, A.</i>	2-3
Actions, gestures and words: multimodality in the early stages of language development <i>Murillo Sanz, E.</i>	4-14
Metodología experimental y tratamiento estadístico. Aplicación al estudio de la prosodia del catalán <i>Borràs-Comes, J.</i>	14-28
Informatzaileen sailkapenerako arazo metodologiko batzuen gainean <i>Gaminde, I.; Eguskiza, N.; Romero, A.; Etxebarria, A.</i>	29-42
Ahots seduzitzaileen ezaugarriak generoaren ikuspegitik <i>Etxebarria, A.; Eguskiza, N.; Gaminde, I.; Romero, A.</i>	43-59
The Acoustic Realization of /s/ and /tʃ/ by L1 and L2 Basque Speakers <i>Beristain, A.</i>	60-72
Igorreko azentuaren korrelatu akustikoez <i>Iglesias, A.; Eguskiza, N.; Gaminde, I.; Unamuno, A.</i>	73-86
Compress-EUS: i(ra)kasleen laburpenak lortzeko tresna <i>Atutxa, U.; Iruskieta, M.; Ansa, O.; Molina, A.</i>	87-98
Zubererazko Dardarkarien azterketa sozio-fonetikoa <i>Etchebest, X.</i>	99-105
Larrabetzuko euskaldun berrien seme-alaba euskaldun zaharren ezaugarri prosodikoez <i>Eguskiza, N.; Gaminde, I.; Olalde, A.; Etxebarria, A.; Gaminde, U.</i>	106-117
Prozedura metodologikoren konparaketa haur elebidunen hasierako garapen lexikoaren informazioaren lorpenerako <i>De Pablo, I.; Romero, A.; Etxebarria, A.; Gaminde, I.</i>	118-125
Euskalkien sailkapen zientifikoa: Lexikoa <i>Aurrekoetxea, G.; Ormaetxea, J.L.; Videgain, X.</i>	126-140
Eranskinak (kartela eta argazkiak)	141-151

SARRERA

Lorea Unamuno, Asier Romero, Aintzane Etxebarria eta Aitor Iglesias 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 bakarria. 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 baitakigu teoria artifizio bat baino ez dela, entelekia bat, errealitatea birsortzeke eta interpretatzeko. Hori dela eta ikertzailea jarrera dogmatikoetatik urundu beharrean dago, nahiz uste osoa izan berea dela egokiena; 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 desberdineko jardunaldiak eratzeari. Beti helburu bakarria gogoan: euskararen bariazioari buruzko ezagutza handitzea, era bateko zein besteko ekarpenak egiteke guneak sortuz. Ekarpak guztiak dira baliagarriak eta denetatik egoten da zer ikasi. Ekarpak eta haien 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 sei jardunaldi edo *workshop* eratu ditu Gasteizen, Bilbon eta Leioan, 2008az geroztik (2009an, 2011n, 2012an, 2015ean, 2016an eta 2017ko 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 urundu behar du, izurritik bezalaxe. Ikusmolde dogmatikoak ez datoz bat zientziarekin. Ikertzailea dogmatismotik urundu 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. 2017ko 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 Beristain, Murillo, Borràs... izan ditugu euren ekarria aurkezten. Guztira hamaika txosten aurkeztu dira egun bateko baina bete-beteko jardunaldian. Eskerrak hemendik parte hartzaile guztiei.

Ekarpenak orotara bost multzotan sailka genitzake:

- a) Hizkuntza bariazioaren inguruan ikusmolde desberdinak erabiliz, hala nola teoria desberdinen ikusmoldeak (Gaminde, Eguskiza, Romero eta Etxebarria “Informatzaileen sailkapenerako arazo metodologiko batzuen gainean”, tresneriaren aldetik (Atutxa, Iruskietza, Ansa eta Molinaren “Compress-EUS: i(ra)kasleen laburpenak lortzeko tresna”), fonetika landuz (Beristainen “The Acoustic Realization of /g/ and /tʃ/ by L1 and L2 Basque Speakers” eta Iglesias, Eguskiza, Gaminde eta Unamunoren “Igorreko azentuen korrelatu akustikoez”), emozioen ikuspuntutik (Etxebarria, Eguskiza, Gaminde eta Romeroren “Ahotsa seduzitzaileen ezaugarriak generoaren ikuspegitik”), bariazio soziolinguistikoa (Ensunzaren “Hizkuntza-aldaketaren inplikazio soziolinguistikoa euskarari: jarrerak eta prestigioa” eta Etchebesten “Zubererazko Dardakarien azterketa sozio-fonetikoa”; eta azkenik bariazio morfologikoaren gaineko lanak: Aurrekoetxea, Videgain eta Ormaetxearen “Euskalkien sailkapen zientifikoa: lexikoa”).
- b) Intonazioaren bariazioa bereziki landu izan da ekarri zenbaitetan, hala nola Borràs-Comesen “Metodología experimental y tratamiento estadístico: aplicación al estudio de la prosodia del catalán”.
- c) Code-switching edo kode-aldaketa landu dute Eguskiza, Gaminde, Olalde, Etxebarria eta Gamindek bere “Larrabetzuko euskaldun berrien seme-alaba euskaldun zaharren ezaugarri prosodikoez” ekarriarekin.
- d) Eta azkenik, haurren keinuak eta lehen hitzak ere landu dira ekarpen birekin: Murilloren “Actions, gestures and words: multimodality in the early stages of language development” eta De Pablo, Romero, Etxebarria eta Gaminderen “Metodología ezberdinak haur euskaldunen komunikazio-garapena neurtzeko”.

ACTIONS, GESTURES AND WORDS: MULTIMODALITY IN THE EARLY STAGES OF LANGUAGE DEVELOPMENT

Eva Murillo Sanz
Universidad Autónoma de Madrid
eva.murillo@uam.es

Abstract

Language development is always a multimodal process. Children coordinate gestures and vocalizations from the early stages of language development. They learn to synchronize different communicative resources even before they are able to produce words. Gestures and vocalizations become parts of the same linguistic system, and multimodal communicative behaviors appear throughout all stages of language development. The ability to coordinate gestural and vocal elements with a single communicative purpose results predictive of subsequent linguistic achievements at different stages of language learning process. This paper focus on the predictive value of multimodal communicative behaviors throughout the early stages of language development.

Keywords: Language development, gestures, vocalizations, multimodal communication

1. Introduction

Language is a multimodal process that develops in a multimodal context. Children learn language through interaction with people who talk, smile, make gestures with their hands, move and look at interesting things in the world. At the same time, children interact with the social context by means of vocalizations, motor actions, gestures or facial expressions. These communicative resources are part of a single language system in which different expression channels act together to convey meanings to others. Gestural and vocal elements are always linked in language development. We can find this gesture and vocalization coordination throughout all the language learning stages. As we will see in the following sections, these multimodal communicative behaviors have a predictive role on subsequent linguistic abilities throughout the early stages of language development.

2. Before the first words: the one-word transition

Children learn language immersed in a multimodal world in which they experience sounds, rhythms, faces and objects around them. At the same time, from the earlier stages of language development they interact with adult partners by means of different channels of communicative expression. They produce a variety of vocal behaviors that affect adult's response (Keller & Schölmerich, 1987; Bloom, Russell & Wassenberg, 1987; Masataka, 1993a,b), and they also produce motor actions in a communicative way. Vocal and motor components not only co-develop, but also interact from the first stages of language development. As claimed by Iverson (2010), changes in motor skills increases the set of opportunities for acting in the world. These opportunities provide contexts that facilitate the development of abilities contributing directly and indirectly to communicative and linguistic development.

Well before infants are able to use words or master the grammatical rules, they use vocal and motor elements in a coordinated way. As early as 2-3 months of age

infants employ the index finger extension more frequently with syllabic vocalizations than with vocalic ones or without vocalizations (Fogel & Hannan, 1985; Masataka, 1995). As syllabic vocalizations are perceived by adults as more speech-like than the vocalic ones, they tend to verbally respond to this type of vocalizations in a contingent way. With this contingent social response, they are indirectly reinforcing the link between index finger extension and syllabic vocalizations.

In the last third of their first year infants begin to produce canonical babbling. Canonical babbling is considered one of the vocal milestones in the process of language development, and is characterized by the production of well-formed syllables that have adult-like spectral and temporal properties (Oller, 1986). Manual rhythmic activity appears related to the emergence of canonical babbling. The moment of canonical babbling emergence is preceded by an increase in manual rhythmic activity (Ejiri, 1998). In addition, the vocalizations accompanied by rhythmic activity have different acoustic properties from those produced without it (Ejiri & Masataka, 2001).

By the end of their first year, infants begin to use manual gestures with a communicative purpose. The use of gestures, and specifically the pointing gesture has predictive value for subsequent lexical development (Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Cadime, Silva, Santos, Ribeiro, Viana, 2017; Camaioni, Castelli, Longobardi & Volterra, 1991; Colonnaesi, Stams, Koster & Nboom, 2010; Rowe, Özçaliskan & Goldin-Meadow, 2008). Before children are able to produce words, they coordinate effectively gestures and vocalizations in their communicative attempts (Murillo & Belinchón, 2013). Studying Spanish speaking children we found that the rate of gestural and vocal elements coordination accompanied by social use of gaze at 12 months of age was a strong predictor of word production three months later (Murillo & Belinchón, 2012). Other studies have found similar results, Wu and Gros-Louis (2014) also found that gesture-vocal coordination was related to infants' linguistic skills at 15 months of age. In the same line, Igualada, Bosch and Prieto (2015), showed that infants who were able to integrate pointing and speech at 12 months of age, showed better vocabulary abilities six months later.

It seems, therefore, that even before children produce words, the multimodal communicative behaviors are related with subsequent linguistic achievements. Iverson and Goldin-Meadow (2005) observed that children aged 10 to 12 months relied heavily on gestures to refer to objects. The items appeared first referred by means of gestures, emerging subsequently in their verbal lexicons.

The differential response from adults to multimodal communicative behaviors could explain this predictive value on lexical development. Behaviors such as showing or pointing towards an object while vocalizing elicit the adult to name the referent in a joint-attention frame. Several studies have shown the differential response from adults to multimodal communicative behaviors from the child (i.e. Fasolo and D'Odorico, 2012). Hence, multimodal communicative behaviors can enhance the opportunities for word learning.

Multimodal communicative behaviors produced by infants develop throughout the transition from the beginning of the intentional communication to the first words use and to the first two-word combinations. As reported by Esteve-Gibert and Prieto, (2014), children are able to temporally synchronize gestures and speech before they

produce the first words. They found that children combine gestures and vocalizations with an adult-like pattern. The prominences of gestures and vocalizations were aligned in children's productions. As in the adult's multimodal productions, in the infant's communicative behaviors the gesture onset precedes the speech onset, gesture stroke onset co-occurs with speech onset and stroke onset precedes the beginning of the accented syllable.

Romero, Etxebarria, de Pablo and Romero (2017) report similar results with Basque-speaking infants. Their results showed that the interval between the gesture onset and the vocalizations onset as well as the interval between the stroke onset and the vocalization onset gradually decreases from 9 to 13 months of age.

In order to explore the changes in the synchrony of the gestural and vocal elements from 9 to 18 months of age, we longitudinally observed 8 Spanish-speaking children every three months in a semi-structured play situation (Murillo, Ortega, Otones & Casla, 2017). We measured the total duration of the communicative behavior and the proportion of temporal overlap between vocalizations and gestures. We also considered the onset distance between elements, that is, the interval between the onset of the first communicative element produced and the second one.

Figure 1 shows the mean duration of the multimodal communicative behavior at each age. Multimodal communicative behaviors tend to be shorter with age.

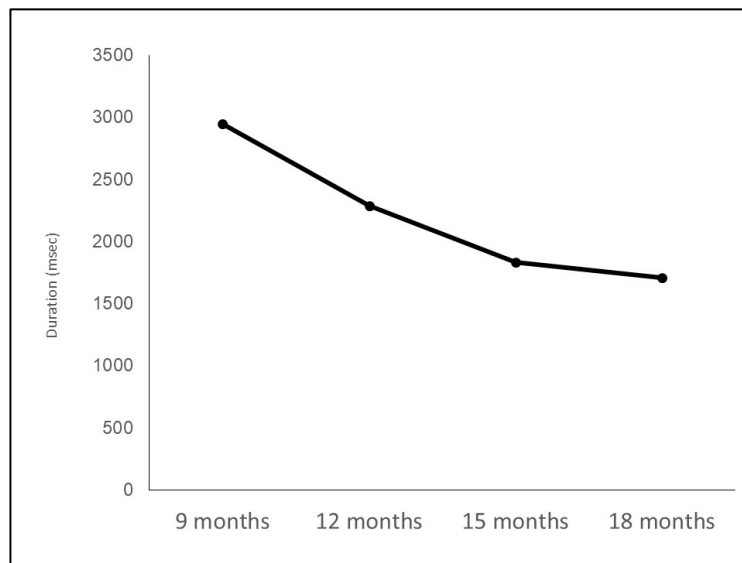


Fig. 1.- Mean duration of multimodal communicative behavior by age

Regarding temporal overlap between elements, we found that the proportion of temporal overlap increases with age. We found significant differences between the overlap proportion at 9 months and at 18 months of age (see Figure 2).

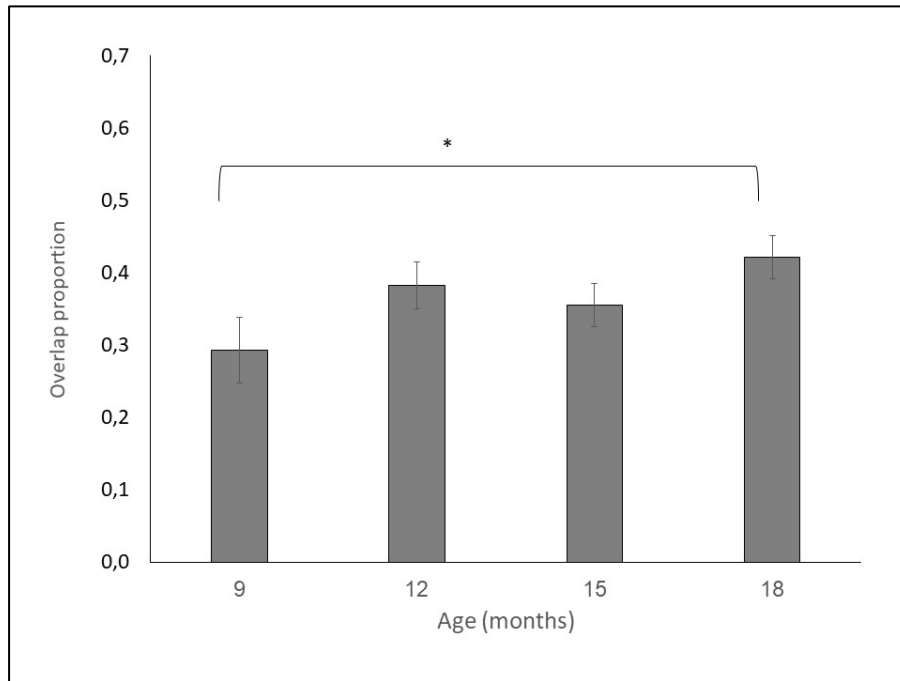


Fig.2.- Mean proportion of temporal overlap between communicative elements.

The onset distance between elements tends to decrease with age, especially after 9 months of age. Figure 3 shows the changes in onset distance with age.

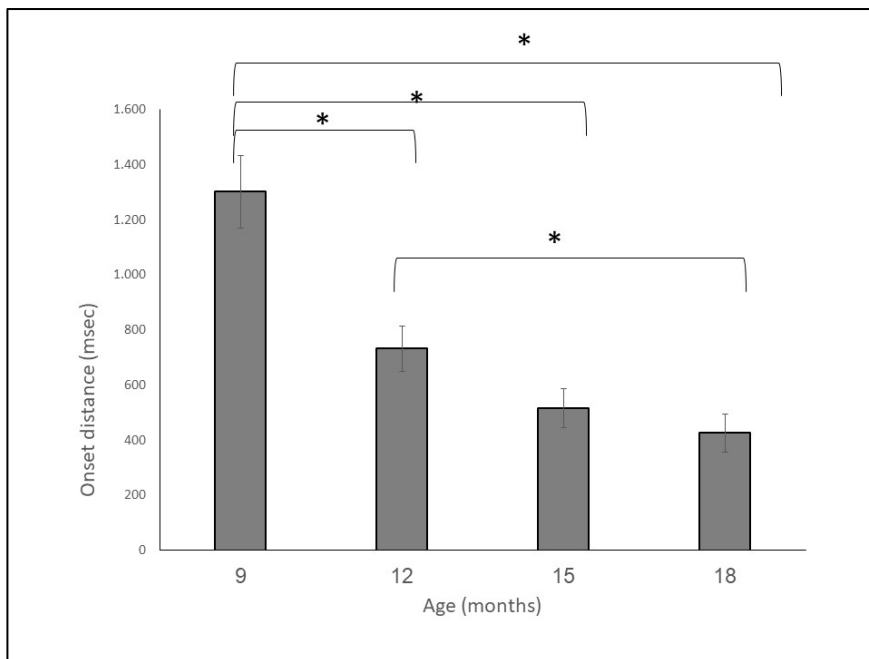


Fig.3.- Mean onset distance by age

Taken together, all these results show that the elements constituting the multimodal communicative behaviors are increasingly integrated, suggesting that communicative elements become parts of a single communication system throughout the period studied.

Moreover, in the transition to first words, the vocalizations produced with gestures have acoustic properties more similar to those of adult language than the vocalizations produced without gesture (Murillo & Capilla, 2016).

It seems, hence, that before children use their first words, they are able to produce multimodal communicative behaviors in a synchronized way. The ability to combine different communicative resources with a single communicative purpose results predictive of subsequent linguistic achievements. We should therefore consider these behaviors as an early index of linguistic development. What can an education professional expect on this period preceding the transitions to first words?. Around the first year of age, we should find: (1) that child produce communicative gestures accompanied by vocalizations; (2) that she produces a variety of gestures (such as pointing, reaching, showing, etc.); (3) that the communicative elements produced are increasingly synchronized; (4) that the gestures produced are mainly accompanied by vocalizations with a canonical structure; (5) that around 12 months of age the child produce pointing + vocalization + social gaze.

3. With the first words: the two-word transition

When children start to produce words, they continue at using gestures in their communicative attempts. Gestural and verbal elements are part of a single system that evolves with communicative and linguistic development. Words production do not replace gesture use in children's communication. In fact, multimodal communicative behaviors also predict subsequent linguistic achievements when children are already able to produce words. The use of gestures when children are aged 14 months, predicts the vocabulary size at 42 months of age (Rowe, Özçalışkan & Goldin-Meadow, 2008). Caprici, Iverson, Pizzuto and Volterra (1996) also found that the productions of gestures accompanied by words at 16 months of age was related to total vocal production four months later.

The combinations of gestures and words not only predict the vocabulary size, but specific gesture +vocal combinations also result predictive of specific subsequent linguistic constructions. Cartmill, Hunsicker and Goldin-Meadow, (2014) showed that the age at which children start to produce pointing + nouns combinations predicts the onset for determiner + noun constructions.

Regarding the onset of two-word combination, several studies relate changes in word-gesture combinations with the transition to two-word constructions. Özçalışkan and Goldin-Meadow (2005) offered a classification of gesture-speech combinations depending on the relationship of the information conveyed by each communicative element. They considered a *reinforcing* relation when gesture conveyed the same information that the accompanying speech (e.g. "car" + point at a car). When the gesture clarified the referent of a pronominal, demonstrative or deictic (e.g. "this" + point at toy) the relation was considered as *disambiguating*. Finally, when gesture added information to the message conveyed in speech (e.g. "mommy" + point at a car), the relation was considered as *supplementary*. By means of supplementary combinations, children convey two pieces of information, that is, they are already combining two elements before they produce two-word constructions. The onset age of gesture+ word supplementary combinations predicts the onset age of two-word constructions (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005). The number of

supplementary combinations at 18 months of age predicts the verbal syntactic complexity at 42 months (Rowe & Goldin-Meadow, 2009). Moreover, children produce particular sentence constructions by means of gesture+speech combinations before they are able of producing the same constructions entirely within speech (Özçalışkan and Goldin-Meadow, 2005). The use of gestures can be less cognitively demanding than remembering a conventional word form, facilitating and anticipating word combination.

Summarizing, it seems that when children are able to produce words, changes in the gesture-speech combinations precede and predict changes in subsequent syntactic development. Multimodal combinations can facilitate the transition to two-elements combinations by lightening the cognitive load in the communicative attempts from the child. What can a professional from education to expect in this period? In this stage of language development we should find: (1) a generalized use of gestures in the communicative attempts; (2) an increasing variety in the use of gestures: deictic gestures but also conventional and representational ones; (3) that are mostly produced with speech and; (4) gesture-speech combinations conveying supplementary meanings.

4. Beyond the two-word stage.

It seems, hence, that gesture and speech are linked from the early stages of language development. This link becomes evident in the adulthood, but there is scarce evidence about how children can reach an adult-like pattern regarding the gesture-speech combination.

Some studies have addressed the role of speech and gesture coordination when school-age children face problem-solving task, such as conservation problems (Church & Goldin-Meadow, 1986), mathematical equivalence problems (Alibali & Goldin-Meadow, 1993; Perry, Church, & Goldin-Meadow, 1986) and balance problems (Pine, Lufkin, & Messer, 2004). However, little is known about how children coordinate gestures and vocalizations when they are starting to master the syntactic rules.

Some studies, using picture-naming tasks suggests that there is a decrease in multimodal combinations after two years of age (Huttunen, Pine, Thurnham & Khan, 2013; Stefanini, Bello, Caselli, Iverson & Volterra, 2009). According to these studies, gesture production tends to decrease with age as verbal resources increase. However, this may not be the case when instead of naming pictures children have to construct sentences. To explore this issue, we designed an experimental task consisting in a “find the odd one” game. The syntactic structure of the target name was modified with the aim of eliciting not only object labeling, but also the expression of different relations between elements (Murillo, Galera & Casla, 2015). Each item included five familiar objects pictures, four of them were identical and one, the target picture was different. The complexity of distinguishing element of the target picture increased in terms of the relationship between elements depicted in each image. We classified the items as easy, medium or difficult according to their complexity. We included in the “easy” category the items in which the verbal response to spot the target was only a name. In the “medium” complexity category, we included items in which the difference was based on a characteristic (color or size) or on the presence of an object above or below the main element. In this category, the verbal answer (in the case of a perfect verbal response) was defined by two semantic elements. In the “difficult” category, we

included the items in which the difference of the target was based on two or more characteristics (color, size and spatial relationship).

Results showed that children from 24 to 35 months of age produce gesture and speech combination as the most frequent communicative resource to solve the task, regardless the complexity of the message to convey and child's vocabulary or morphosyntactic levels. Even when children are able to combine words in their spontaneous speech, they produce gesture-speech combinations as the main communicative resource when the task is linguistically challenging. When the message to convey increases in complexity, the semantic load tend to lean on the gesture. Moreover, children use different gesture-speech combination depending on the complexity of the message to convey (Murillo, Galera & Casla, 2015). It seems, therefore that children can flexibly adapt their gesture-speech combinations to the complexity of the message to express. Moreover, preliminary data suggests that children also adapt the characteristics of the gesture according to the type of combination produced (Murillo & Morales, 2015). As can be seen in Figure 4, the relation between pointing duration and accompanying speech duration varies depending on the type of gesture-speech combination.

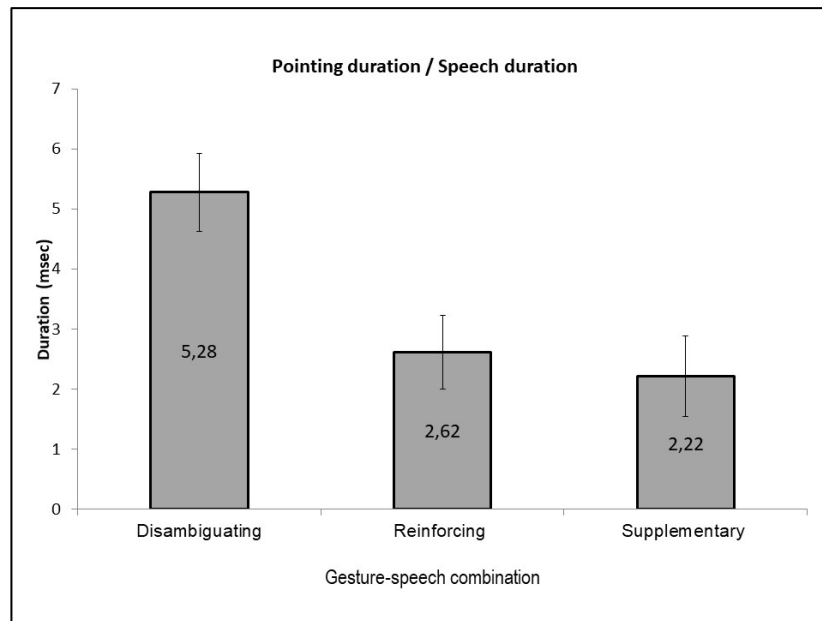


Fig.4.- Mean pointing duration according to gesture-speech combination type.

When considering stroke duration, (e.g. the duration of the main phase of the gesture), we found the same pattern: the stroke phase of the pointing gestures produced in disambiguating combinations is significantly longer than the stroke phase produced in reinforcing or supplementary combinations (see Figure 5).

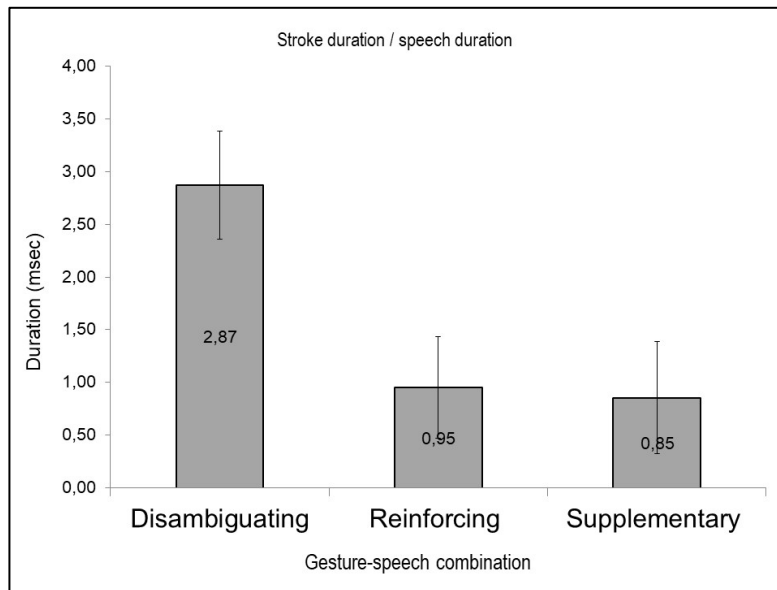


Fig. 5.- Mean stroke duration according to gesture-speech combination type.

Disambiguating combination requires that the observer look at the pointing to be understood. The target object cannot be deduced by means of the verbal message. This suggests that children flexibly adapt the duration of the gestures according to its combination with speech and not only according to the verbal message duration (Murillo & Morales, 2015).

Considering the previous research, it seems that multimodal combination does not decrease when children are able to combine words, but they learn to use these resources flexibly in a communicative way. Although much more research is needed on the change in the use of multimodal combinations towards the adult-like pattern, research suggests that gestures and speech evolve together throughout all the stages of language development.

5. Conclusions

Evidence from multimodal communication research provide some key ideas about how language is learnt and how we should consider its development. Language is a multimodal process that develops in a multimodal context. Instead of considering speech as the main focus of study, we should include in our language conception other communicative channels that are more than mere companions of the verbal modality (Vigliocco, Perniss & Vinson, 2014). They are part of the same language system from the early stages of development and they evolve together throughout the process of language learning.

The link between the gestural and vocal elements can be observed from the first intentional communicative acts throughout the process of language development. This tight relationship allows anticipating and predicting subsequent linguistic achievements, so it could be considered as an index of language development. In the very early stages of language development, the coordination of vocalizations and gestures anticipates word production. Recent studies emphasize the crucial role of gesture use in language development and its potential value as a screening tool. Lüke, Grimminger, Rohlfing, Liszkowski, and Ritterfeld, (2017) show that the index finger point at 12 months of age

is a good screening tool for primary language delay at 24 months. In the same line, Hsu and Iyer (2016) found that children at risk for language delay at 3 years of age, showed less gesture use at 15 months of age.

Gesture and speech are produced in a synchronized way, and this coordination predicts not only lexical development, but also specific syntactic achievements. Even children that are already able to combine words, produce multimodal communicative behaviors when facing linguistically challenging tasks. In sum, most of the traditional language development milestones, such as canonical babbling onset, first words production, word combination, etc., are preceded and anticipated by changes in the gestural-vocal system.

At the educational level, this multimodal conception of language invites us to observe not only what children say, but also what they do when they try to communicate with others. At every stage of language development, we can notice how this multimodal communicative system evolves for the sake of communication. The opportunities to learn language have to do not only with structural rules but also with movement and action in an interactive context, and this should be taken into account when designing educational programs. In the same way, language should be considered as a multimodal system when developing intervention programs addressed to children with language and/or communicative impairments.

6. References

- Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.
- Bloom, K., Russell, A. & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, 14, 211-27.
- Cadime, I., Silva, C., Santos, S., Ribeiro, I., & Viana, F. L. (2017). The interrelatedness between infants' communicative gestures and lexicon size: A longitudinal study. *Infant Behavior and Development*.
- Camaioni, L., Castelli, M. C., Longobardi, E. & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language*, 11, 345-59.
- Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language*, 23(3), 645-673.
- Cartmill, E. A., Hunsicker, D. & Goldin-Meadow, S. (2014). Pointing and naming are not redundant: children use gesture to modify nouns before they modify nouns in speech. *Developmental Psychology*, 50, 1660.
- Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43-71.
- Colonna, C., Stams, J. M., Koster, I. & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, 30, 352-66.
- Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, 55(4), 226-237.
- Ejiri, K. & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4, 40-8.

- Esteve-Gibert, N. & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, 57, 301-16.
- Fasolo, M. & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, 12, 1-15.
- Goldin-Meadow, S. & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 85-107. New Jersey: LEA.
- Huttunen, K. H., Pine, K. J., Thurnham, A. J., & Khan, C. (2013). The changing role of gesture in linguistic development: a developmental trajectory and a cross-cultural comparison between British and Finnish children. *Journal of Psycholinguistic Research*, 42, 81-101.
- Hsu, H. C., & Iyer, S. N. (2016). Early gesture, early vocabulary, and risk of language impairment in preschoolers. *Research in Developmental Disabilities*, 57, 201-210.
- Igualada, A., Bosch, L. & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42-52.
- Iverson, J. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, 37, 229-61.
- Iverson, J. & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16, 367-71.
- Keller, Heidi & Axel Schölmerich, (1987). Infant vocalizations and parental reactions during the first 4 months of life. *Developmental Psychology*, 23(1), 62-67.
- Lüke, C., Grimminger, A., Rohlfing, K. J., Liszkowski, U., & Ritterfeld, U. (2017). In Infants' Hands: Identification of Preverbal Infants at Risk for Primary Language Delay. *Child Development*, 88(2), 484-492.
- Masataka, Nobuo (1993a). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language*, 20(2), 303-312.
- Masataka, Nobuo (1993b). Relation between pitch contour of prelinguistic vocalizations and communicative functions in Japanese infants. *Infant Behavior & Development*, 16(3), 397-401.
- Masataka, N. (1995). The relation between index-finger extension and the acoustic quality of cooing in three-month-old infants. *Journal of Child Language*, 22, 247-57.
- Murillo, E. & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture*, 12, 16-39.
- Murillo, E. & Belinchón, M. (2013). Patrones comunicativos multimodales en la transición a las primeras palabras: cambios en la coordinación de gestos y vocalizaciones. [Multimodal communicative patterns on the transition to first words: Changes in the coordination of gesture and vocalization]. *Infancia y Aprendizaje*, 36, 473-87.
- Murillo, E., & Capilla, A. (2016). Properties of vocalization-and gesture-combinations in the transition to first words. *Journal of Child Language*, 43(4), 890-913.
- Murillo, E., Galera, N., & Casla, M. (2015). Gesture and speech combinations beyond two-word stage in an experimental task. *Language, Cognition and Neuroscience*, 30(10), 1291-1305.
- Murillo, E. & Morales, A. (2015) No todos los pointing son iguales: diferencias en el gesto de señalar en coordinación con el habla en niños de 24 a 34 meses: datos

- preliminares. *Poster presented at Jornadas de la Revista de Investigación en Logopedia. Talavera de la Reina. 12 June.*
- Murillo, E., Ortega, C., Otones, A. & Casla, M. (2017). Changes in the synchrony of multimodal communicative behaviors in early language development. *Poster presented at DCOMM Conference . Rome 21-23 June.*
- Özçalışkan, Ş. & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*, 101-13.
- Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: children learning about balance. *Developmental Psychology, 40*, 1059-1067.
- Romero, A., Etxebarria, A., de Pablo, I., & Romero, A. (2017). The interrelation of gestures and vocalization in early communication functions: Evidence from Basque language. *Revista signos, 50*(93), 96-123.
- Rowe, M. L. & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*, 182-7.
- Rowe, M., Özçalışkan, Ş. & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*, 182-99.
- Stefanini, S., Bello, A., Caselli, M. C., Iverson, J. M., & Volterra, V. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes, 24*(2), 168-189.
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a Multimodal Phenomenon. *Philosophical Transactions of the Royal Society of London, Series B, Biological sciences, 369*.
- Wu, Z. & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*, 72-90.

METODOLOGÍA EXPERIMENTAL Y TRATAMIENTO ESTADÍSTICO. APLICACIÓN AL ESTUDIO DE LA PROSODIA DEL CATALÁN

Joan Borràs-Comes
Universitat Autònoma de Barcelona
joan.borras@uab.cat

Laburpena

Este artículo introduce diferentes conceptos y estrategias sobre diseño experimental y análisis estadístico que serán de utilidad para el investigador en lingüística que se encuentre en sus primeros años de carrera. Asimismo, se ponen ejemplos de experimentos de producción y de percepción y se establecen los rasgos principales que permiten distinguir un enfoque cualitativo de uno cuantitativo. Especialmente pensando en estos últimos, se introducen diferentes técnicas de muestreo, de tipo probabilístico y no, los principios y características fundamentales de un diseño experimental y sus variaciones, y lo que entendemos por variables y cómo debemos modelarlas en el proceso de aplicar un análisis estadístico.

Gako-hitzak: diseño experimental, estadística aplicada, tratamiento cuantitativo de datos, formación para la investigación, prosodia

1. Introducció

Al estudio de la entonación catalana se le reconocen, al menos, cien años de historia. De hecho, ya en 1916, Pere Barnils publicó el artículo «De l'entonació en els nostres dialectes» en el *Butlletí de Dialectologia Catalana*, pero cabe suponer que la documentación sobre la estructura suprasegmental del catalán siguió la misma suerte que su uso público, y no fue hasta la década de 1970 cuando vieron la luz nuevos trabajos sobre la materia. Sin embargo, poder realizar en la década de 1980 un estudio sobre producción o percepción prosódica poco tenía que ver con hacerlo hoy en día, cuando, si nos lo proponemos, no necesitamos ni salir de casa para disponer de múltiples dispositivos y herramientas para su análisis y descripción. Así ha sido, pues, que en los últimos veinte años tal disponibilidad y generalización de soluciones electrónicas ha empujado hacia la realización de múltiples trabajos sobre la prosodia del catalán (e.g., Prieto y Cabré, 2015).

La generalización de estudios experimentales sobre prosodia en la última década, así como el acceso a estudios de posgrado por parte de un número significativo de estudiantes año tras año, conlleva tener que formarse adecuadamente en determinadas técnicas y procedimientos. Esas habilidades, que se relacionan con la psicología experimental y la estadística aplicada, son con frecuencia extrañas en los planes de estudios de los grados que conforman la base de conocimiento de dichos estudiantes y que constituyen la puerta de acceso a los estudios de tercer ciclo. El caso es que esas técnicas tampoco son fáciles de adquirir mediante la autoformación y la lectura atenta de artículos sobre lingüística o ciencias afines en las que hayan sido aplicadas. Como podría parecer lógico, nos podemos encontrar que los países que producen más ciencia son también los que destinan más recursos a personal dedicado a la investigación, y eso conlleva que los centros pioneros en la aplicación de determinadas soluciones analíticas y procedimentales hayan podido contratar a terceros para realizar tales tareas o a especialistas pertenecientes a otras ramas —que no siempre compartirían “lenguaje” con el resto del equipo investigador—, lo cual, en última instancia, ha podido acabar

suponiendo una falta de vocación formativa en los outputs científicos que dichos centros hayan generado.

Sin perseguir ni mucho menos la exhaustividad, el presente artículo tiene como finalidad introducir una serie de conceptos y nociones sobre diseño experimental y análisis estadístico que puedan ser de utilidad para que aquellos que se inicien en la investigación empírica en diferentes ramas de la lingüística.

2. Tipos de estudios conductuales sobre prosodia

A rasgos generales, los estudios experimentales sobre prosodia pueden clasificarse en dos tipos: los que analizan la producción y los que analizan la percepción, y es frecuente que encontremos publicaciones en las que ambos tipos de exploración aparezcan conjuntamente para que la argumentación de los autores tenga más peso. Veamos unos ejemplos de preguntas de investigación —parecidas a las que figuran en Borràs-Comes, Sichel-Bazin y Prieto (2015)— en las que el participante en el experimento tenía que imaginar que llamaba por su nombre a un interlocutor con el cual estaba relacionado con más o menos poder y distancia social (conceptos que pertenecen a la teoría de la cortesía de Brown y Levinson).

En una situación comunicativa en particular...

Producción	Percepción
¿cuánta altura tonal (Hz) utilizarías en tu enunciado?	dado un contorno entonativo particular, ¿cuán adecuado te parecería utilizarlo, en una escala del 0 al 100?
¿qué contorno entonativo utilizarías?	dado un contorno entonativo particular, ¿te parecería adecuado o inadecuado, utilizarlo?

Si comparamos las cuatro preguntas de investigación que acabamos de esbozar (que podrían desembocar perfectamente en cuatro experimentos distintos), la relación que establecen los de producción y los de percepción entre el participante y la lengua oral es netamente distinta (mientras que en unos producen lengua oral, en otros, en cambio, se pide a los participantes que se posicionen con respecto a una producción que ya estaba disponible). Sin embargo, mientras que en los experimentos que aparecen en la primera fila obtendríamos una respuesta continua (un valor en Hz; un valor en una escala del 0 al 100), en los que aparecen en la última sería categórica (un contorno entonativo concreto entre el abanico posible de contornos con que cuenta una lengua; una respuesta sin matices sobre si al participante le parece adecuado o inadecuado), lo cual tiene consecuencias importantes de cara al análisis estadístico de los datos.

Tradicionalmente, los estudios que se han encargado del estudio de las propiedades físicas-acústicas del sonido (o de los movimientos asociados a su articulación) se han llevado a cabo en el marco de estudios de producción con respuestas continuas (y se han relacionado más con el campo de la fonética que con el de la fonología). Por otra parte, los estudios que recogen la actividad eléctrica cerebral que provoca la exposición a ciertos estímulos, como la mayor parte de los psicolingüísticos y neurolingüísticos, se preocupan de respuestas continuas en el contexto de experimentos de percepción. No obstante, el análisis estadístico de unos y otros puede llegar a ser muy parecido, y es que los dos estudian respuestas de carácter

continuo, agrupables según el participante que las haya proporcionado, y provocadas por cierta estimulación oral, escrita o de otro tipo. En la siguiente sección tratamos de arrojar luz sobre diferentes aspectos de su diseño, el cual nos servirá como vehículo para que entendamos mejor su estructura.

3. Las orientaciones científicas cualitativa y cuantitativa

Muy probablemente, la tarea más frecuente de un científico del lenguaje —y quizá de cualquier científico— es explicar la variación, o, al menos, una pequeña parcela de ella. Para eso, el investigador suele utilizar un conjunto de constructos teóricos, a veces pertenecientes a otras ramas del conocimiento, con los que idea una hipótesis. El cumplimiento (o no) de esa hipótesis supone que nuestro objeto de análisis muestre un comportamiento distinto, de manera sistemática (y sin poder ser explicado, consecuentemente, por el azar). Entonces, el científico puede decir que tal tratamiento explica en parte la variación, y que por lo tanto causa una diferencia en el comportamiento (o que estructura, en alguna medida, lo que antes solo era, supuestamente, variación y desorden). Eso es lo que más o menos conocemos como método hipotético-deductivo.

Por ejemplo, en Prieto y Borràs-Comes (aceptado: 2017), los autores tratan de establecer pautas de carácter semántico-pragmático que expliquen, en alguna medida, la variación de (cuatro) contornos entonativos que tiene a su alcance cualquier hablante de catalán central para realizar un enunciado interrogativo directo. Una de sus hipótesis es si las suposiciones del hablante causan que se use más un contorno u otro. Para ello, los autores construyeron diferentes ejemplos de situaciones comunicativas en las que el hablante tuviera más o menos conocimiento sobre lo que estaba preguntando (con tres niveles, que iban desde la ignorancia hasta haber tenido percepción sensorial de un hecho relacionado). Los participantes del experimento leyeron la descripción de esos contextos y, a continuación, oyeron a un hablante produciendo uno de los cuatro posibles contornos; su tarea era evaluar, del 0 al 100, cuán adecuado les parecía haberse usado un contorno entonativo determinado en un contexto comunicativo en particular. Los resultados indicaron que el comportamiento de dos contornos, justamente, tenía mucho que ver con la manera en que habían sido diseñadas las situaciones comunicativas (pues uno de ellos era más aceptado cuando el hablante no tenía información, y el otro cuando el hablante sabía ya prácticamente la respuesta); en cambio, para los otros dos contornos solo se obtuvo una baja aceptación general.

Así, Prieto y Borràs-Comes (aceptado: 2017) tomaron como muestra las respuestas de 119 participantes y con ellas pretendían extrapolar los resultados a la población total de hablantes de catalán central, a través de un diseño experimental planificado *a priori* y mediante el uso de técnicas de análisis estadístico inferencial. Estas no son características de cualquier tipo de investigación científica, sino que constituyen algunas de las propiedades de la orientación cuantitativa en la investigación (o *nomotética*), en contraposición con la orientación cualitativa (o *ideográfica*). Petri y Govern (2012) definen la dualidad de la siguiente manera: mientras que el enfoque nomotético pretende establecer leyes universales que permitan explicar la motivación de una manera general para todos los seres, el enfoque ideográfico pretende entender la conducta motivada, examinando en qué se distinguen los individuos de una especie entre sí. La tabla siguiente aporta algunos detalles sobre las dos orientaciones a las que nos estamos refiriendo, recogidos de distintos manuales:

Orientación cualitativa	Orientación cuantitativa
La objetividad no tiene que ser necesariamente una meta	Búsqueda de objetividad, sin negar la intersubjetividad (cf. marco teórico)
En general, se lleva a cabo en entornos naturales	Suele llevarse a cabo en entornos estructurados
Ideográfica: trata de describir acontecimientos o hechos particulares	Nomotética: busca leyes generales que se puedan aplicar a muchas personas diferentes
Diseño emergente no preestablecido	Diseño planificado <i>a priori</i>
La información se produce de manera natural	Generalmente, la producción de información se provoca (se elicitada)
Generalmente, el investigador es el elemento principal para obtener información	Por lo común, la información es recogida por instrumentos estandarizados y validados
Vocación exploratoria	Vocación confirmatoria
Datos “ricos y profundos”, no generalizables	Datos “sólidos y repetibles”, generalizables
Análisis conceptual/filosófico de los resultados	Análisis estadístico
Método inductivo, comprensión, fenomenología	Método hipotético-deductivo

Cabe señalar que las dos orientaciones no son mutuamente excluyentes, sino que se pueden combinar —en lo que se conocen como orientaciones mixtas— para así utilizar las fortalezas de cada una de ellas y conseguir un resultado superior en la investigación.

De cara al tratamiento de los datos, una de las principales diferencias entre orientación cualitativa y cuantitativa es la vocación de generalización de los resultados, claramente presente en la orientación cuantitativa, pero con serias limitaciones para poder ser llevada a cabo en la cualitativa. O, más bien dicho: cómo una y otra afrontan esa vocación que muchas veces, en realidad, comparten.

Pensemos en un estudio clásico de análisis del discurso. El procedimiento de análisis es típicamente cualitativo: se lleva a cabo en un entorno natural, los datos se pueden describir con todo lujo de detalles porque fueron grabados con gran calidad, y se practica una vocación exploratoria clara por parte del investigador. No obstante, suele haber una motivación más general, como encontrar algún tipo de patrón que explique por qué, por ejemplo, en algunas oraciones que produce el hablante este mira a su interlocutor y, en cambio, en otras, no. Al investigador cualitativo y al cuantitativo les puede interesar contestar esa misma pregunta; es decir, qué hace que un hablante unas veces mire y otras no, pero enfocarán su análisis de maneras claramente distintas. El investigador cualitativo, por su parte, analizará al detalle los enunciados producidos naturalmente; fijándose en el detalle aparentemente más insignificante, para fijarse nuevamente si el hecho de que en la conversación se diera tal o tal condicionante generaba, a su parecer, un comportamiento distinto del hablante. Por otro lado, el investigador cuantitativo, después de un examen numérico de la frecuencia en la que se dan esos casos y situaciones en el corpus, tomará las conclusiones a las que había

llegado su homólogo cualitativo y las convertirá en hipótesis de un nuevo estudio. En ese estudio aparte, típicamente, el investigador preverá más control experimental (y por lo tanto menos naturalidad), un número mayor de enunciados, una frecuencia de enunciados constante para cada uno de los participantes, y unas condiciones de experimentación en las que el tratamiento numérico-estadístico de la hipótesis pueda ser abordado de manera más controlada.

Por lo tanto, la investigación cuantitativa pretende generalizar a la población total los resultados que obtenga de una parte de la población, y lo hace a través del análisis estadístico. Esta parte de la población se denomina *muestra*, y para que sus resultados se puedan generalizar a la población total deberá ser *representativa*, lo cual se consigue si se cumplen una serie de principios a la hora de seleccionar los individuos.

4. La muestra

La muestra tiene que ser representativa de la población total, para lo que existen diferentes estrategias de muestreo clasificables en dos grupos: las probabilísticas y las no probabilísticas. Dentro del muestreo probabilístico encontramos el muestreo *aleatorio simple*, en el que todas las unidades tienen la misma probabilidad de pasar a formar parte de la muestra, y el *aleatorio estratificado*, en el que, ya que nos interesa que la muestra tenga una composición análoga a la población, dividimos la población en clases o estratos homogéneos antes de realizar, dentro de cada estrato, un muestreo aleatorio simple de menor escala.

Pongamos, por ejemplo, que nos interesa saber si el aprendizaje del inglés está condicionado por el hecho de estar cursando las asignaturas del grado en ese idioma. Para eso hemos seleccionado 50 estudiantes de psicopedagogía, 50 de arquitectura y 50 de ingeniería informática. En el grado de psicopedagogía hay un 90% de mujeres, en arquitectura un 50% y en ingeniería informática solo un 10%. En este caso, el investigador debe decidir qué hacer con el factor SEXO en el contexto de su investigación, para lo que tiene dos opciones de cara a establecer el muestreo, y la elección de una u otra dependerá estrictamente de cuál sea su pregunta de investigación:

Supuesto 1: No le interesa saber si el SEXO tiene efectos sobre el aprendizaje, pero no le está de más controlar su peso de manera global. Lo que simplemente hizo fue escoger tres estudios, entre todos los posibles, y se sirve de que, globalmente, entre el total de 150 participantes en su experimento, haya la mitad de hombres y la mitad de mujeres. Considera que el SEXO no va a ser un factor que tenga que preocuparle en exceso cuando vaya a analizar los datos. Así, pues, selecciona 10 hombres y 40 mujeres en psicología, 25 hombres y 25 mujeres en arquitectura, y 40 hombres y 10 mujeres en ingeniería informática. Eso le permitirá contar con una muestra con una composición similar a la de la población total (cf. más adelante el concepto de *bloqueo*).

Supuesto 2: Sí que le interesa saber si el SEXO tiene que ver con el aprendizaje del inglés. Pongamos que tiene hipótesis muy concretas sobre la interacción entre SEXO y ESTUDIO DE GRADO, como por ejemplo que los hombres se comportarán de manera distinta con respecto a la adquisición del inglés en los grados en los que conforman una mayoría de la población y en los que quedan en minoría. Al investigador, en este caso, lo que le conviene es tomar grupos cuanto más homogéneos mejor, y seleccionará 25

hombres y 25 mujeres de cada uno de los tres grados en los que llevará a cabo el estudio.

En el muestreo no probabilístico, las muestras se seleccionan en un proceso en el que no se da a todos los individuos de la población iguales oportunidades de ser seleccionados. Como técnicas no probabilísticas se recogen el muestreo *estratégico o de conveniencia*, en el cual se seleccionan los participantes según su accesibilidad/disponibilidad para el investigador, el muestreo *por criterio*, en el que la inclusión de un participante depende de que cumpla determinados criterios (cosa que podríamos relacionar con el concepto de *screening* o cribado).

En muchos estudios lingüísticos se da que se utiliza un muestreo de conveniencia cuando la muestra se limita a estudiantes de grado de la misma institución en la que trabaja el investigador y después se pretenden generalizar esos resultados al total de hablantes de esa lengua. También es muy frecuente que se usen técnicas de cribado para establecer quién pasa a formar parte de la muestra y quien no; pensemos por ejemplo en el caso que queramos estudiar la cualidad vocálica de las vocales átonas del catalán central: tendremos que quitar de la muestra a los distintos hablantes no nativos de catalán, también aquellos que no sean hablantes del dialecto específico que nos interesa estudiar, y quizás también establezcamos un nivel de dominancia con respecto al uso diario de catalán y castellano por debajo del cual un determinado participante también quedaría excluido de la muestra.

De todos modos, hay que tener presente que en el proceso del muestreo no solo hay que aguardar aspectos de variación interna de la muestra sino, sin ir más lejos, cuántos individuos deben formar parte de ella. El tamaño muestral es crucial, tanto para la representatividad como para después tratar de interpretar los resultados estadísticos (v., e.g., Rodríguez Osuna, 1991). Por ejemplo, puede que la estadística nos diga que dos poblaciones no son significativamente diferentes, pero que esa falta de significación venga dada justamente porque no hubimos tomado suficientes observaciones de cada uno de los grupos que más tarde compararíamos. Por otra parte, un tamaño de muestra demasiado elevado podría conllevar una pérdida innecesaria de tiempo, de dinero o de otros recursos invertidos para la realización de la investigación. El establecimiento del tamaño muestral a partir de técnicas estadísticas queda fuera del alcance del presente trabajo, con lo que, para empezar, se aconseja seguir proporciones similares a las que hayan seguido anteriormente estudios ya publicados.

5. Tipos de variables

Antes de seguir, conviene definir lo que entendemos por *variable*, que es una característica analizada de un elemento de la muestra que puede asumir diferentes valores. (Si solo pudiese asumir uno, no lo llamaríamos variable sino *constante*.) Cada uno de esos valores que puede tomar se denomina *nivel*. En el ejemplo de la sección anterior habíamos hablado de la variable SEXO, y esa variable contaba con dos niveles: hombre y mujer. Una variable puede ser de diferentes tipos, y su tipología dependerá, por lo menos, de tres puntos de vista distintos referidos a sus niveles internos.

En primer lugar, los niveles internos de una variable pueden ser ordenables o no. En este caso, una variable puede ser *cuantitativa* o *categorica*. Las variables cuantitativas pueden ser *continuas*, si sus niveles pueden tomar un número infinito de

valores (p. ej., temperatura medida en °C, o superficie medida en m²) o *discretas*, si solo pueden tomar un número limitado de valores (p. ej., escalas de valoración del 1 al 5 para indicar de menos a más la adecuación contextual percibida para una serie de enunciados). Por otra parte, las variables categóricas, que también pueden recibir los nombres de cualitativas o nominales, contarían con niveles no ordenables (p. ej., diferentes tipografías con las que puedo representar un texto), o incluso con niveles ordenables en los que la distancia entre cada par de ellos no es el mismo (p. ej., escalas de valoración del tipo “muy de acuerdo, más bien de acuerdo, sin un posicionamiento claro, más bien en desacuerdo, muy en desacuerdo”).

En segundo lugar, las variables se pueden clasificar dependiendo de si han sido ‘manipuladas’ por el experimentador o de si son medidas a partir de los sujetos. Por un lado, tenemos las llamadas *variables independientes* o *factores*, dentro de las cuales tendremos que distinguir entre los *factores fijos* y los *factores aleatorios*, que se describirán más adelante en la sección sobre análisis estadístico y que se diferencian, en primer lugar, porque sus niveles internos hayan sido elegidos conscientemente o, por el contrario, fueron determinados por el azar. Por el otro lado, las *variables dependientes*, que suelen ser aquellas medidas del experimento que el investigador analiza como respuesta a una determinada pregunta (esta última determinada a partir de los factores).

En tercer lugar, las variables también se pueden clasificar dependiendo de si sus niveles se examinan usando uno o varios grupos de sujetos (lo que llamamos *estrategia de comparación*). Así, podemos distinguir una variable *intra-sujetos* o *de medidas repetidas* cuando un solo sujeto nos proporciona más de un valor en nuestra base de datos (p. ej., cuando en un concurso culinario un mismo juez da valoración a varios platos) y una variable *inter-sujetos* cuando un sujeto solo nos aporta un valor (p. ej., cuando indago si la lengua materna de los sujetos permite predecir una diferencia en la respuesta: cada sujeto solo puede tener una lengua materna, o si estuviera haciendo una encuesta electoral en la que cada participante solo me proporcionara una respuesta referida al partido que piensa votar si en aquel momento se celebrasen unas elecciones).

6. El diseño del experimento

Existen diferentes tipos de aproximaciones cuantitativas a la investigación (v., e.g., Portell Vidal y Vives Brosa, 2014): experimentales, cuasiexperimentales, *ex post facto*, etc., pero la que probablemente tenga un uso superior en la investigación actual en prosodia y fonología sea la primera de ellas. Un experimento es un procedimiento mediante el cual testamos una hipótesis causal (o varias), y eso es una cuestión de suficiente importancia como para que a un experimento se le pida una gran validez interna y control. Con *diseño* de un experimento nos referimos a la estructura lógica que sigue, y esta se caracteriza por dos hechos: la manipulación de la variable independiente (VI) y la aleatorización.

La manipulación de la variable independiente (VI) tiene que ver con que el investigador crea las condiciones para que tengan lugar ciertos niveles de la VI, y que esos niveles se aplicarán a los participantes antes de medir la variable dependiente (VD). Es decir: el investigador manipula los niveles de una variable TIPOGRAFÍA cuando para examinar su incidencia en la legibilidad decide incluir los niveles específicos *Arial*, *Times New Roman*, *Courier New* y *Lucida Handwriting*. Manipular una VI implica también determinar atributos como la intensidad, la frecuencia o la duración de los

niveles de la VI, establecer el momento en el que se aplicarán esos niveles a los participantes, y decidir qué participantes recibirán un nivel u otro de la VI.

La aleatorización se refiere a la asignación aleatoria de los participantes a las diferentes condiciones experimentales (si es inter-sujeto) o a los diferentes órdenes de presentación (si es intra-sujeto). La razón de ser de la aleatorización es procurar que las variables extrañas (conocidas o desconocidas), aquellas variables que o bien no se prevén o bien no forman parte de los constructos teóricos de nuestras hipótesis, queden suficientemente equilibradas en las diferentes condiciones experimentales.

Los diseños experimentales podrán ser de diferentes tipos: según el número de VD que midan (univariadas, multivariadas), según el número de VI que incluyan (unifactorial, [multi]factorial), según la estrategia de comparación que utilicen para cada VI inter-sujeto (intra-sujeto, mixto), y según el tipo de técnica de control —relacionada específicamente con la aleatorización— que apliquen, las cuales pasan a ser descritas a continuación.

En una investigación se pueden distinguir dos tipos de técnicas de control: las basadas en alguna forma de aleatorización, que son específicas del diseño experimental, y las que no se basan en la aleatorización, que son además aplicables a diseños no experimentales. Entre las que se relacionan con la aleatorización se incluye, en primer lugar, la *asignación al azar*, o asignación aleatoria sin restricciones, en la que los individuos de la muestra se distribuyen aleatoriamente entre los diferentes grupos participantes en un experimento.

En segundo lugar, se puede hacer uso del *bloqueo*, un procedimiento en el que se forman grupos homogéneos de individuos en función de los valores obtenidos en una o más variables, llamadas *variables de bloqueo*, relacionadas con la VD. Después, una vez creados los bloques, los sujetos de cada bloque se asignan aleatoriamente a las diferentes condiciones experimentales. Pensemos, por ejemplo, que en un diseño experimental cuento con un grupo que recibirá un tratamiento y otro que recibirá un placebo (tipos de grupos que reciben los nombres de *experimental* y *control*, respectivamente). Como investigador, puede parecerme que habiendo asignado aleatoriamente los participantes de la muestra a cada uno de ellos ese hecho automáticamente hará que los grupos sean suficientemente comparables entre sí para poder medir el efecto del tratamiento. Sin embargo, en muchas investigaciones no tiene por qué ser el caso, porque, por ejemplo, las habilidades cognitivas de un grupo y otro podrían ser demasiado diferentes, porque un grupo contara de media con un coeficiente de memoria de trabajo (v. *working memory*) superior al otro, o porque uno y otro tuvieran una distribución demasiado dispar de hombres y mujeres como para asegurar que eso no pudiera estar ejerciendo de *factor de confusión* (v. *confounding factor*). Es decir: Si los grupos me proporcionan resultados significativamente diferentes, ¿cómo sé que es a causa de mis factores fijos y no, en cambio, de las diferencias morfológicas que existen entre ellos dos? La técnica del bloqueo hace que las oportunidades de que tenga que hacerme esta pregunta se reduzcan considerablemente.

En tercer lugar, otra estrategia de control es el *contrabalanceo*, la cual se aplica en diseños con un componente intra-sujeto. En este caso, el sujeto mismo actúa como procedimiento de control, puesto que la asignación aleatoria no solo se aplica a los individuos de la muestra, sino también a los diferentes estímulos o ítems que tiene que

recibir cada sujeto. El uso de este procedimiento conlleva, lógicamente, decidir una estrategia concreta para el orden de presentación de los ítems y condiciones, entre las cuales tendremos la realización de un contrabalanceo completo o incompleto y diferentes estrategias concretas como los diseños en cuadrado latino (v. *latin square design*).

Las técnicas de control no basadas en la aleatorización incluyen por lo menos tres procedimientos. En primer lugar, la *eliminación*, que supone mantener la variable que se quiere controlar en su valor nulo. Pensemos, por ejemplo, en un experimento que plantea evaluar cómo se interpretan los gestos de un hablante mientras produce una serie de enunciados. En un primer momento contamos con grabaciones que incluyen tanto información visual como acústica. Una técnica de control basada en la eliminación consistiría en testear, aparte, una condición experimental unimodal, es decir, la misma serie de estímulos en los que hubiéramos eliminado toda la información acústica (para poder asegurarnos que lo que realmente provoca que la respuesta varíe es la gestualidad). En segundo lugar, disponemos de la *constancia*, según la cual una variable tomaría el mismo valor para todos los participantes. Por ejemplo, si pretendo estudiar cómo afecta la entonación a la interpretación de unas frases, quizá me interese mantener la misma estructura sintáctica para las dos condiciones, para así después poder decir que la sintaxis no tuvo un efecto crucial en las diferencias que encontrara entre las dos condiciones y que estas eran causadas en efecto por la variación entonativa. En tercer lugar, también dispondría de la técnica del simple-ciego o doble-ciego, según la cual el participante (o además el investigador) no sabría, en el momento de la experimentación, si se le estaría administrando una intervención activa o un placebo, es decir, si pertenecía a un grupo de control o a un grupo de tratamiento.

7. La estadística. Conocer nuestras variables

Se pueden describir dos grandes tipos de estadística, la *descriptiva* y la *inferencial*. La estadística descriptiva tiene como objetivo resumir el conjunto de datos que nos ha proporcionado la muestra. Dentro de ella contamos con parámetros de *posición* (promedio, mediana, moda...), parámetros de *dispersión* (rango, variancia, desviación típica o estándar...) y parámetros de *forma* (asimetría, apuntamiento o curtosis). La segunda de ellas, la estadística inferencial, en cambio, nos permite deducir si las características que pueda obtener de la muestra son extrapolables a la población total.

En el contexto de la estadística inferencial es muy importante el concepto de *modelo estadístico*. Un modelo estadístico es una ecuación matemática que parte de los datos suministrados y de la influencia del azar en estas observaciones para reproducir los fenómenos que observamos en nuestros datos de la forma más exacta posible. O, dicho de otro modo: un modelo estadístico es una aproximación a la realidad de los datos en forma de predicción idealizada.

Lógicamente, el modelado de los datos requiere tener en cuenta los distintos aspectos de nuestro diseño experimental; fundamentalmente: cuáles son las variables dependientes e independientes y de qué tipo son (es decir, qué distribución de probabilidad siguen). La distribución de probabilidad puede ser *normal* (valores reales), *exponencial* o *gamma* (valores reales positivos), *Poisson* (valores enteros), *Bernoulli* o *binomial* (solo dos valores; típicamente 0 y 1), *ordinal* (categorías ordenables), o

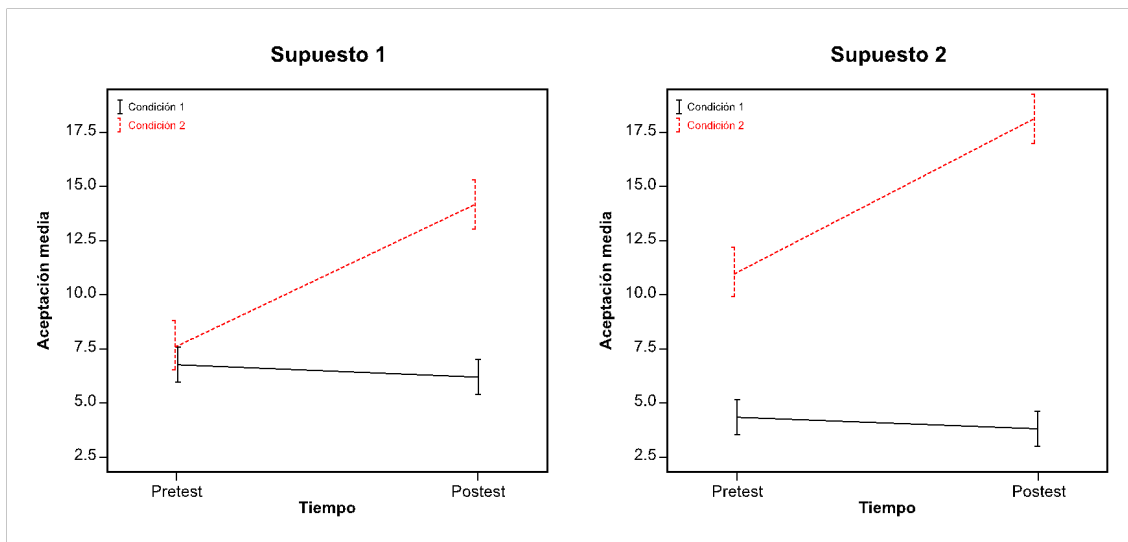
categorica, también llamada *nominal* o *multinomial* (categorías no ordenables). El ejemplo más típico de investigación es aquel en el que se dispone de una o varias variables independientes que siguen una distribución categórica de dos o más niveles y una variable dependiente con distribución normal, de recuento (Poisson) o binaria. Sin embargo, antes de poder poner ejemplos de diferentes configuraciones aún nos quedan conceptos para establecer.

Como ya indicamos dos secciones atrás, existen dos tipos de variables independientes según su composición interna y el modelado que reciben: los factores fijos y los factores aleatorios.

7.1. Factores fijos

Un *factor fijo* (*fixed factor*) es una variable para la que se prueban varios niveles con la finalidad de ver cómo influencia en los resultados. Sus niveles son considerados de interés específico en la investigación y son seleccionados a consciencia. Hay que tener en cuenta que las conclusiones que obtengamos de un factor fijo son específicas de sus niveles: es decir, que una lengua tenga una estructura silábica más compleja que otra no nos dice nada sobre las lenguas que no hayamos examinado.

Conviene saber que hay dos clases de efectos asociables a los factores fijos: los efectos principales (*main effects*), referidos a la relevancia de una variable por sí misma para explicar una diferencia de resultados de la VD, y las interacciones, según las cuales existe una interacción cuando el efecto de un factor es diferente según el nivel en el que se encuentre otro factor. Circula una especie de leyenda urbana según la cual, cuando vayamos a interpretar los resultados, si hallamos que una interacción es significativa eso implica que tengamos que obviar la descripción de los efectos principales o que estos carecen de relevancia. Nada más lejos de la realidad. Considérese el gráfico siguiente, en el que se muestran dos resultados posibles para un experimento en el que valorábamos la aceptación media de una serie de estímulos pertenecientes a dos CONDICIONES (Condición 1, Condición 2) examinados en dos TIEMPOS distintos (Pretest, Postest). Tanto en un supuesto como en el otro, obtenemos una interacción significativa CONDICIÓN×TIEMPO, que nos indicaría que la diferencia de TIEMPO sería relevante para la Condición 2, pero no para la Condición 1. Si mi descripción se limitara a la que acabo de hacer, me valdría para describir tanto el Supuesto 1 como el Supuesto 2, y no es que ellos dos sean equivalentes, puesto que el Supuesto 2, además, muestra un efecto principal de CONDICIÓN: la Condición 2 es superior a la 1 tanto en el Pretest como en el Postest, mientras que en el Supuesto 1 ese efecto solo se da en el Postest. Así, mientras que en el Supuesto 1 solamente son relevantes los resultados de la interacción, en el Supuesto 2, además, debo añadir los del efecto principal de CONDICIÓN.



Una covariable (*covariate*), que es aquella variable que sigue una distribución continua y para la cual también se espera que varíe de acuerdo con los factores, también sería considerada como fija (hablaríamos de una *covariable fija*), en tanto que sus efectos también interesan al investigador. Imaginemos que realizamos un experimento en el cual el participante tiene que establecer cuán adecuadas ve unas determinadas entonaciones cuando son insertadas en unos contextos comunicativos determinados. Podríamos considerar que, en este experimento, hay unas condiciones que son más difíciles que otras, y que un reflejo de esa dificultad se podría obtener mediante el análisis del tiempo de reacción (es decir, el tipo que tarda el participante en dar su respuesta). En este caso, nuestra VD puede continuar siendo la adecuación reportada por el hablante, pero uno de mis predictores puede el TIEMPO DE REACCIÓN, modelado como covariable fija, para más tarde poder saber si la dificultad de dar una respuesta tenía alguna influencia en la respuesta que se escogió finalmente.

7.2. Factores aleatorios

Un *factor aleatorio* (*random factor*) es una variable independiente cuyos niveles fueron asignados aleatoriamente. El investigador quiere controlar cuánta variación causa; puede interesarle cuantificarla o simplemente tenerla en cuenta para que no influya en los resultados de los efectos fijos. Sobre todo, lo que le interesa al investigador es generalizar a la población total las conclusiones extraídas por ese factor.

El caso más típico es el de la variación asociada a los diferentes individuos que conforman la muestra (*variación de sujeto*), o, por ejemplo, el tipo de variación que iría asociada al conjunto de elementos léxicos seleccionados para estudiar el efecto de una estructura sintáctica específica o del tipo de contorno entonativo que asignó (*variación de ítem*): en este caso el investigador querrá también generalizar sus hallazgos al conjunto de palabras de una lengua o a todas las palabras que sean de un tipo gramatical determinado, por ejemplo. Son aleatorios en el sentido que son miembros aleatorios del conjunto posible de enunciados o participantes. Son variables en el sentido que podrían influenciar los resultados del estudio, pero su efecto no tiene un interés primario para el investigador: no importa qué participante o qué ítem en particular obtuvo puntuaciones menores dentro de su categoría. Otras relaciones de anidamiento (como estudiantes

pertencientes a diferentes clases) también son factores aleatorios si el propósito del estudio no es ver qué diferencias hay entre clases, familias, etc.

Sobre los factores aleatorios hay que tener también en cuenta que si ese factor solo tiene dos niveles (aunque lo creamos aleatorio y no nos interese de cara a la confirmación de alguna de nuestras hipótesis), hará falta que lo especifiquemos como factor fijo, puesto que su variancia no es calculable. Este siempre es el caso del SEXO, para el cual, consecuentemente, habrá que decidir si lo modelamos como factor fijo o si, directamente, no lo incluimos en el modelo.

8. La necesidad de un modelo que tenga en cuenta todo lo importante

Según el diseño de nuestro experimento, según el tipo de variables independientes que diseñamos para llevarlo a experimentación, deberemos aplicar un modelo de efectos fijos, de efectos aleatorios o de efectos mixtos (cuando incluyan ambos tipos de efectos).

Hay diferentes pruebas estadísticas que son recurrentes en la literatura científica sobre lingüística, como las pruebas *t* de medidas independientes, los ANOVA o análisis factoriales, o las regresiones lineales múltiples. Todos ellos asumen que los datos son independientes, es decir, que un participante no podría aportar más de una respuesta al conjunto de datos, ni siquiera una respuesta para una de las condiciones distintas que se examinen. Eso implica que diseños experimentales como los intra-sujetos, los longitudinales o los pretest-postest, en los que cada participante aporta más de una respuesta por condición, deban buscar soluciones alternativas para el tratamiento de sus datos. De los pocos que escapan a esta descripción y que sí permiten medidas repetidas son las pruebas *t* por parejas, lo cual los hace similares a los modelos de efectos mixtos.

Se puede hacer una clasificación bastante clara de diferentes tipos de modelos univariantes (aquellos que tienen una sola VD) atendiendo a la distribución de esta VD y al hecho de que tengan o no factores aleatorios además de fijos. La base de todos estos modelos es la regresión lineal, por lo que, si pretendemos computar un modelo en el que la VD no tenga una distribución normal, en el modelo se aplicará lo que se denomina una función de enlace (*link*). Las más utilizadas son las Logit o Probit, para distribuciones Binomiales (binarias), y las Log, para distribuciones Poisson (de recuento). Los modelos que no necesitan aplicar una función de enlace se conocen como *modelos lineales generales*; los que sí lo necesitan, *modelos lineales generalizados*.

	Distribución de la variable dependiente (VD)	
	Normal	No normal (+ <i>link</i>)
Solo factores fijos	General Linear Model	Generalized Linear Model
Fijos y aleatorios	General Linear Mixed Model (LMM)	Generalized Linear Mixed Model (GLMM)

Todos estos modelos tienen propiedades en común. Su estructura es la misma: todos son modelos en los que se establece una relación entre una única VD y uno o más predictores; con la variación del modelo quedando establecida en los residuales. Algo realmente útil es que mediante el uso de una misma interfaz de análisis en el paquete

estadístico que utilicemos (SPSS, STATA) el mismo procedimiento nos valdría para calcular cualquiera de ellos mientras especificáramos en su interior las opciones pertinentes.

Cuando los participantes de nuestro experimento nos proporcionan respuestas, estas están agrupadas (*clustered*) según el individuo que las haya proporcionado. Se espera que las repuestas de un mismo individuo tengan alguna propiedad en común; se espera que estén correlacionadas. La puntuación de una persona en un pretest es muy poco probable que sea independiente de la que obtenga en el postest. Una persona con una puntuación relativamente alta en el pretest seguramente tendrá también una medida alta en el postest, y esto tiene que ser tomado en consideración. De modo similar, en un experimento de tiempo de reacción, algunas personas pueden ser más rápidas que otras, con lo cual se espera una correlación entre cada sujeto y sus respuestas. Además de la agrupación por sujetos, puede haber otros tipos de agrupaciones: miembros de una misma familia, estudiantes de una misma clase, los datos recogidos por un científico en el marco de una investigación en la que han participado varios, ítems experimentales que pertenezcan a un mismo subgrupo o bloque, etc. Cuando los datos se obtienen de manera natural, sin un control experimental rígido, es completamente esperable que algunas personas proporcionen más datos que otras; si un participante es especialmente hablador, eso podría provocar que influenciara demasiado los resultados globales de nuestro estudio si no tenemos en cuenta la relación entre enunciado y hablante cuando vayamos a analizar los datos. Por razones como estas es importante establecer correctamente cuáles son los factores aleatorios.

En el contexto de un modelo estadístico, llamamos *residuos* o *errores* a las distancias entre los distintos puntos representados en los datos y la predicción del modelo (el cual, en una regresión lineal, por ejemplo, representaría esta predicción mediante una línea recta). Es decir: los residuos constituyen aquella parte de variabilidad de los datos que el modelo ha sido incapaz de capturar. Es lógico pensar, pues, que aquellos modelos con valores residuales menores constituirán modelos mejores, pues su predicción se ajustará en mayor medida a los datos reales. Los residuos se pueden reducir en gran parte gracias al modelaje adecuado de los participantes y de los ítems experimentales.

En un modelo paramétrico, la variable dependiente está vinculada a las independientes a través de una ecuación matemática (el modelo en sí) que implica cantidades denominadas parámetros del modelo. En una regresión lineal simple, la base matemática de muchos modelos, esos parámetros son la intercepción (*intercept*) y la pendiente (*slope*). La interceptación y la pendiente son propiedades de cada uno de los factores aleatorios que se definan en un modelo mixto.

Si el modelo contiene interacciones triples, los efectos principales y las interacciones de orden menor deben estar presentes en el modelo para que la interacción triple esté especificada correctamente, y cabe tener presente que los diseños jerárquicos tienen efectos anidados (*nested effects*), como la variación de sujeto dentro de un grupo (pacientes anidados en doctores, doctores anidados en hospitales, etc.). Comparados con los modelos de efectos repetidos más sencillos (*repeated measures ANOVA*), los GLMM proporcionan una serie de ventajas: (1) admiten variables dependientes que no sigan una distribución normal, (2) pueden trabajar con valores perdidos, (3) especifican y computan correctamente los efectos aleatorios, (4) correlacionan los errores y (5)

permiten que los términos de error exhiban variabilidad no constante, lo cual flexibiliza el modelaje de la variable dependiente.

La especificación de los GLMM en paquetes estadísticos como SPSS queda fuera del alcance del presente trabajo, y sería necesario dedicar bastantes más secciones para que quedaran claros diferentes conceptos y procedimientos de validación antes de poder afrontarla con suficiente seguridad. Todo eso queda pendiente para futuros trabajos sobre estadística aplicada a las ciencias del lenguaje, por lo que, por el momento, se recomienda al lector que quiera profundizar en su aplicación el manual de Eddington (2015) para SPSS, *Statistics for linguists*, y los trabajos de Winter (2013) y Winter y Wieling (2016) para su aplicación en R.

9. Referencias

- Borràs-Comes, J., Sichel-Bazin, R., & Prieto, P. (2015). Vocative intonation preferences are sensitive to politeness factors. *Language and Speech*, 58(1), 68–83.
- Eddington, D. (2015). *Statistics for Linguists. A Step-by-Step Guide for Novices*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics*, 65, 145–159.
- Petri, H. L., & Govern, J. M., (2012). *Motivation: Theory, research, and application* (6^a ed.). Belmont, CA: Wadsworth.
- Portell Vidal, M., & Vives Brosa, J. (2014). *Introducció als dissenys experimentals, quasiexperimentals i ex post facto*. Col·lecció «Materials», 229. Cerdanyola del Vallès: Servei de Publicacions de la Universitat Autònoma de Barcelona.
- Prieto, P., & Borràs-Comes, J. (aceptado). Question intonation contours as dynamic and gradient epistemic operators. *Natural Language & Linguistic Theory*.
- Prieto, P., & Cabré, T. (Eds.) (2013). *L'entonació dels dialectes Catalans*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Rodríguez Osuna, J. (1991). *Métodos de muestreo*. Primera reimpresión: 2001. Colección «Cuadernos Metodológicos», 1. Madrid: Centro de Investigaciones Sociológicas.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499*.
- Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, 1(1), 7–18.

INFORMATZAILEEN SAILKAPENERAKO ARAZO METODOLOGIKO BATZUEN GAINEAN¹

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

inaki.gaminde@ehu.eus, naia.eguskiza@ehu.eus, a.romero@ehu.eus,
aintzane.etxebarria@ehu.eus

Laburpena

Azken urteotan euskararen azterketetan gero eta maizago erabiltzen dira zenbakizko metodoak. Arlo honetan ezin uka daiteke sano aurreratu dena: dialektometriak metodo berriak ekarri ditu eta horiei esker hizkuntzaren barietateak beste era batera sailkatzen dira. Metodo horien artean multzokatze-azterketa da arrakastatsuenetakoa. Ez ditugu eztabaidatuko hemen metodo horren abantailak, hain zuzen ere, metodoa bera aplikatu aurretik sumatu ditugun arazo metodologiko batzuk agerian utzi nahi genituzke eztabaidarako. Gure lana aurreko lan batzuen ekarpenetan oinarritzen da (Léonard et al., 2015 eta Gaminde et al., 2016a) eta datuen prestakuntzan jarraitu beharreko urratsez arduratuko gara. Bestalde, erkatzen diren ezaugarriak matrize dikotomikoen bidez aztertzen direnean, datu guztiak maila berean ezartzen dira. Gure ustez datuen modelizazio linguistikoa ostean, modelizazio matematikoa bidez, ezaugarri bakoitzari agertzeko daukan probabilitatea esleitu beharko litzaioke. Era honetara datuak doitu geratuko lirateke gero multzokatze azterketa egiteko. Gure azterketa honetarako Larrabetzuko datuak erabiliko ditugu (Etxebarria et al., 2016) eta metodo biak, tradizionala eta proposatzen duguna, erabilia balizko aldeak erakutsiko ditugu.

Hitz-gakoak: bariazio geolinguistikoa, metodologia, dialektometria, euskara

1. Sarrera

Hizkerak sailkatzeko metodo kuantitatiboak gurera etortzen hasi zirenetik hona (Aurrekoetxea, 1992, 1995) nolabaiteko arrakasta izan dutela pentsa daiteke, gero eta maizago aplikatu direlako (Aurrekoetxea, 2003, 2004, 2005, 2006, 2008, 2009; Aurrekoetxea eta Ormaetxea, 2006, 2010; Aurrekoetxea et al., 2014a,b; Gaminde, 2007, Iglesias, 2014, Ensunza, 2015, Lujanbio, 2016). Metodo hauek hizkerak sailkatzeko ez ezik, informatzaileak sailkatzeko ere erabili dira (Gaminde et al. 2012, 2016a). Beste arlo batzuetan ere zenbakiak askotan ikusten ditugu eta horrenbestez euskal azterketak beste dimentsio bat hartzen ari direla esan dezakegu.

Hizkeren aldakortasuna ikertzeko, estatistikatik datozen baliabide eta teknika asko daude eta batzuk oso erabilgarriak izaten dira. Gure artean gehien erabiltzen direnak hiru multzo nagusitan banatu daitezke: Deskriptiboak, Inferentzialak eta Sailkatzaileak.

Estatistiko deskriptiboetan gehien erabiltzen direnak, maiztasunak, batez bestekoak eta desbiderapenak ditugu. Inferentzietan khi karratua, studenten t, Anova, Mannen U, Kruskal-Wallis H, Pearsonen korrelazioa, Erregresioa eta aldakuntza-koefizientea sarritan agertzen dira. Azkenik, sailkatzaileetan multzokatze-azterketa da gehien erabiltzen dena, azken aldion dimentsio aniztun eskala ere ikusten hasi den arren. Hemen multzokatze azterketa egitean kontuan har daitezkeen gorabehera batzuekin arduratuko gara bereziki.

¹ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

Estatistikaren eta matematikaren erabilerak lagundu diezagukete aurreikuspenak egiteko eta ulertzeko zein den hizkuntzaren aldakortasunaren norabidea. Gisa honetara hizkuntzaren dinamika sistema konplexuen dinamikarekin parekatzen dugu (Beckner et al., 2011; Ellis, 2011; López Rivera, 2013; Moreno Cabrera, 2008). Sistema konplexuen eta korapilatsuen arteko aldea ikusteko, Monterok (2017) burmuinaren konplexutasuna ulertzeko darabilen analogia ikus daiteke. Hau onartzen badugu, sistema horiek aztertzeko proposatu den metodologia bera erabil daiteke (Léonard et al., 2015). Hau da, datu multzoa oinarritzat hartuta, lehenengo eta behin datuen azterketa linguistikoa egiten da, geroago azterketa matematikoa eta beronen bidez datuen azpian ezkutuan dautzan egiturak azaleratzen dira. Alta, gure ikerketetan susmatu ditugun arazo metodologiko bi aipatu gura genituzke aurrera egin baino lehen; bata, laginaren arazoa da eta, bestea, neurriarena.

Estatistikaren erabileran askotan aipatzen da lagina zoriz hartu beharra (Herrera et al., 2003; Díaz de Rada, 2009; Etxebarria, 2011; Murillo eta Martínez-Garrido, 2012, Romero, 2016; besteak beste). Aldakortasunaz egiten diren lan gehienetan laginak ez dira zoriz hartzen; gehien jota, laginketa estratifikatuak izaten dira eta askotan euren denbora eta jakitatea emateko prest egoten diren informatzaileekin lan egiten da baldintza zehatz batzuk jarrita. Lagina horrela jasotzeak lanaren emaitzak baldintzatzen ditu eta hori ezinbestean eduki behar da gogotan emaitzak interpretatzean.

Bestalde, neurriaren zehaztasunaren arazoa dugu. Esaten da ze (Gell-Man, 1994), objektu edo gertakari bat behatze hutsak objektuaren edo gertakariaren egoera aldarazten duela. Honi behatzailearen paradoxa esaten izan zaio (Labov, 1972) eta ekiditeko edo apaltzeko ahaleginak egin diren arren, ez dirudi guztiz ekidinezina denik² (Gell-Man, 1994). Beraz, probabilitateen eremuan mugitzen gara eta horren ondorioz errakuntza marjina bat onartu beharra dago.

Aurreko guztia kontuan hartuta, ez dugu multzokatze azterketaren baliagarritasuna kolokan jarri nahi. Gure helburua da beronen erabileraren kritikatik abiatuta eztabaidarako proposamen metodologiko batzuk egitea. Gure proposamena egiteko Larrabetzuko aditz morfologiaren datuak erabiliko ditugu (Etxebarria et al., 2016). Euskal aditzaren gainean lan asko egin den arren, ez dira hainbeste aldagai sozialen eragina kontuan hartzen dutenak (Laka et al., 2008, Gaminde et al., 2012, Etxebarria et al., 2016).

Gure aurreko lanean (Etxebarria et al., 2016) esaten genuen bezala, azterketa linguistikoa itxurazko desberdintasunetatik harago doa eta agerian jartzen ditu sakoneko aldeak eta kidesunak, berau ezinbestekoa da azterketa kuantitatiboari ekiteko eta ulertzeko zer den aldatzen dena.

Lana bost ataletan banatuta aurkezten dugu. Sarrera honen osteko bigarren atala corpora osatzen duten 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 Kontuan hartu behar dira probabilitateak kalkulatzeko zenbakien dezimalak biribiltzean egiten diren zehaztasun ezak.

2. Corpora eta metodologia

Gure lanaren helburuak betetzeko aztertu dugun corpusaren ezaugarriak aurkeztuko ditugu atal honetan (ikus Etxebarria et al., 2016). Gure azterketan 24 informatzaile izan ditugu, 12 gizona eta 12 andrazko hiru belaunalditan banatuta. Nagusiak 1941-1960 urteen artean jaiotakoak izan dira, helduak 1961-1980 urte artean eta gazteak 1981-2000 urteen artean.

Datuak jasotzeko galdekizunak erabili ziren eta Marantz PMD610 izeneko grabagailuarekin eta kanpo mikrofono batekin jaso dira. Grabaketak eginda, “Praat” (Boersma eta Weenink, 2016) izeneko programaren bidez transkribatu eta etiketatu dira datu basea osatzeko.

Denetara 19 adizki laguntzaileen ordainak jaso dira. Adizkiak ohiko sailen arabera sailkatuta, ondoko era honetara banatzen dira (euskara estandarraren adizkiak erabiltzen dira etiketa gisa):

Nor-Nori (4): “zaio”, “zait”, “zaizkigu” eta “zitaizkion”

Nork-Nor (10): “dute”, “ditut”, “ditu”, “dituzte”, “naute”, “gaituzte”, “ninduen”, “nituen”, “zituen” eta “genituen”

Nork-Nori-Nor (5): “dizkiot”, “zidan”, “genion”, “zenion” eta “nizkion”

Datuen prozesaketan lehen urratsa adizkien azterketa linguistikoa banan-banan egitea izan da. Hurrengo urratsean matrizeak osatu dira; kasu batean matrize dikotomikoa egin da, baina bestean, matrizea egin orduko adizki bakoitzaren probabilitateak kalkulatu dira. Azkenik, multzokatze azterketa hierarkikoa egin da SPSS programaren bidez; horrela distantzia matrizeak eta dendogramak lortu dira.

3. Azterketa linguistikoa

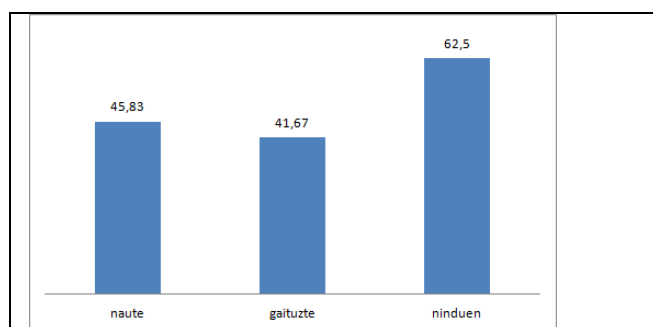
Gure aurreko lanean egin genuen azterketa xehea ez dugu berriz errepikatuko hemen (Etxebarria et al., 2016). Azterketa linguistikoaren gainean han irudikatu genituen ohar garrantzitsu bi egin gura ditugu; batetik, azterketa linguistikoa ezinbestekoa da ezaugarri bakoitzaren probabilitatea kalkulatzeko eta, azaleko formatik harago, aldagai sozialen arabera zein joera gailentzen den jakiteko; bestetik, ezaugarri horien garrantziaren arabera bakoitzari esleia dakioke “pisu” zehatz bat Léonard et al.-en (2015) lanean egiten den bezala. Aitortu behar dugu ezin izan dugula kalkulu hau egin, ez dugulako argi ikusten doitzeko ezaugarri bakoitzari eman beharoko litzaiokeen garrantziaren arabera pisua zein den. Aipatu dugun lan horretan *had hoc* egineko eskala bat erabiltzen da baina ez diogu zehatzegi iritzi gure lan honetan aplikatzeko.

Adizki baten aldakien artean egon daitezkeen aldaketak era bitakoak izan daitezke; batzuk morfologikoak eta beste batzuk fonologikoak eta fonetikoak. Alde morfologikoan morfema desberdinak erabiltzea, adizki batzuk ordezkatzeta eta morfemen kokagunea aldatzea da garrantzitsuena. Alde fonologikoan arauak betetzea edo aldatzea daukagu. Azkenik, alde fonetikoetan kontsonanteen eta bokalen aldaketak, gehitzeak eta kentzeak aipa daitezke.

Horrela bada, Nor-Nori sailean bost ezaugarritan agertzen zaizkigu aldakiak; euretako hiru morfologikotzat har daitezke eta beste biak fonetikotzat. Morfologikoak

“niri” eta “guri” morfemen aldakiei dagozkie; “niri” denean hiru aukera agertu dira “-t” (“yat”), “-te” (“yate”, “date”) eta “-ste” (“daste”), “guri” denean bi aukera daude “-ku” (“yaku”) eta “-sku” (“dasku”). “zitzaizkion” adizkiaren kasuan hasierako morfema “Ø” (“akosen”), “y” (“yakosen”) eta “d” (“dakosen”) izan daiteke. Aldaketa fonetikoetan orain aldiko adizkien oinaren hasieran “d” (“dako”) edo “y” (“yako”) ager daitezke. Bestalde, iraganeko adizkietan “-n”ren aurreko bokala “a” (“akosan”) edo “e” izan daiteke (“akosen”).

Nork-Nor sailean aldaketak 12 ezaugarritan gertatzen dira, horietako 6 morfologikoak dira, 1 fonologikoa eta gainerako 5ak fonetikoak. Aldaketa morfologikorik garrantzitsuenak analogiaz sortuak dira. Horrela, sail honetako adizki batzuekin joera handia dago sail honetako ordez Nork-Nori-Nor sailekoak erabiltzeko, paradigma aldaketa “naute”, “gaituzte” eta “ninduen” adizkiekin gertatzen da batez ere. Aldaketa ez da hiru kasuetan neurri berean gertatzen, “naute”rekin % 45,83 kasutan paradigma aldaketa gertatzen da, “gaituzte”rekin % 41,67 kasutan eta “ninduen”ekin % 62,5 kasutan (1. irudia).



1. irudia: Paradigma aldaketaren ehunekoak adizki bakoitzeko

Subjektuaren leku-aldaketa kasu bakar batean baino ez zaigu agertu (“dusen” “gendusen”en ordez). Pluralaren morfemaren erabileran hiru aukera dauzkagu: “-it” (“deitue”, “situen”, “gaitue”), “-s” (“deus”, “neusen”, “seusen”) eta pluralgile biak batera (“deitues”, “neitusen”, “seitusen”). Pluralgileak “ie” (“deurie”, “neurie”) eta “e” (“deure”, “neure”) izan daitezke. “zituen” adizkiarekin iraganaren morfema “Ø” izan daiteke (“eusen”) edo “s-” (“seusen”). Azkenik, iraganen paradigma-aldaketa egiten denean, morfema “y-” (“yosten”), “d-” (“dosten”) eta “Ø” (“osten”) izan daiteke.

Arau fonologikoetan aldaketa bakarra “-t” morfemari dagokiona da. Informatzaile batzuek(1) multzoko arauak dauzkate (“doas”):

- (1) t > da/___s [pluralgilea]
 d > Ø/o___a

Beste informatzaile batzuek, ostera, “-t” morfemari eusten diote (“dotas”, “dotes”). Aldaketa fonetikotzat jo behar dugu loturazko bokala “a” edo “e” izatea. Gainerako aldaketa fonetikoak “-de” eta “-die” morfemen “d”ren ordez “r” erabiltzea, erroaren “eu” diptongoari eragiten diotenak, “eu”ren ordez “au” (“dau”). Antzera gertatzen da “ei” diptongoarekin beronen ordez “i” (“ditus”) eta “o” erabiltzen dira (“dotues”) eta batzuetan lehen silaba osoa gal daiteke (“tues”). Azkenik “gend” morfemaren “d” gal daiteke (“genusen” “gendusen”en ordez) beharbada euskara estandarren eraginez.

Nork-Nori-Nor sailean aldaketak hamar ezaugarritan gertatzen dira; 2 morfologia arlokoak dira, hiru fonologia arlokoak eta bost fonetikoak. Subjektuaren leku aldatzea hainbatean gertatzen da (“*dotzetesen*”, “*dotzegun*”, “*dotzesun*”). “*zidan*” adizkiaren iraganaren morfema “Ø” izan daiteke (“*osten*”), “*y*” (“*yosten*”) edo “*d*” (“*dosten*”). Fonologia arloko hiru arau agertzen dira; batean (2) bokal ertaina altu bihurtzen da bokal baxu baten aurrean (“*dotzias*”).

(2) $e > i / _a$

Bada beste arau bat “*tzo*” morfemaren “*o*” edozein kontsonanteren aurrean “*e*” bihurtzen duena; informatzaile batzuek, ostera, ez dute erabiltzen (“*geuntzon*” “*geuntzen*” en ordez). Azken arau fonologikoa “*t*” morfemari eragiten diona da, gorago ikusi dugun (1) arau berbera da eta hemen “*dizkiot*” adizkiaren “*t*” dagokio, pluralgilearen aurrean “*da*”ren ordez (“*dotze(d)as*”) informatzaile batzuek “*t*”ri berari eusten diote (“*dotzetes*”).

Aldaketa fonetikoetan “*a*” eta “*e*” lotura hizkien alternantzia aurkitzen dugu kasu batzuetan (“*dotzetas*”, “*dotzetes*”). Erroaren “*eu*”ren ordez “*o*” (“*gontzen*”) eta “*e*” (“*nentzesan*”) agertzen dira. “*neun*” (“*neuntzesan*”) oinarriaren azken “*n*” gal daiteke (“*neutzesan*”). Azkenik, silaba galera bi dauzkagu “*genion*” adizkiaren ordain batzuetan hasierako silaba galtzen da eta azken silabaren murriztapena ere gerta daiteke (“*tzun*”).

Aurreko lanean adierazi genuen bezala, aztertu ditugun aldaketa morfologikorik garrantzitsuenak analogiaz egiten direnak dira. Joera handia dago iragan aldiko adizki tradizionalen ordez 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 sailean joera hau nahikoa azkarra da.

4. Azterketa kuantitatiboa

Informatzaileak egindako adizkien aldakien araberako sailkapena egiteko, multzokatze azterketa hierarkiko izeneko aldagai anizkun teknika erabiliko dugu. Teknika hau sailkatzailea da eta beronen helburu nagusia da elementu batzuk multzokatzea talde homogeneotan, euren artean dauzkaten antzekotasunetan edo desberdintasunetan oinarrituta. Oso garrantzitsua da ulertzea teknika deskribatzailea dela eta berak ematen dituen multzoak dakigunarekin eta teknika kualitatiboekin eta geografikoekin erkatu behar direla:

El Análisis Cluster no tiene bases estadísticas sobre las que deducir inferencias estadísticas para una población a partir de una muestra, es un método basado en criterios geométricos y se utiliza fundamentalmente como una técnica exploratoria, descriptiva pero no explicativa. (De la Fuente, 2011)

Teknika hau aplikatzean urrats batzuk egin behar dira, Ayugari (d.g.) jarraikirik honako hauek proposa daitezke:

- (1) Aldagaien aukeraketa
- (2) Multzokatze metodoa
- (3) Antzekotasun eta distantzia neurriak

- (4) Sailkatze algoritmoa
- (5) Multzoak lortu
- (6) Emaitzen interpretazioa

Aldagai asko edukiz gero, multzokatzea egin aurretik azterketa estatistiko batzuk egin daitezke aldagai garrantzitsuenak aukeratzeko. Gure artean egin diren ikerketetan azterketa faktoriala proposatu da (Aurrekoetxea, 1995) aldagai kopurua murrizteko; aldagaiak dikotomikoak direnean korrespondentzia bakunen edo anizkunen azterketak gomendatzen dira (Pérez López, 2005). Aldagai dikotomikoak erabiltzen direnean, aldagaia erabiltzen den kasuan “1” agertuko da datu matrizean eta erabiltzen ez denean “0” (2. irudian adibide bat ematen da Nor-Nori saileko adizkiekin).

Lek	yako	dako	daste	date	yat	yate	dakus	daskus	yakus	akosan	akosen	dakosen	yakosan	yakosen
A11	1	0	0	0	1	0	0	0	1	0	0	0	1	0
A12	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A13	0	1	0	1	0	0	1	0	0	0	1	0	0	0
A14	1	0	0	0	1	0	0	0	1	0	0	0	0	1
G11	1	0	0	0	1	0	0	0	1	0	0	0	1	0
G12	0	1	0	1	0	0	1	0	0	0	0	0	1	0
G13	1	0	0	0	0	1	0	0	1	0	0	0	0	1
G14	0	1	0	1	0	0	1	0	0	0	0	0	1	0
A21	0	1	0	1	0	0	1	0	0	0	0	0	1	0
A22	0	1	0	1	0	0	1	0	0	1	0	0	0	0
A23	1	0	0	0	0	1	0	0	1	0	0	0	1	0
A24	0	1	1	0	0	0	0	1	0	0	0	0	0	1
G21	0	1	0	1	0	0	1	0	0	0	0	0	1	0
G22	0	1	0	1	0	0	0	0	1	0	0	0	1	0
G23	1	0	0	0	1	0	0	0	1	0	0	0	1	0
G24	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A31	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A32	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A33	0	1	0	1	0	0	0	1	0	0	0	0	1	0
A34	0	1	0	1	0	0	1	0	0	0	0	0	1	0
G31	0	1	0	1	0	0	0	0	1	0	0	0	0	1
G32	0	1	0	1	0	0	1	0	0	0	0	0	1	0
G33	0	1	1	0	0	0	0	1	0	0	0	0	1	0
G34	0	1	1	0	0	0	0	1	0	0	0	0	1	0

2. irudia: Nor-Nori sailean informatzaileek emandako adizkiekin sortutako matrize dikotomikoa

Multzokatze azterketetan mota bi bereizten dira gehienetan (Pardo eta Ruiz, 2002, Pérez López, 2005), multzokatze hierarkikoak eta ez hierarkikoak. Bariazioaren azterketetan metodo hierarkikoak erabiltzen dira (Goebel, 2010). Era berean, metagarriak edo urrigarriak izan daitezke (Aurrekoetxea, 1995; Terrádez, d.g.).

“pueden ser acumulativos (se forman grupos haciendo *clusters* cada vez más grandes) o disminutivos (partiendo de un solo grupo se separan los elementos en *clusters* cada vez más pequeños)”

Bariazioaren gaineko ikerketetan metagarriak erabiltzen dira, hau da, multzokatze urrats bakoitzean hurbilen daudenak multzo berean sartzen dira beste multzo bat osatzeko eta horrela harik eta multzo bakarra lortu arte.

Multzokatze-metodoa aukeratu ondoren, kasuen artean erkaketak egin ahal izateko, antzekotasunen eta desberdintasunen distantzia neurria aukeratu behar da. Hauetariko asko agertzen dira bibliografian (Pardo eta Ruiz, 2002, Pérez López, 2005), baita bariazioa aztertzeke erabili direnetan ere (Goebel, 1992, 2010; Aurrekoetxea, 1995, 2005; Clua, 2010).

Lexikoa konparatzean proposatu den distantzia neurrietako bat Leveinstein distantzia izan da (Nerbonne et al., 1999, 2010; Heeringa, 2004; Heeringa et al., 2006, Iglesias, 2014, Ensunza, 2015). Distantzia hau interesgarria izan daiteke neurtzeko berben aldakien arteko alde fonetikoak, baina lexema desberdina denean arazoak sorrarazten ditu; izan ere, batetik bestera dagoen distantzia % 100ekoa da beti horrelakoetan. Edozelan ere, teknika esploratzailea denez gero, distantziak probatu beharko liriateke erabili orduko: izan ere, bakoitzeko emaitzak desberdinak izaten dira.

SPSSk ematen dituen aukeretako batzuk erabiliz gero, ikus ditzakegu kalkuluak egindakoan dauden aldeak. Esate baterako, 3. irudian lehen belaunaldiko informatzaileen distantzia euklidearra erabilia agertzen diren emaitzak erakusten dira, distantzia honen eta Minkowskyren distantziaren emaitzak bat datoz.

Caso	1:A11	2:A12	3:A13	4:A14	5:G11	6:G12	7:G13	8:G14
1:A11	,000	1,395	1,834	1,228	1,611	2,122	1,197	1,997
2:A12	1,395	,000	1,844	1,088	1,394	2,071	1,321	1,749
3:A13	1,834	1,844	,000	1,899	1,545	1,325	1,604	1,162
4:A14	1,228	1,088	1,899	,000	1,314	2,204	1,286	1,891
5:G11	1,611	1,394	1,545	1,314	,000	1,944	1,297	1,454
6:G12	2,122	2,071	1,325	2,204	1,944	,000	1,884	1,384
7:G13	1,197	1,321	1,604	1,286	1,297	1,884	,000	1,733
8:G14	1,997	1,749	1,162	1,891	1,454	1,384	1,733	,000

3. irudia: 1. belaunaldiko informatzaileen arteko distantzia-matrizea distantzia euklidearra erabilia.

Caso	1:A11	2:A12	3:A13	4:A14	5:G11	6:G12	7:G13	8:G14
1:A11	,000	1,946	3,363	1,507	2,595	4,502	1,432	3,989
2:A12	1,946	,000	3,399	1,185	1,943	4,288	1,745	3,059
3:A13	3,363	3,399	,000	3,606	2,387	1,756	2,573	1,351
4:A14	1,507	1,185	3,606	,000	1,726	4,859	1,654	3,575
5:G11	2,595	1,943	2,387	1,726	,000	3,778	1,683	2,115
6:G12	4,502	4,288	1,756	4,859	3,778	,000	3,548	1,915
7:G13	1,432	1,745	2,573	1,654	1,683	3,548	,000	3,004
8:G14	3,989	3,059	1,351	3,575	2,115	1,915	3,004	,000

4. irudia: 1. belaunaldiko informatzaileen arteko distantzia-matrizea distantzia euklidear karratua erabilia.

Badaude beste distantzia mota batzuk, hala nola Hiri Blokeena (5. irudian emaitzak ikus daitezke) edo Chebychevena (6. irudian emaitzak erakusten dira). Kasu hauetan guztietan ikusten den moduan, emaitzak bat ez etortzean, distantzia mota aukeratzean kontuan hartu behar da zertan oinarritzen den, hau da, berdintasunetan edo desberdintasunetan eta elementuen arteko hurbiltasunari edo urruntasunari garrantzia ematen zaion. Distantzia euklidearra aukeratzen denean, taldeen barruko elementuak antzekoak izango dira eta euren arteko distantzia txikiak.

Caso	1:A11	2:A12	3:A13	4:A14	5:G11	6:G12	7:G13	8:G14
1:A11	,000	5,957	8,166	4,416	6,873	10,209	4,208	9,832
2:A12	5,957	,000	8,955	4,457	6,164	9,832	4,831	8,123
3:A13	8,166	8,955	,000	9,248	6,291	4,209	7,206	4,250
4:A14	4,416	4,457	9,248	,000	4,957	11,541	4,792	9,498
5:G11	6,873	6,164	6,291	4,957	,000	8,750	4,999	5,623
6:G12	10,209	9,832	4,209	11,541	8,750	,000	8,499	4,209
7:G13	4,208	4,831	7,206	4,792	4,999	8,499	,000	8,040
8:G14	9,832	8,123	4,250	9,498	5,623	4,209	8,040	,000

5. irudia: 1. belaunaldiko informatzaileen arteko distantzia-matrizea Hiri Blokeen distantzia erabilia.

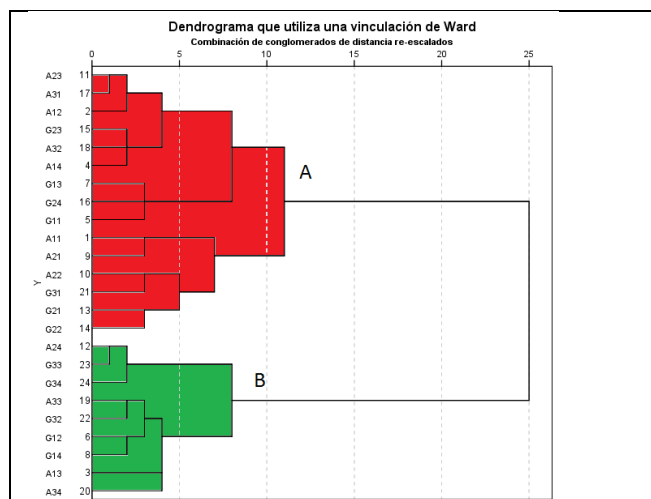
Caso	1:A11	2:A12	3:A13	4:A14	5:G11	6:G12	7:G13	8:G14
1:A11	,000	,667	,667	,667	,667	,917	,667	,667
2:A12	,667	,000	,583	,458	,583	,917	,583	,583
3:A13	,667	,583	,000	,583	,583	,917	,583	,542
4:A14	,667	,458	,583	,000	,583	,917	,583	,583
5:G11	,667	,583	,583	,583	,000	,917	,542	,583
6:G12	,917	,917	,917	,917	,917	,000	,917	,917
7:G13	,667	,583	,583	,583	,542	,917	,000	,583
8:G14	,667	,583	,542	,583	,583	,917	,583	,000

6. irudia: 1. belaunaldiko informatzaileen arteko distantzia-

matrizea Chebychev distantzia erabilia.

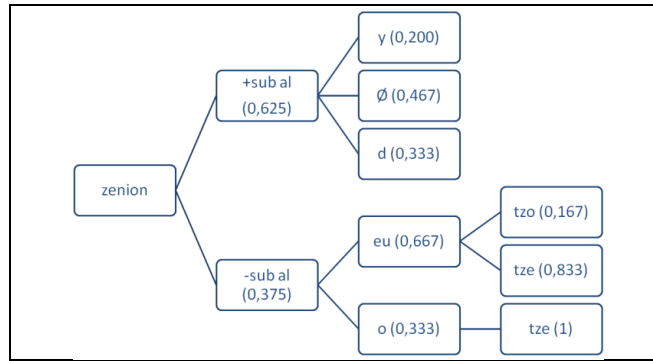
Azkenik, sailkatze algoritmo asko daude (Pardo eta Ruiz, 2002, Pérez López, 2005), baina bariazioaren azterketetan gailendu dena Ward-en metodoa da (Goebel, 1992, 2010; Aurrekoetxea, 1995, 2005; Gaminde, 2007; Gaminde et al., 2016b), euskararen gainean egineko lanetan ere algoritmo hau izan da erabiliena.

Edozelan ere, matrizea eraikitzean datu dikotomikoak erabiltzen badira, aldaki guztiak maila berean jartzen dira. Gure datuak horrelako matrize bat erabilia multzokatzen ditugunean, 7. irudian agertzen den dendograma lortzen da. Bertan agertzen diren multzo nagusien osaketa aztertzen badugu, ikusten dugu A multzoan hamabost informatzaile kokatzen direla eta B multzoan bederatzia. Informatzaileen belaunaldien arabera, A multzoan 1. belaunaldiko bost informatzaile kokatzen dira 2. belaunaldiko zazpi eta 3. belaunaldiko hiru. B multzoan 1. belaunaldiko hiru informatzaile kokatzen dira, 2. belaunaldiko bat eta 3. belaunaldiko bost. Generoaren arabera, A multzoan zortzi neska agertzen dira eta zazpi mutil eta B multzoan lau neska eta bost mutil. Beraz esan dezakegu ez belaunaldika ez eta generoka ez dagoela multzo homogeneorik.



7. irudia: Informatzaileen sailkapena matrize dikotomiko bat erabilia.

Aldakien probabilitatearen azterketa egiteko urrats batzuk egitea proposatzen da. Batetik, aurreko behar batean (Gaminde et al., 2016a) proposatzen genuen bezala, adizkien morfemen hurrenkeraren arabera sortzen diren txantiloiak (Artiagoitia, 2000; Gaminde et al., 2012) oinarritzat hartu ditugu eta gero adizkiak eraikitzeke erabili behar diren eragiketak hierarkikoki antolatu ditugu. Eragiketa horiek ordenatuta egin 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 zedarritzen dira. Zuhaitz-diagrametan urkulu bakoitzaren adarren probabilitateak adierazten dira. (8. irudian “zenion” adizkiaren urkulketa eta adarren probabilitateak erakusten dira).



8. irudia: Aldagai bakoitzaren aldakiak eta probabilitateak.

Informazio honekin, lan honetatik kanpo geratzen den arren (ikusi Etxebarri et al., 2016 eta Gaminde et al. 2016a), aldaki bakoitzaren probabilitateak kalkula daitezke. Esate baterako, “yotzesun” adizkiaren probabilitatea ondoko formularen arabera kalkulatu litzateke:

$$P (“+sub al” \cap “y”)$$

“+sub al” subjektuaren leku aldaketari dagokion ezaugarria da, “y” iragan aldiko hasierako morfemari dagokio. Kalkulua ondoko era honetara egiten da.

$$P (“+sub al”) * P (“y”)$$

Termino bakoitzaren probabilitateak 8. irudian erakusten direnak dira; beraz, adizki honen agerpen probabilitatea 0,125koa da. Probabilitateak honela kalkula ditzakegu ezaugarri bakoitzarena zenbatekoa den jakin gura dugunean edo egin dezakegu adizki bakoitzarena kalkulatu zuzenean, bakoitzaren maiztasunean oinarrituta (kontuan hartu bedi azken emaitza, dezimalen biribiltzearen zehaztasun ezak eragiten dituen gorabeherengatik ere bertsua dela era bitara eginda). 9. irudian erakusten da zelan kalkulatu diren Nor-Nori saileko adizkien probabilitateak kalkulu orri arrunt batean.

Lek	yako	dako	daste	date	yat	yate	dakus	daskus	yakus	akosan	akosen	dakosen	yakosan	yakosen
A11	1	0	0	0	1	0	0	0	1	0	0	0	1	0
A12	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A13	0	1	0	1	0	0	1	0	0	0	1	0	0	0
A14	1	0	0	0	1	0	0	0	1	0	0	0	0	1
G11	1	0	0	0	1	0	0	0	1	0	0	0	0	1
G12	0	1	0	1	0	0	1	0	0	0	0	1	0	0
G13	1	0	0	0	0	1	0	0	1	0	0	0	0	1
G14	0	1	0	1	0	0	1	0	0	0	0	1	0	0
A21	0	1	0	1	0	0	1	0	0	0	0	0	1	0
A22	0	1	0	1	0	0	1	0	0	1	0	0	0	0
A23	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A24	0	1	1	0	0	0	0	1	0	0	0	0	0	1
G21	0	1	0	1	0	0	1	0	0	0	0	0	0	1
G22	0	1	0	1	0	0	0	0	1	0	0	0	0	1
G23	1	0	0	0	1	0	0	0	1	0	0	0	0	1
G24	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A31	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A32	1	0	0	0	0	1	0	0	1	0	0	0	0	1
A33	0	1	0	1	0	0	0	1	0	0	0	1	0	0
A34	0	1	0	1	0	0	1	0	0	0	0	1	0	0
G31	0	1	0	1	0	0	0	0	1	0	0	0	0	1
G32	0	1	0	1	0	0	1	0	0	0	0	1	0	0
G33	0	1	1	0	0	0	0	1	0	0	0	1	0	0
G34	0	1	1	0	0	0	0	1	0	0	0	1	0	0
	10	14	3	11	4	6	8	4	12	1	1	7	12	3
	0,416666667	0,583333333	0,125	0,458333333	0,166666667	0,25	0,333333333	0,166666667	0,5	0,041666667	0,041666667	0,291666667	0,5	0,125

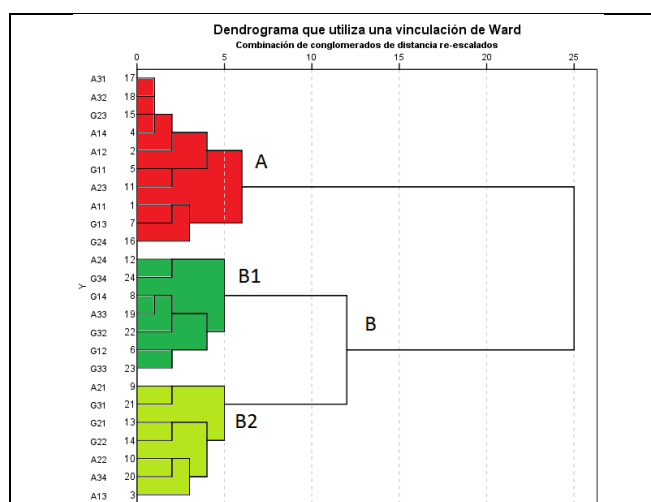
9. irudia: Nor-Nori saileko probabilitateen kalkulua.

Adizki bakoitzaren probabilitatea erabiltzen badugu “1” ordeztzeko, 10. irudian agertzen den bezala, multzokatze azterketa egiteko behar den matrizea lortzen dugu.

Lek	yako	dako	daste	date	yat	yate	dakus	daskus	yakus	akosan	akosen	dakosen	yakosan	yakosen
A11	0,417	0	0	0	0,167	0	0	0	0,5	0	0	0	0,5	0
A12	0,417	0	0	0	0	0	0,25	0	0	0,5	0	0	0	0,5
A13	0	0,583	0	0,458	0	0	0,333	0	0	0	0,042	0	0	0
A14	0,417	0	0	0	0,167	0	0	0	0,5	0	0	0	0,5	0
G11	0,417	0	0	0	0,167	0	0	0	0,5	0	0	0	0,5	0
G12	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0,292	0	0
G13	0,417	0	0	0	0	0,25	0	0	0,5	0	0	0	0,5	0
G14	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0,292	0	0
A21	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0	0,5	0
A22	0	0,583	0	0,458	0	0	0,333	0	0	0,042	0	0	0	0
A23	0,417	0	0	0	0	0,25	0	0	0,5	0	0	0	0,5	0
A24	0	0,583	0,125	0	0	0	0	0,167	0	0	0	0	0	0,125
G21	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0	0,5	0
G22	0	0,583	0	0,458	0	0	0	0	0,5	0	0	0	0,5	0
G23	0,417	0	0	0	0,167	0	0	0	0,5	0	0	0	0,5	0
G24	0,417	0	0	0	0	0,25	0	0	0,5	0	0	0	0	0,125
A31	0,417	0	0	0	0	0,25	0	0	0,5	0	0	0	0,5	0
A32	0,417	0	0	0	0	0,25	0	0	0,5	0	0	0	0	0,125
A33	0	0,583	0	0,458	0	0	0	0,167	0	0	0	0,292	0	0
A34	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0,292	0	0
G31	0	0,583	0	0,458	0	0	0	0	0,5	0	0	0	0,5	0
G32	0	0,583	0	0,458	0	0	0,333	0	0	0	0	0,292	0	0
G33	0	0,583	0,125	0	0	0	0	0,167	0	0	0	0,292	0	0
G34	0	0,583	0,125	0	0	0	0	0,167	0	0	0	0,292	0	0

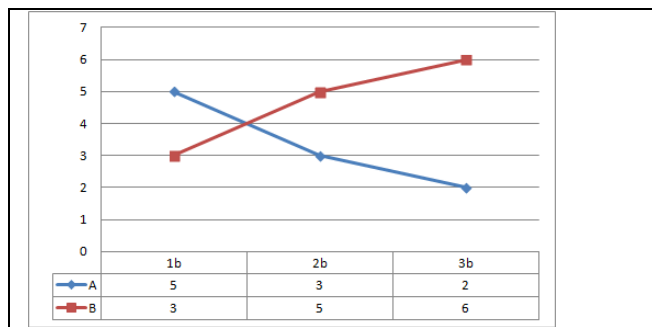
10. irudia: Nor-Nori saileko probabilitateen araberako matrizea.

Orain multzokatze azterketa egiten dugunean, gorago adierazitako parametro berberak erabilia 11. irudiko dendograma lortzen dugu. Ikusten den moduan, A multzoan hamalau informatzaile kokatzen dira eta B multzoan hamar. Informatzaileen belaunaldien araberako banaketari begira A multzoan 1. belaunaldiko bost informatzaile kokatzen dira 2. belaunaldiko hiru eta 3. belaunaldiko bi. B multzoan 1. belaunaldiko hiru informatzaile kokatzen dira, 2. belaunaldiko bost eta 3. belaunaldiko sei. Generoaren arabera, A multzoan sei neska agertzen dira eta lau mutil eta B multzoan sei neska eta zortzi mutil.

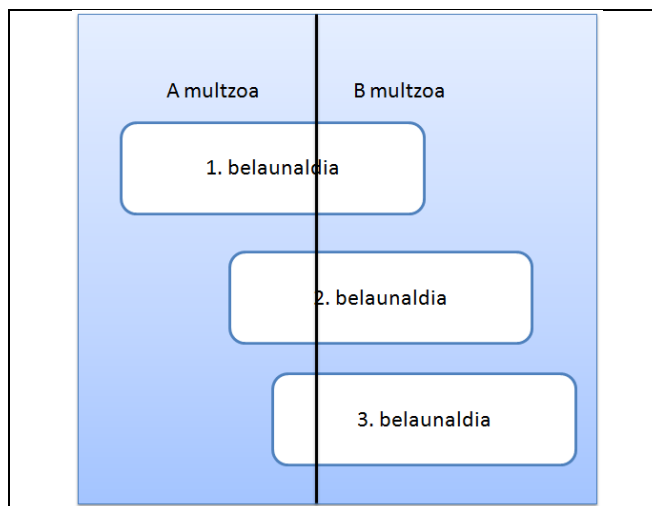


11. irudia: Informatzaileen sailkapena probabilitateen matrizea erabilia.

Emaitzak ikusita argi agertzen da belaunaldiaren arabera progresio argia agertzen dela informatzaileen sailkapenean, 12. irudian ikus daitekeen bezala. Datu hauek ikusita esan dezakegu bigarren eta hirugarren belaunaldietako informatzaileak elkarrengandik hurbilago daudela lehen belaunaldiko informatzaileengandik baino (13. irudia).



12. irudia: Informatzaileen hedapena multzoetan euron belaunaldien arabera.



13. irudia: Belaunaldi bakoitzeko informatzaileen hedadura multzo bakoitzean.

5. Ondorioak

Hona helduta aurreko atalean eztabaidatu ditugun emaitzetatik ondorio nagusiak laburbilduko ditugu.

Batetik, multzokatze-azterketa egiteko erabiltzen den datu-matrizearen ezaugarriek azken emaitza baldintzatzen dutela argi geratu da. Argi geratu da “0” eta “1” erabilia sortzen den matrize dikotomikoaren bidez askozaz gutxiago azaltzen dena molde bakoitzaren probabilitateak kalkulatu eraikitzen den matrizearen bidez baino.

Beste alde batetik, probabilitateen kalkulua alde aurreko azterketa linguistikoan oinarritzen denean, aldaketak eurak ere pisatu egiten dira eta bakoitzari bere balioa ematen zaio. Era honetara, ezaugarri linguistikoaren balioetan oinarrituta, azterketa doituagoa da.

Azkenik, era honetara eginez gero, azpian dautzan egiturak azaleratzen dira; izan ere, matrize dikotomikoaren bidez egindako azterketak ez du ezer erakusten; alta, probabilitateen araberrako azterketa egiten denean, informatzaileen banaketan belaunaldiak daukan garrantzia agerian jartzen da. Hizkerak maila diatopikoan sailkatzen direnean metodo berberak erabil daitezke.

Edonola ere, hemen ere ikuspuntu ekliptikoa aldarrikatu behar dugu. Estatistikak teknikak ipintzen ditu gure esku, baina guri dagokigu unean-unean gure datuak azaltzeko teknirik onenak aukeratzea eta aplikatzea.

6. Bibliografia

- Arteagagoitia, X. (2000). *Hatsarreak eta parametroak lantzen*. AFA.
- Aurrekoetxea, G. (1992). Nafarroako Euskara: Azterketa Dialektometrikoa. *Uztaro* 5, 59-109.
- Aurrekoetxea, G. (1995). *Bizkaieraren egituraketa geolinguistikoa*. Bilbo: EHU.
- Aurrekoetxea, G. (2005). Nafarroako Euskararen Sailkapenaz. In Etxebarria, P. eta Knör, H. (2005) *Nerekin yaio nun. Txillardegiri omenaldia* (109-124 orr.). Bilbo: Euskaltzaindia eta EHU.
- Aurrekoetxea, G. (2009). Bariazio sinkronikoa aztertzeko metodologia(k). *Lapurdum* 13, 43-59.
- Aurrekoetxea, G. / J. L. Ormaetxea (Edtk.). (2010). *Tools for Linguistic Variation*, Bilbao: UPV-EHU, Supplements of the Anuario de Filología Vasca "Julio de Urquijo", 53.
- Aurrekoetxea, G., & Ormaetxea, J.L. (2006). Euskararen Atlas Sozio-Geolinguistikoa ikerketa proiektua, *Euskalingua* 9.
- Aurrekoetxea, G. (2003). Euskalkiak estandarren uholdepean, Arratiako kasua. In *Mendebalde Kultura Alkartea: Ahozotasuna aztergai*.
- Aurrekoetxea, G. (2004). Estandar eta dialektoen arteko bateratze-joerak, ikuspuntu teorikotik begirada bat. *Uztaro*, 50.
- Aurrekoetxea, G. (2008). Bariazio soziolinguistikoa Dimako euskararen. *Euskalingua*, 12.
- Aurrekoetxea, G., Gandarias, L., Gaminde, I., & Iglesias, A. (2014a). Variación prosódica en vasco: áreas acentuales. In Yolanda Congosto Martín, María Luisa Montero Curiel y Antonio Salvador Plans (eds.) *Fonética Experimental, Educación Superior e Investigación*. Madrid: Arco Libros.
- Aurrekoetxea, G, Gaminde, I., Gandarias, L., & Iglesias, A. (2014b). Prosodic variation in the Basque language: intonational areas. In Díaz-Negrillo, A. & Díaz Pérez Francisco Javier (edtk) *Specialisation and Variation in Language Corpora*. Berna: Peter Lang AG, International Academic Publishers.
- Ayuga, E. (d.g.). Análisis de Conglomerados.
http://ocw.upm.es/estadistica-e-investigacion-operativa/matematicas-y-estadistica-aplicada/contenidos/OCW/Anal_Multivar/Mat_Clase/anal_mult_2.pdf
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2011). La lengua es un sistema adaptativo complejo. *Lingüística en la Red*.
http://www.linred.es/articulos_pdf/LR_articulo_04092011.pdf
- Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer, Version 5.1*. [Computer program]. <<http://www.praat.org>>
- Clua, E. (2010). Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal. In Aurrekoetxea, G. & Ormaetxea, J. L. (2010). *Tools for Linguistic Variation*. Bilbo: UPV/EHU.
- De la Fuente, S. (2011). *Análisis Conglomerados*. Universidad Autónoma de Madrid.
<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>
- Díaz de Rada, V. (2009). *Análisis de datos de encuesta: Desarrollo de una investigación completa utilizando SPSS*. Bartzelon: UOC.
- Ellis, N. C. (2011). The Emergence of Language as a Complex Adaptive System. In James Simpson (Ed.), *Routledge Handbook of Applied Linguistics*. Routledge/Taylor Francis.

- Ensunza, A. (2015). *Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan*. [Doktoretza tesia]. Leioa: UPV/EHU.
- Etxebarria, A., Gaminde, I., Olalde, A., & Gaminde, U. (2016). Hizkuntza aldakortasuna Larrabetzuko aditz morfologian. In Iglesias, A.; Romero, A. eta Ensunza, A. *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. Bilbo: UPV/EHU.
- Etxeberria, J. (2011). *Estatistika Aplikatua: Teoria eta Praktika*. Donostia: Elhuyar.
- Gaminde, I. (2007). *Bizkaian Zehar: Euskararen Ikuspegi Orokorra*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I., Romero, A., & Legarra, H. (2012) *Gramatika eta hizkuntza bariazioa Bermeo*. Bermeo: Bermeoko Udala.
- Gaminde, I., Romero, A., Etxebarria, A., & Eguskiza, N. (2016a). Bizkaiko aditz laguntzaileen bilakaeraren azterketaz. In Iglesias, A.; Romero, A. eta Ensunza, A. *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. Bilbo: UPV/EHU.
- Gaminde, I., Etxebarria, A., Eguskiza, N., Romero, A. & Unamuno, L. (2016b). Lexikoarenbariazioa eta multzokatze-azterketa. *Euskalingua*, 28, 19-3.
- Gell-Mann, M. (1994). *El quark y el jaguar: Aventuras en lo simple y lo complejo*. San Francisco: EpubLibre.
- Goebel, H. (1992). Problèmes et méthodes de la dialectométrie actuelle (avec application à l' AIS). In Euskaltzaindia/Académie de la Langue Basque éd., *Nazioarteko dialektologia biltzarra. Agiriak/Actes du Congrès International de Dialectologie* (429-475 orr.), Bilbo: Euskaltzaindia.
- Goebel, H. (2010). Introducción a los problemas y métodos según los principios de la Escuela Dialectométrica de Salzburgo. In Aurrekoetxea, G. & Ormaetxea, J. L. (2010) *Tools for Linguistic Variation*. Bilbo: UPV/EHU.
- Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Doktorego tesia. University of Groningen, Groningen. <http://www.let.rug.nl>
- Heeringa, W., Kleiweg, P., Gooskens, Ch., & Nerbonne, J. (2006). Evaluation of String Distance Algorithms for Dialectology. http://www.let.rug.nl/gooskens/pdf/publ_lingdist_2006.pdf
- Herrera, H., Martínez, R., & Amengual, M. (2001). *Estadística aplicada a la investigación lingüística*. Madril: EOS Universitaria.
- Iglesias, A. (2014). *Igorreko hizkeraren azterketa dialektologikoa*. [Doktorego tesia]. Leioa: UPV/EHU
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: U. of Pennsylvania Press.
- Laka, E., Olondo L., & Gaminde, I. (2008). Bermeoko gazteen euskararen aditz morfologiaz, *Euskalingua*, 13.
- Léonard, J. L., Heinsalu, E., Patriarca, M., & 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* (50-58 orr.). Leioa: 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.
- Lujanbio, O. (2016). *Hizkuntza-aldakortasuna euskararen Nafarroa ipar-mendebaldeko bi udalerritan egindako azterketa*. [Doktoretza-tesia]. Leioa: UPV/EHU.
- Montero, J. (2017) *Permiso para quejarse*. Bartzelona: Ariel.
- Moreno Cabrera, J. C. (2008) *El nacionalismo lingüístico: Una ideología destructiva*. Bartzelona: Ediciones Península.

- Murillo, F. J., & Martínez-Garrido, C. (2012). *Análisis de datos cuantitativos con SPSS en investigación socioeducativa*. Madril: UAM.
- Nerbonne, J., Prokic, J., Wieling, M., & Gooskens, Ch. (2010). Some further dialectometrical steps. In Aurrekoetxea, G. & Ormaetxea, J. L. (2010) *Tools for Linguistic Variation*. Bilbo: UPV/EHU.
- Nerbonne, J., Heeringa, W., Kleiweg, P. (1999). Comparison and Classification of Dialects *Proceedings of EACL '99*, <http://www.aclweb.org/anthology/E99-1048>.
- Pardo, A., & Ruiz, M. A. (2002). *SPSS 11 Guía para el análisis de datos*. Madril: McGraw-Hill.
- Pérez López, C. (2005). *Métodos estadísticos avanzados con SPSS*. Madril: Thomson.
- Romero, E. (2016). *Estadística para todos*. Madril: Pirámide.
- Terrádez, M. (d.g.) *Análisis de Conglomerados*. Bartzelona: Universitat Oberta de Catalunya. <http://www.uoc.edu/in3/emath/docs/Cluster.pdf>

AHOTS SEDUZITZAILEEN EZAUGARRIAK GENEROAREN IKUSPEGITIK³

Aintzane Etxebarria, Naia Eguskiza, Iñaki Gaminde eta Asier Romero
Euskal Herriko Unibertsitatea
aintzane.etxebarria@ehu.eus, naia.eguskiza@ehu.eus, inaki.gaminde@ehu.eus,
a.romero@ehu.eus

Laburpena

Ahotsaren bitartez era ezberdinetako funtzioak bete ditzakegu; batzuk linguistikoak dira, azentua, intonazioa, fokoa eta esaldi egitura zehazteko balio dute; beste batzuk paralinguistikoak dira eta jarreraren, intentzioen eta hizkera estiloen berri ematen digute; azkenik, badaude beste zenbait ez direnak linguistikoak eta fisikoaren, egoera emozionalaren zein pertsonalaren berri ematen digutenak. Adierazitako funtzio horiek betetzeko nahiz pertzibitzeko ezinbestekoak dira tasun akustikoak, oinarritzko maiztasuna edo F0, iraupena, energia, abiada eta ahots mota edo kalitatea. Lan honetan, hiru mutilen grabazioak erabiliz eta generoa aldagaiari garrantzia emanez, emakumeen ikuspuntua kontuan hartuko baita, ahots seduzitzailearen ezaugarriak zehaztean aipatutako lehen lau tasunetatik zeinek duen eraginik handiena adieraziko da, horretarako pertzepzio testak burutu dituzte Bilboko Irakasleen Unibertsitate Eskolako ehun emakume elebidunek.

Gako-hitzak: Ahots seduzitzailea, azentua, intonazioa, emozioak, euskara

1. Sarrera

Ahotsa hiru osagai-motaren ondorioz sortzen da: soinu-ezaugarriak, mekanismo fisiologikoak eta produkzio akustikoa. Morenok eta lankideek (2010) hiru alde horiek kontuan hartuta, era honetan azaltzen dute zer den fenomeno hau:

El fenómeno vocal es el resultado de la combinación de cualidades sonoras como el tono, la intensidad y el timbre y es desarrollado a través de mecanismos fisiológicos mediante tareas relacionadas con la producción acústica. (Moreno eta lank., 2010: 9)

Ahotsaren bidez hainbat funtzio lortzen ditugu; esaterako, entzulea asaldatzea, lasaitzea edota iraintzea, geure geureak diren ezaugarriak ere ematen ditugu ezagutaraztera; esaterako, hitz egiten duenaren generoaren berri izan dezakegu, eta horretan badaude zenbait parametro berezile; adibidez, Mora eta Cobeta (2013) autoreek esaten dutenaren ildotik, ahots-tonua (emakumeetan gizonezkoetan baino altuagoa da), bolumena, ahots tinbrea, intonazio patroiak eta ezaugarri linguistikoak nahiz paralinguistikoak, baita hizkuntzarekin zerikusirik ez duen morfologia fisikoa ere.

Arazo fisikoen edota gaixotasunen berri ere izan dezakegu ahotsaren bidez, batez ere, aipatutako hiru osagai-motaren batean aldaketak gertatzen direlako; esaterako, Alzheimerra, Parkinsona edota esklerosia gaixotasunek disprosodia edo prosodia-eza izaten dute:

La enfermedad de Parkinson se caracteriza también por disprosodia y disartria. La prosodia es la melodía del lenguaje, es decir, la combinación de la entonación,

³ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

pronunciación, ritmo de habla y estrés silábico, que aportantanto información lingüística como no lingüística. (Sanabria, García, Ahumada, 2013: 285).

Gainera, aurretiko lanetan ikusi izan dugu ahotsak jarrerren, emozioen eta intentzioen berri ematen digula (Gaminde eta lank., 2014; Gaminde, 2007, 2010; Garay eta lank., 2011; Sainz eta lank., 2008; Navas eta lank., 2004, 2005, 2007), aipatutako informazio paralinguistiko nahiz ez linguistiko horiek aztertzeo tasun akustikoak erabili izan dira. Oraingo honetan, hiru ahotsek zer sugeritzen duten aztertzen da izaerari, itxurari eta estatusari dagokionez, jakina da, gizartean ikasitako edota jasotako aurreiritziek (Del Olmo, 2005) horretan eragina izan dezaketela, gainera, tasun akustikoetan oinarrituz (F0, energia eta iraupena) ahots horiek deskribatzen dira.

2. Metodologia

Lan hau burutzeko bi baliabide erabili izan dira: hiru mutil gazteren ahotsa esaldi bera adieraziz grabatuta eta galdetegia. Grabazioari dagokionez, Bilboko Irakasleen Unibertsitate Eskolako hiru ikasle gazte hautatu ziren, guztiak mutilak, horrela aukeratuak izan ziren ahots-mota ezberdina pertzibitu zitekeelako, guztiek hauxe esaten zuten:

“Errigoiti eta Gernika artean Citroen Xara motako auto bat bidetik irten eta irauli egin da”.

Grabazio horietan *Praat* softwarea (Boersma eta Weenink, 2017) erabiliz hiru izari akustikoren bost tasunak aztertu ziren.

Galdetegiari dagokionez, bi atal ditu; alde batetik, erantzuleen datu orokorreari buruzko informazioa eskatu zaie (jaiotze-urtea, generoa, herria, ama-hizkuntza eta ikas-eredua); bestetik, hiru ahotsei buruzko kalifikatzaile zerrenda eman zaie Likert eskalaren araberrako ezer ez, apur bat, nahiko ala asko markatzeko. Entzun ahala, kalifikatzaile bakoitzaren ondoan, pertzibitzen dutena ezer ez, apur bat, nahiko ala asko den markatu behar dute. Kalifikatzaile guztiak hiru multzotan agertzen dira, bat izaerari dagokio eta beste biak itxurari eta estatusari.

Otsailean zehar Bilboko Irakasleen Unibertsitate Eskolan hirugarren mailan ikasten ari ziren hirurogeita hamabost neskari pasatu zitzaaien galdetegia eta grabazio bakoitza entzun ondoren, galdetegiko kalifikatzaile bakoitzaren ondoan Likert eskalako maila bakoitzean X bat jarri behar zuten. Datuak Excel eta SPSS programen bidez aztertu ziren.

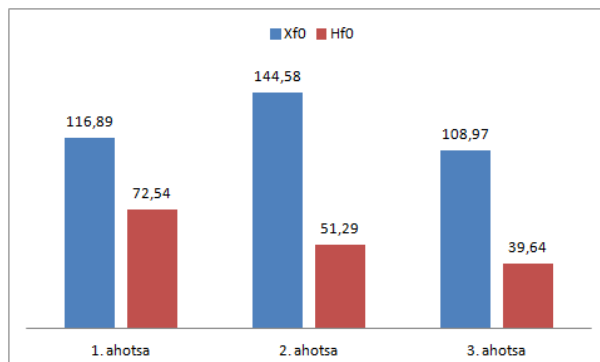
3. Ahotsen ezaugarri prosodikoak

Ahotsen ezaugarri prosodikoak aztertzeo makro prosodia eta mikro prosodia bereiziko ditugu. Makro prosodian testu osoaren neurketak aztertuko ditugu eta mikro prosodian, berriz, silaben neurketak. Erabili ditugun izari akustikoak oinarritzko maiztasuna edo f0 (Hz.), energia (db.) eta iraupena (ms.) dira.

F0ri dagokionez, batezbesteko osoa (Xf0) eta f0ren heinen batezbestekoa (Hf0) aztertu ditugu. Ahots bakoitzaren emaitzak 1. taulan agertzen dira (1. irudiko grafikoa).

Ahotsa	Xf0	Hf0
1	116,89	72,54
2	144,58	51,29
3	108,97	39,64

1. taula: F0ren batez besteko osoa eta heinen batez bestekoa.

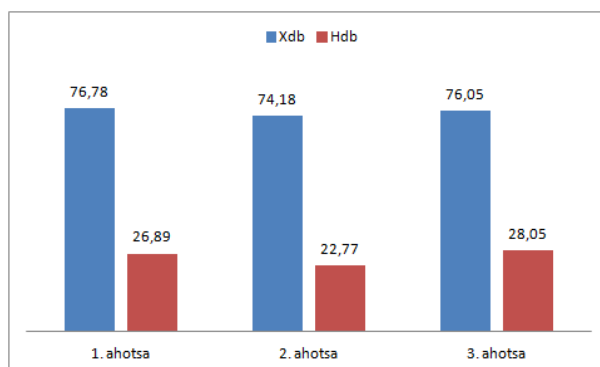


1. irudia: Hiru ahotsetan Xf0ren tea Hf0ren emaitzak.

Energiari dagokionez, batezbesteko osoa (Xdb) eta energiaren heinen batezbestekoa (Hdb) aztertu ditugu. Ahots bakoitzaren emaitzak 2. taulan agertzen dira (2. irudiko grafikoa).

Ahotsa	Xdb	Hdb
1	76,78	26,89
2	74,18	22,77
3	76,05	28,05

2. taula: Energiaren batezbesteko osoa eta heinen batezbestekoa.



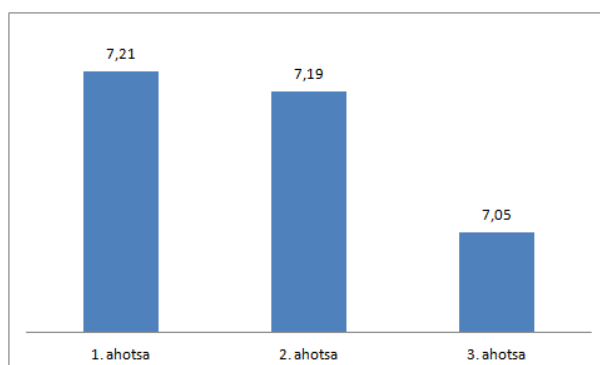
2. irudia: Hiru ahotsetan Xdbren eta Hdbren emaitzak.

Iraupenaren datuak 3. taulan ematen dira; hizketa denbora (H.D.), beronen ehunekoa (% H. D.), etenen denbora (E.D), beronen ehunekoa (% E.D.) eta testuaren denbora osoa ematen ditugu. Lehen informatzaileak denborarik luzeena erabili du eta etenen denborarik luzeena ere bai; bigarren informatzaileak denbora osorik laburrena erabili du eta etenen denborarik laburrena hirugarren informatzaileak erabili du.

	H.D.	% H.D	E.D.	% E.D.	Osoa
1. ahotsa	4440	91,17	430	8,83	4870
2. ahotsa	4370	96,47	160	3,53	4530
3. ahotsa	4700	98,33	80	1,67	4780

3. Taula: Iraupenaren datuak hiru ahotsetan.

Azkenik, hizketa abiada kalkulatu dugu informatzaile bakoitzeko. Kalkulu hau egiteko, etenen denbora kendu dugu eta talde prosodikoen denbora baino ez dugu erabili. Emaitzak antzekoak dira hiru ahotsekin: 1. ahotsak 7,21 silaba segundoko erabiltzen ditu, bigarrenak 7,19 eta hirugarrenak 7,05 (3. irudiko grafikoa)



3. irudia: Ahots bakoitzaren hizketa abiada.

Mikro prosodian datu zehatzagoak aurki ditzakegulakoan, silaben neurketak egin ditugu; neurtu ditugun izari akustikoak iraupena, f_0 eta energia izan dira.

4. taulan informatzaile bakoitzaren silaben iraupenaren batezbestekoak, desbideratzeak, minimoak, maximoak eta heinak ematen dira. Ikus daitekeen bezala, bigarren informatzailearen silabak laburrenak dira eta hirugarrenarenak, berriz, luzeenak; hala ere, aldeak ez dira estatistikoki esanguratsuak.

	N	x	sd	min	max	heina
1	32	139,06	41,14	50	240	190
2	32	132,50	33,89	60	210	150
3	33	141,52	38,42	40	240	200

4. taula: Silaben iraupenaren batezbestekoak, desbideratzeak, minimoak, maximoak eta heinak.

Xf_0 ren datuak 5. taulan agertzen dira, hirugarren informatzaileak grabeena dauka eta bigarren informatzaileak akutuena. ANOVA probaren arabera alde horiek estatistikoki esanguratsuak dira ($F = (a.m.: 2) 158,883$; $p = 0,000$), Tukeyren post hoc proba eginez gero, informatzaile guztien arteko aldeak estatistikoki esanguratsuak dira eta denak bereizten dira.

	N	x	sd	min	max	heina
1	32	117,98	9,08	94,44	135,23	40,79
2	32	143,42	7,20	125,33	163,09	37,76
3	33	109,11	7,71	88,61	124,11	35,50

5. taula: F_0 ren datuak.

Hf_0 ren datuak 6. taulan ikus daitezke; batezbestekorik handiena lehen informatzaileak dauka eta bigarrenaren eta hirugarrenaren arteko aldea txikia da. Aldea ANOVAREN arabera estatistikoki esanguratsua da ($F = (a.m.: 2) 3,818$; $p = 0,025$). Hala ere, Tukeyren post hoc probaren arabera alde hau esanguratsua da lehen eta bigarren informatzailearen artean bereizteko soilik.

	N	x	sd	min	max	heina
1	32	22,04	18,30	0,79	67,01	66,22
2	32	13,70	10,49	3,68	52,51	48,83
3	33	15,04	7,97	1,90	34,16	32,26

6. taula: Hf0ren datuak.

Xf0 eta Hf0 tasunen korrelazioak kalkulatu baditugu ($r = -0,410$, $p = 0,000$) ikusten dugu, berori esanguratsua den arren, nahikoa txikia dela.

Xdbren datuak 7. taulan ikus daitezke; batezbestekorik handiena lehen informatzaileak dauka eta txikiena bigarrenak, hirugarren eta lehen informatzaileen arteko aldeak oso txikiak dira. Aldea ANOVAREN arabera estatistikoki esanguratsua da ($F = (a.m.: 2) 6,043$; $p = 0,003$). Hala ere, Tukeyren post hoc probaren arabera alde hau esanguratsua da bigarren informatzailea beste bietatik bereizteko.

	N	x	sd	min	max	heina
1	32	76,92	3,81	70,39	85,91	15,52
2	32	73,84	3,60	66,99	81,58	14,59
3	33	76,32	3,88	67,72	83,29	15,57

7. taula: Xdbren datuak.

Azkenik, 8. taulako Hdb datuak aztertzen baditugu, ikusten dugu heinaren batezbestekorik altuena lehen informatzaileak daukala eta txikiena bigarrenak. Aldea ANOVAREN arabera ez da estatistikoki esanguratsua.

	N	x	sd	min	max	heina
1	32	13,78	6,75	3,15	27,71	24,56
2	32	10,68	6,14	1,63	20,53	18,90
3	33	12,64	7,53	1,09	24,86	23,77

8. taula: Hdbren datuak.

Xdb eta Hdb tasunen korrelazioak kalkulatu baditugu ($r = 0,983$, $p = 0,000$) ikusten dugu, korrelazio positibo eta handia dela.

Hiru izari akustikoetan oinarritutako bost tasunen eragina eta aldeak aztertuta argi agertzen da informatzaileak gehien bereizten dituen Xf0 dela.

4. Erantzunen azterketa

Erantzunen azterketa hiru azpiataletan banatuta aurkeztuko dugu; lehen azpiatalean informatzaileen balizko izaerari dagozkionak emango ditugu, bigarrenan itxura fisikoaren gainekoak eta hirugarrenean estatus sozialari dagozkionak.

4.1. Izaera

Ahotsak entzuleengan eragin ditzakeen ahotsaren jabearen izaeraren gaineko iritziak eta usteak neurtzeko bost galdera erabili ditugu: (a) Atsegina, (b) Ausarta, (c) Lasai, (d) Harroa eta (e) Zentzuduna. Jarraian galdera bakoitzeko entzuleen erantzunen emaitzak aurkeztuko ditugu.

(a) galderari dagozkion erantzunak 9. taulan ikusten dira; ahots guztien ehunekorik altuenak 2 kategorian kokatzen dira eta aldeak ez dira estatistikoki

esanguratsuak. Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (189) da, gero bigarrenarena (177) eta baxuena lehenarena (164).

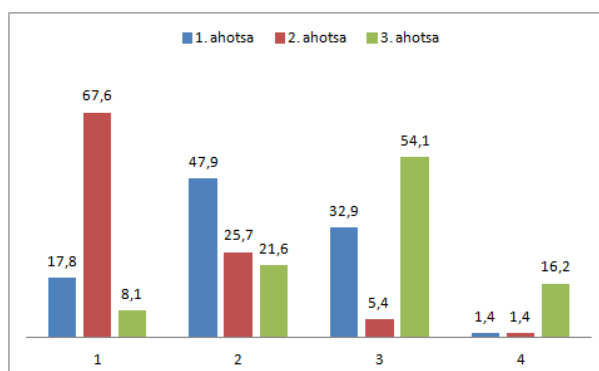
Ahotsa	1		2		3		4	
	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	15	20,27	32	43,24	23	31,08	4	5,41
2	14	18,92	29	39,19	19	25,68	12	16,22
3	8	10,81	29	39,19	25	33,78	12	16,22

9. taula: Atsegina izaerari dagozkion datuak hiru ahotsak kontuan izanda.

(b) galderari dagozkion erantzunak 10. taulan aurkezten dira; lehen ahotsaren ehunekorik altuena 2 kategorian kokatzen da, bigarren ahotsarena 1 kategorian eta hirugarren ahotsarena 3 kategorian (4. Irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 =$ (a.m.: 6) 105,543, $p = 0,000$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (206) da, gero lehenarena (159) eta baxuena bigarrenarena (104).

Ahotsa	1		2		3		4	
	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	13	17,81	35	47,95	24	32,88	1	1,37
2	50	67,57	19	25,68	4	5,41	1	1,35
3	6	8,11	16	21,62	40	54,05	12	16,22

10. taula: Ausarta izaerari dagozkion datuak hiru ahotsak kontuan izanda.

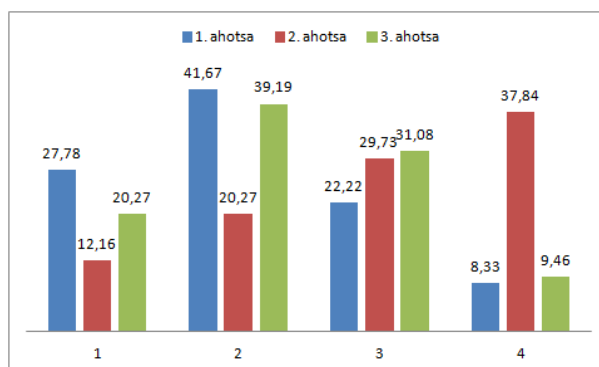


4. irudia: (b) galderari dagozkion erantzunak.

(c) galderari dagozkion erantzunak 11. taulan aurkezten dira; lehen ahotsaren ehunekorik altuena 2 kategorian kokatzen da, bigarren ahotsarena 4 kategorian eta hirugarren ahotsarena 2 kategorian (5. irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 =$ (a.m.: 6) 33,722, $p = 0,000$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena bigarren informatzailearena (217) da, gero hirugarrenarena (170) eta baxuena lehenarena (152).

Ahotsa	1		2		3		4	
	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	20	27,78	30	41,67	16	22,22	6	8,33
2	9	12,16	15	20,27	22	29,73	28	37,84
3	15	20,27	29	39,19	23	31,08	7	9,46

11. taula: Lasaia izaerari dagozkion datuak hiru ahotsak kontuan izanda.

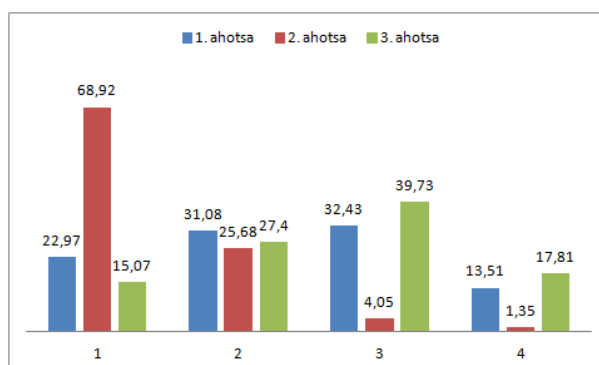


5. irudia: (c) galderari dagokion erantzunak.

(d) galderari dagozkion erantzunak 12. taulan aurkezten dira; lehen ahotsaren ehunekorik altuena 3 kategorian kokatzen da, bigarren ahotsarena 1 kategorian eta hirugarren ahotsarena 3 kategorian (6. irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 = (a.m.: 6) 65,841, p = 0,000$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (190) da, gero lehenarena (175) eta baxuena bigarrenarena (102).

Ahotsa	1		2		3		4	
	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	17	22,97	23	31,08	24	32,43	10	13,51
2	51	68,92	19	25,68	3	4,05	1	1,35
3	11	15,07	20	27,40	29	39,73	13	17,81

12. taula: Harroa izaerari dagozkion erantzunak.

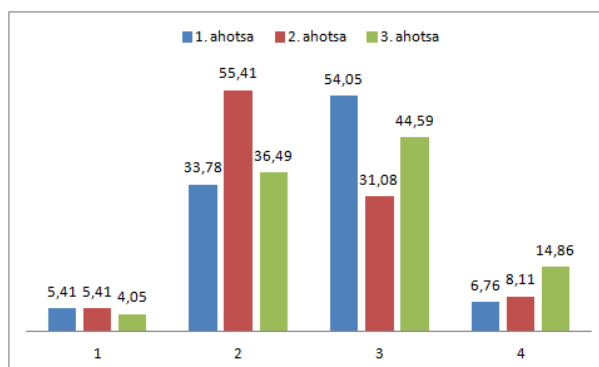


6. irudia: (d) galderari dagokion erantzunak.

(e) galderari dagozkion erantzunak 13. taulan ikusten dira; lehen ahotsaren ehunekorik altuena 3 kategorian kokatzen da, bigarrenarena 2 kategorian eta hirugarrenarena 3 kategorian (7. irudiko grafikoa); aldeak ez dira estatistikoki esanguratsuak. Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (200) da, gero lehenarena (194) eta baxuena bigarrenarena (179).

Ahotsa	1		2		3		4	
	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	4	5,41	25	33,78	40	54,05	5	6,76
2	4	5,41	41	55,41	23	31,08	6	8,11
3	3	4,05	27	36,49	33	44,59	11	14,86

13. taula: Zentzuduna izaerari dagozkion datuak hiru ahotsak kontuan izanda.



7. irudia: (e) galderari dagokion erantzunak.

Azpiatal honetako galderen puntuazio absolutuak erabilia ikus dezakegu zein den, hemen galdetu diren ezaugarrien arabera, bakoitzaren ahotsaren bidez islatzen den informatzaileen balizko izaera (14. taula). Horren arabera f0rik baxuena daukan hirugarren informatzailea da atseginena, ausartena, harroena eta zentzudunena. Lehen informatzailea, f0 ertaina daukana, bigarren lekuan agertzen da (a), (b), (d) eta (e) galderetan. Azkenik, f0rik altuena daukan informatzailea da lasaiena, baina gainerako galdera guztietan puntuaziorik baxuenak dauzkana.

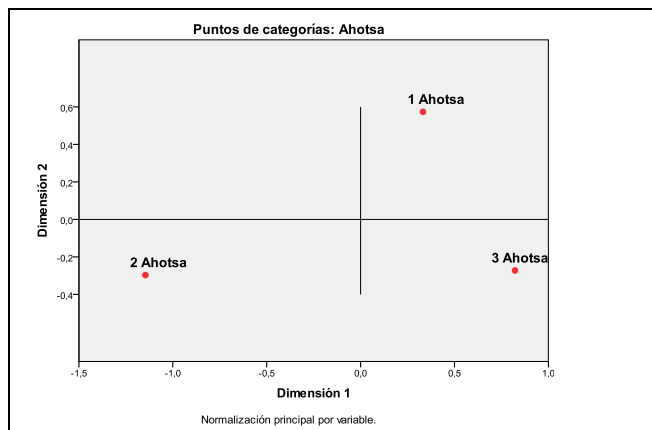
Informatzailea	(a) galdera	(b) galdera	(c) galdera	(d) galdera	(e) galdera
1	164	159	152	175	194
2	177	104	217	102	179
3	189	206	170	190	200

14. taula: Ahotsaren bidez islatzen den informatzaileen balizko izaera.

Informatzaileen ahotsen bidez egiten diren izaeren gaineko inferentzietan jakiteko ahotsak ondo bereizten diren eta aldagai bakoitzaren garrantzia kalkulatzeko korrespondentzia aniztunen azterketa egin dugu. 15. taulan ereduaren laburpena erakusten da. 8. irudiko grafikoan erakusten den bezala, planoan ahots bakoitza ondo bereizita agertzen da.

Dimentsioa	Cronbachen α	Azalduko bariantza	
		Denetara (Autobalioak)	Inertzia
1	0,593	1,901	0,380
2	0,523	1,720	0,344
Denetara		3,622	0,724
Batezbestekoa	0,560	1,811	0,362

15. taula: Ereduaren laburpena.

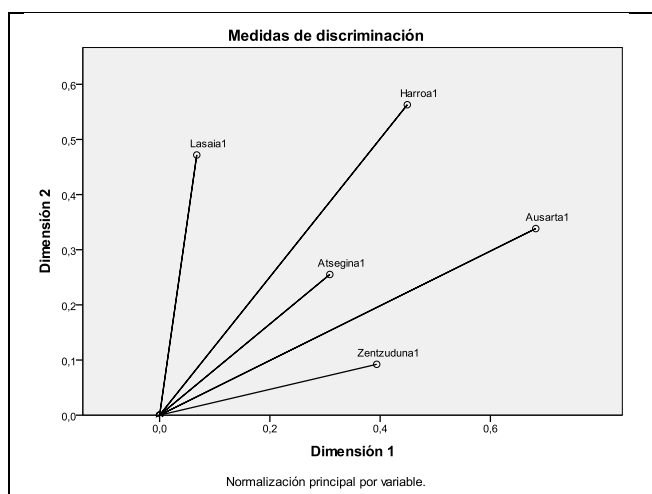


8. irudia: Ahots bakoitzaren tokia planoan.

Diskriminatze neurriak egin den galdera bakoitzaren arabera aztertzen baditugu 16. taulako datuak lortzen dira; Ausarta eta Harroa izatea dira gehien eragiten dutenak. 9. irudiko grafikoan denen eragina erakusten da.

	1. Dimentsioa	2. Dimentsioa	Batezbestekoa
Atsegina	0,309	0,255	0,282
Ausarta	0,682	0,338	0,510
Lasaia	0,067	0,472	0,270
Harroa	0,449	0,563	0,506
Zentzuduna	0,394	0,092	0,243

16. taula: Izaera ezberdinen balioak dimentsio bakoitzean.



9. irudia: Izaera ezberdinen eragina diskriminazioan.

4.2. Itxura

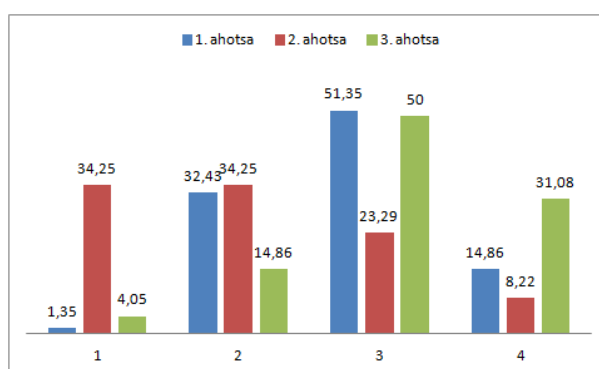
Informatzaileen itxura fisikoaren gainean bost galdera egin ditugu: (a) altua, (b) gihartsua, (c) argala, (d) beltzarana eta (e) erakargarria. Jarraian galdera bakoitzeko entzuleen erantzunen emaitzak aurkeztuko ditugu.

(a) galderari dagozkion erantzunak 17. taulan aurkeztzen dira; ahots bakoitzaren ehunekorik altuena zein kategoriatan gertatzen den aztertzen badugu, ikus dezakegu lehen ahotsaren ehunekorik altuena 3 kategorian kokatzen dela, bigarren ahotsarena 1 eta 2 kategorietan eta hirugarren ahotsarena 3 kategorian (10. irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 = (a.m.: 6) 63,584, p = 0,000$). Puntuazio absolutuei

bagagozkie, puntuaziorik altuena hirugarren informatzailearena (228) da, gero lehenarena (207) eta baxuena bigarrenarena (150).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	1	1,35	24	32,43	38	51,35	11	14,86
2	25	34,25	25	34,25	17	23,29	6	8,22
3	3	4,05	11	14,86	37	50	23	31,08

17. taula: Altua itemari dagozkion erantzunak ahots bakoitzean.

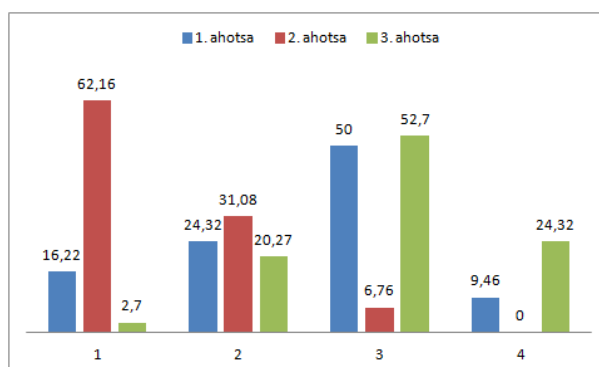


10. irudiko grafikoa: (a) galderari dagozkion erantzunak.

(b) galderari dagozkion erantzunak 18. taulan ematen dira; lehen ahotsaren ehunekorik altuena 3 kategorian kokatzen da, bigarren ahotsarena 1 kategorian eta hirugarren ahotsarena 3 kategorian (11. irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 =$ (a.m.: 6) 101,673, $p = 0,000$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (221) da, gero lehenarena (187) eta baxuena bigarrenarena (107).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	12	16,22	18	24,32	37	50	7	9,46
2	46	62,16	23	31,08	5	6,76	0	0
3	2	2,70	15	20,27	39	52,70	18	24,32

18. taula: Gihartsua itemari dagozkion erantzunak ahots bakoitzean.



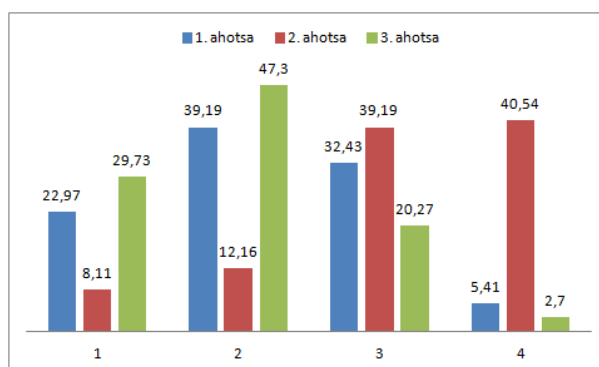
11. irudia: (b) galderari dagozkion erantzunak.

(c) galderari dagozkion erantzunak 19. taulan agertzen dira; lehen ahotsaren ehunekorik altuena 2 kategorian kokatzen da, bigarren ahotsarena 4 kategorian eta hirugarren ahotsarena 2 kategorian (12. irudiko grafikoa); aldeok estatistikoki esanguratsuak dira ($X^2 =$ (a.m.: 6) 69,274, $p = 0,000$). Puntuazio absolutuei bagagozkie,

puntuaziorik altuena bigarren informatzailearena (231) da, gero lehenarena (163) eta baxuena hirugarrenarena (145).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	17	22,97	29	39,19	24	32,43	4	5,41
2	6	8,11	9	12,16	29	39,19	30	40,54
3	22	29,73	35	47,30	15	20,27	2	2,70

19. taula: Argala itemari dagozkion erantzunak ahots bakoitzean.



12. irudia: (c) galderari dagozkion erantzunak.

(d) galderari dagozkion erantzunak 20. taulan ikusten dira; ahots guztien ehunekorik altuenak 1 kategorian kokatzen dira; aldeak, hala ere, estatistikoki esanguratsuak dira ($X^2 = (a.m.: 6) 19,047, p = 0,004$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (138) da, gero lehenarena (109) eta baxuena bigarrenarena (98).

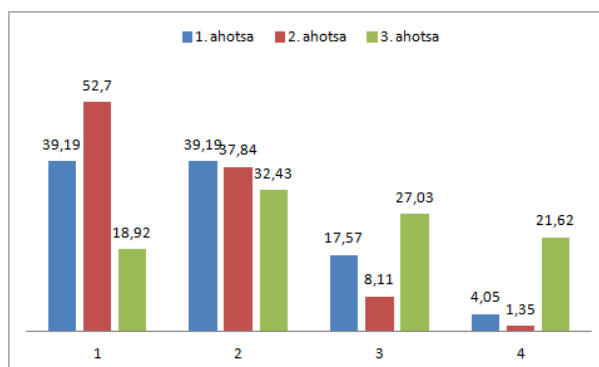
	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	54	72,97	8	10,81	9	12,16	3	4,05
2	56	75,68	12	16,22	6	8,11	0	0
3	39	52,70	16	21,62	9	12,16	10	13,51

20. taula: Beltzarana itemari dagozkion erantzunak ahots bakoitzean.

Azkenik, (e) galderari dagozkion erantzunak 21. taulan erakusten dira; lehen ahotsaren ehunekorik altuenak 1 eta 2 kategorietan kokatzen dira, bigarren ahotsarena 1 kategorian eta hirugarren ahotsarena 2 kategorian (13. irudiko grafikoa); aldeak estatistikoki esanguratsuak dira ($X^2 = (a.m.: 6) 39,542, p = 0,000$). Puntuazio absolutuei bagagozkie, puntuaziorik altuena hirugarren informatzailearena (186) da, gero lehenarena (138) eta baxuena bigarrenarena (117).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	29	39,19	29	39,19	13	17,57	3	4,05
2	39	52,70	28	37,84	6	8,11	1	1,35
3	14	18,92	24	32,43	20	27,03	16	21,62

21. taula: Erakargarria itemari dagozkion erantzunak ahots bakoitzean.



13. irudia: (d) galderari dagozkion erantzunak.

Azpiatal honetako galderen puntuazio absolutuak erabilita, ikus dezakegu zein den, hemen galdetu diren ezaugarrien arabera, informatzaile bakoitzaren itxura fisikoa ahotsaren bidez islatuta (22. taula). Ikusten den moduan, f0rik baxuena daukan informatzailea da altuena, gihartsuena, beltzaranena eta erakargarriena. Galdera hauetan guztietan bigarren lekuan agertzen da lehen informatzailea, hau da, f0 ertaina daukana. Bigarren informatzailea f0rik altuena daukana galdera hauetan azken postuan agertzen da eta argala izatearen kategorian baino ez da agertzen lehenen.

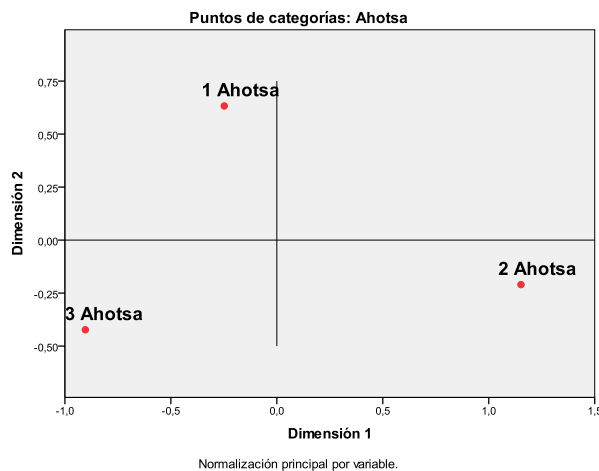
Informatzailea	(a) galdera	(b) galdera	€ galdera	(d) galdera	€ galdera
1	207	187	163	109	138
2	150	107	231	98	117
3	228	221	145	138	186

22. taula: Informatzaile bakoitzaren itxura fisikoa ahotsaren bidez islatuta.

Informatzaileen ahotsen bidez egiten diren euron itxura fisikoaren gaineko inferentzietan jakiteko ahotsak ondo bereizten diren eta aldagai bakoitzaren garrantzia kalkulatzeko korrespondentzia aniztunen azterketa egin dugu lehen bezala. 23. taulan ereduaren laburpena erakusten da. 14. irudiko grafikoan erakusten den bezala, planoan ahots bakoitza ondo bereizita agertzen da.

Dimentsioa	Cronbachen α	Azalduko bariantza	
		Denetara (Autobalioak)	Inertzia
1	0,674	2,172	0,434
2	0,507	1,682	0,336
Denetara		3,853	0,771
Batezbestekoa	0,601	1,927	0,385

23. taula: Ereduaren laburpena.

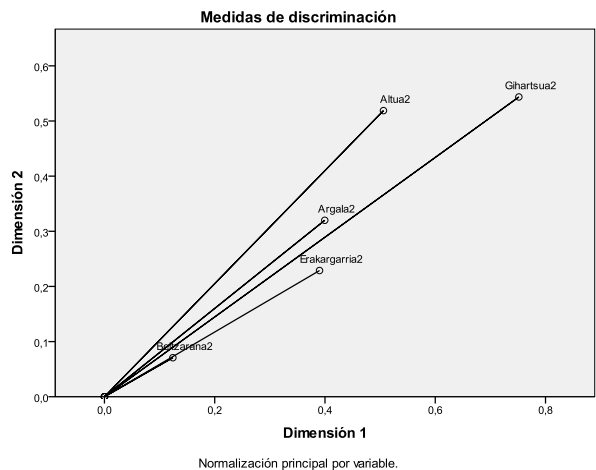


14. irudia: Ahots bakoitzaren tokia planoan.

Diskriminatze neurriak egin den galdera bakoitzaren arabera aztertzen baditugu, 24. taulako datuak lortzen dira; bertan agertzen den moduan, Gihartsua eta Altua izatea dira gehien eragiten dutenak. 15. irudiko grafikoan denen eragina erakusten da.

	1. Dimentsioa	2. Dimentsioa	Batezbestekoa
Altua	0,506	0,519	0,512
Gihartsua	0,752	0,544	0,648
Argala	0,400	0,320	0,360
Beltzarana	0,124	0,071	0,097
Erakargarria	0,390	0,229	0,309

24. taula: Itxuraren itemen balioak dimentsio bakoitzean.



15. irudia: Itxurari dagozkion itemen diskriminazio-neurriak.

4.3. Estatusa

Informatzaileen estatus sozialaren ganean entzuleei egin dizkiegun galderak lau izan dira: (a) goi mailako ikasketak, (b) goi mailako lanbidea, (c) euskara maila eta (d) Bizilekua hau da herri txikia, handia edo hiria.

(a) galderari dagozkion erantzunak 25. taulan batzen dira; ahotsek isla dezaketen ikasketa mailen arabera informatzaile gehienak 2 eta 3 kategorietan banatuta agertzen

dira, agertzen diren aldeak ez dira oso handiak eta ez dira estatistikoki esanguratsuak. Puntuazio absolutuei bagagozkie, ikasketa mailarik altuena hirugarren informatzaileak (198) dauka, gero bigarrenak (189) eta baxuena lehenak (182).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	4	5,48	32	43,84	34	46,58	3	4,11
2	6	8,11	28	37,84	33	44,59	7	9,46
3	5	6,76	21	28,38	41	55,41	7	9,46

25. taula: Goi mailako ikasketak dituzten galderari erantzunak.

(b) galderari dagozkion erantzunak 26. taulan batzen dira; ahotsek isla dezaketen ikasketa mailen arabera informatzaile gehienak 2 kategorian kokatzen dira; aldeak ez dira oso handiak eta ez dira estatistikoki esanguratsuak. Puntuazio absolutuei bagagozkie, lanbiderik onena hirugarren informatzaileak (178) dauka, gero lehenak (171) eta txarrena bigarrenak (157).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	8	10,81	36	48,65	29	39,19	1	1,35
2	17	22,97	32	43,24	24	32,43	1	1,35
3	8	10,81	33	44,59	28	37,84	5	6,76

26. taula: Goi mailako lanbideak dituzten galderari erantzunak.

Euskara mailari dagozkion erantzunak 27. taulan agertzen dira; kasu honetan erantzun gehienak 3 kategorian batzen dira; alabaina, aldeak ez dira estatistikoki esanguratsuak. Puntuazio absolutuei begira euskara mailarik onena hirugarren informatzaileari (211) egozten zaio, gero lehenari (199) eta txarrena bigarrenari (195).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	5	6,76	24	32,43	34	45,95	11	14,86
2	4	5,41	26	35,14	37	50,00	7	9,46
3	5	6,76	15	20,27	40	54,05	14	18,92

27. taula: Ahots ezberdinek duten euskara maila galderari erantzunak.

(d) galderari dagozkion erantzunak 28. taulan batzen dira; ahotsek isla dezaketen ikasketa mailen arabera informatzaile gehienak 2 eta 3 kategorietan banatuta agertzen dira, agertzen diren aldeak ez dira oso handiak eta ez dira estatistikoki esanguratsuak. Edozelan ere, zenbaki absolutuei bagagozkie, ikusten dugu bigarren informatzailea (211), eduki dituen puntuazioen arabera, hiritarragotzat jotzen dela hirugarrena (201) eta lehena baino (190).

	1		2		3		4	
Ahotsa	kopurua	%	kopurua	%	kopurua	%	kopurua	%
1	14	18,92	21	28,38	22	29,73	17	22,97
2	12	16,22	15	20,27	19	25,68	28	37,84
3	15	20,27	17	22,97	16	21,62	26	35,14

28. taula: Ahots ezberdinen bizilekuaren inguruko erantzunak.

Azpiatal honetako aldagaiak estatistikoki esanguratsuak ez diren arren, ikus dezakegu zein den arrakasta sozialaren indizeetan informatzaile bakoitzaren lekua

rankingean puntuazio absolutuen arabera (29. taula). Hirugarren informatzaileak, hau da, f0rik baxuena daukanak puntuaziorik altuenak dauzka (a), (b) eta (c) galderetan.

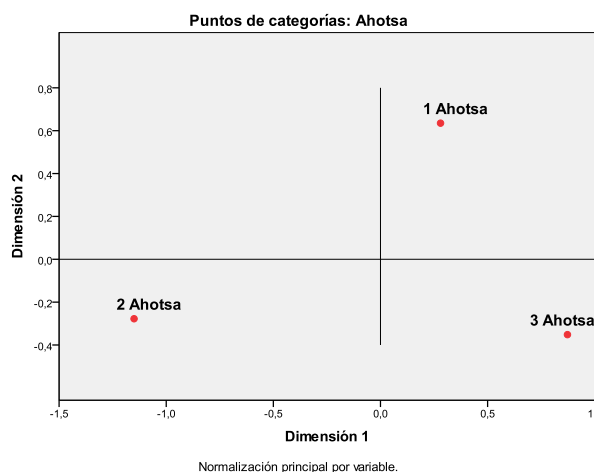
Informatzailea	(a) galdera	(b) galdera	(c) galdera	(d) galdera
1	182	171	199	190
2	189	157	195	211
3	198	178	211	201

29. taula: arrakasta sozialaren indizeetan informatzaile bakoitzaren lekua rankingean puntuazio absolutuen arabera.

Atal honekin amaitzeko, informatzaileen izaeran eta itxura fisikoan estatistikoki esanguratsuak diren datuak aztertuta, korrespondentzia aniztunen azterketaren emaitzak aurkeztuko ditugu jarraian ahotsak ondo bereizten diren ikusteko baita jakiteko bereizkuntzan zeintzuk diren gehien eragiten duten galderak ere. Izaeraren gainean egineko galderetan estatistikoki esanguratsuak (b) Ausarta, (c) Lasaia eta (d) Harroa izan dira eta itxura fisikoaren gaineko galderetan bostak. 30. taulan ereduaren laburpena erakusten da. 16. irudiko grafikoan erakusten den bezala, planoan ahots bakoitza ondo bereizita agertzen da.

Dimentsioa	Cronbachen α	Azalduko bariantza	
		Denetara (Autobalioak)	Inertzia
1	0,797	3,306	0,413
2	0,605	2,125	0,266
Denetara		5,431	0,679
Batezbestekoa	0,722	2,716	0,339

30. taula: Ereduaren laburpena.

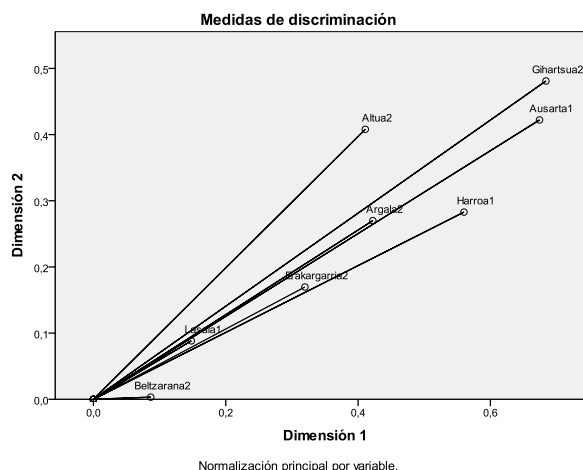


16. irudia: Planoan ahots bakoitzaren tokia.

Diskriminatze neurriak egin den galdera bakoitzaren arabera aztertzen baditugu 31. taulako datuak lortzen dira; Gihartsua eta Ausarta izatea dira gehien eragiten dutenak. 17. irudiko grafikoan den eragina erakusten da.

	1. Dimentsioa	2. Dimentsioa	Batezbestekoa
Ausarta	0,674	0,422	0,548
Lasaia	0,148	0,088	0,118
Harroa	0,560	0,283	0,421
Altua	0,411	0,408	0,409
Gihartsua	0,684	0,481	0,583
Argala	0,422	0,270	0,346
Beltzarana	0,087	0,003	0,045
Erakargarria	0,320	0,170	0,245

31.taula: Diskriminatze neurriak egin den galdera bakoitza kontuan izanda.



17. irudia: Itemen diskriminazio-maila.

5. Ondorioak

Hiru izari akustikoetan oinarrituta (F0, energia eta iraupena) mutil gazteen grabazioen bost tasunen eragina eta aldeak aztertu dira (Xf0, Hf0, Xdb, Hdb eta iraupena), horrenbestez, ahots bakoitzaren ezaugarriak ondorengo lerroetan azaltzen dira.

Lehen ahotsak du iraupenik luzeena, batez ere etenen denborarik luzeena egiten du, gainera, hizketa abiadarik handiena ere badu eta energia batez besteko altuena. Bigarren ahotsak, batez besteko oinarrizko maiztasun altuena du, ahots agudoena da, eta lasaiena, argalena eta besteak baino hiritarragotzat jotzen da. Hirugarren ahotsari dagokionez, batez besteko oinarrizko maiztasunik baxuena du, hizketa denborarik luzeena ere badu, baina eten denborarik txikiena, hizketa abiadarik txiena ere badu, esan daiteke ahotsik grabeena duela. Izaerari, itxurari eta estatusari dagokionez, jasotako puntuazio absolutuen arabera, era honetan pertzibitu dute entzun duten neskek: atseginena, ausartena, harroena, zentzudunena, altuena, gihartsuena, beltzaranena, erakargarriena, ikasketa mailarik altuena, goi mailako lanbidea duen mutila, euskara mailarik onena duena.

Beraz, ondorioztatu daiteke batez besteko oinarrizko maiztasunik baxuena eta horrenbestez ahotsik grabeena duen mutilaren ahotsak puntuazio absoluturik altuenak lortu dituela hiru ataletan; alegia, izaeran, itxuran eta estatusean.

6. Bibliografia

- Boersma, P. eta Weenink, D. (2013). *Praat: doing phonetics by computer*, Version 5.1. [Computer program]. Recuperado de <http://www.praat.org>
- Del Olmo, M. (2005). Prejuicios y estereotipos: un replanteamiento de su uso y utilidad como mecanismos sociales. *XXI Revista de Educación*, 7, 13-23.
- Gaminde, I. (2007). Tonuak eta etenak Bilboko gaztelaniaz eta euskaraz. *Fontes linguae vasconum: Studia et documenta*, 106, 455-474.
- Gaminde, I. (2010). *Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz*. Bilbao: Mendebalde Kultura Alkartea y Bizkaiko Foru Aldundia.
- Gaminde, I., Aurrekoetxea, G., Etxebarria, A., Garay, U., & Romero, A. (2014). *Ahoskera lantzeko argibideak eta jarduerak: Laguntzarako materiala: teoria eta praktika*. Leioa: UPV/EHU.
- Garay, U., Etxebarria, A., Gaminde, I. Larrea, K. eta Romero, A. (2011). Kultura arteko komunikaziogaitasuna: emozioen pertzepzioa hizkuntza etorkinetan. *Tantak*, 23(2), 77-96.
- Mora, E., & Cobeta, I. (2013). Voz en el cambio de género. In Cobeta, I., Nuñez, F., Fernández, S. (Eds.), *Patología de la voz* (pp. 313-322). Barcelona: ICG Marge.
- Moreno A., Álvarez, M., Bejarano, M.A., & Pulido, C.A. (2010). parámetros acústicos de la voz en el adulto mayor. *Umbral Científico*, 17, 9-17.
- Navas, E., Hernáez, I., Castelruiz, A., Sánchez J., & Luengo, I. (2004). Acoustic Analysis of Emotional Speech in Standard Basque for Emotion Recognition. In A. Sanfeliu et al. (Eds.), *CIARP 2004, LNCS 3287* (pp. 386-393). Berlin: Springer-Verlag.
- Navas, E., Hernáez, I., Luengo, I., Sánchez, J. & Saratxaga, I. (2005). Analysis of the Suitability of Comm on Corpora for Emotional Speech Modeling in Standard Basque. In V. Matousek et al. (Eds.), *Lecture Notes on Artificial Intelligence* (265-272), vol. 3658.
- Navas, E; Hernáez, I.; Luengo, I.; Sainz, I; Saratxaga, I. & Sanchez, J. (2007). Meaningful Parameters in Emotion Characterisation. In A. Esposito et al. (Eds.), *Verbal and Nonverbal Commun Behaviours. Lecture Notes on Computer Science* (74-84) vol. 4775.
- Sainz, I., Saratxaga, I., Navas, E., Hernáez, I., Sánchez, J., Luengo, I., & Odriozola, I. (2008). Subjective evaluation of an emotional speech database for Basque. En *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*.
- Sanabria, J., García, P., & Ahumada, F. (2013). Disfonías de origen neurológico. En Cobeta, I., Nuñez, F., Fernández, S. (Eds.), *Patología de la voz* (pp. 283-294). Barcelona: ICG Marge.

THE ACOUSTIC REALIZATION OF /s/ AND /tʃ/ BY L1 AND L2 BASQUE SPEAKERS.

Ander Beristain
 University of Illinois at Urbana-Champaign
anderb2@illinois.edu

Laburpena

The Basque language possesses a relatively rich system of sibilant consonants, including three voiceless fricatives, orthographically represented as <z> (dental), <s> (alveolar) and <x> (prepalatal), and the three affricates (<tz>, <ts> and <tx>). In addition, there is a voiceless palatal stop <t> and some dialects have added voiced sibilants. This system has gone through multiple neutralization processes regarding some of the contrasts in several geographical areas one of which is Gernika Basque, in Bizkaia. Two of these processes, *seseo* (merging of /s/ <s> and /s/ <z> into /s/) and *tzetzeo* (merging of /tʃ/ <ts> and /tʃ/ <tz> into /tʃ/) are relatively old (Hualde, 2010). A recent neutralization is that, whereas older speakers from Gernika distinguish six phonemes, /s/, /ʃ/, /z/, /tʃ/, /tʃ/ and /c/, some younger speakers may have a single fricative /s/ and a single affricate /tʃ/ instead (Ensunza, 2012). The aim of this project is two-fold: (1) To identify the main acoustic differences in the production of /s/ and /tʃ/ by L1-Basque (G1) and L2-Basque (G2) speakers in Gernika Basque and Batua speakers in Bizkaia; and (2) to determine a possible influence from Spanish. The participants are five females: one L1-Gernika Basque speaker, and four L2-Batua speakers from the Greater Bilbao Area. The data come from natural speech data (Rodríguez-Ordóñez, 2016). Via Praat, the first 100 tokens of /s/ and /tʃ/ in the interviews were coded. The dependent variables are the center of gravity (COG), intensity, duration and amount of voicing of the fricative. The independent variables are the L1 (Basque or Spanish), linguistic context (initial, intervocalic, pre- and post-consonant and final) and phoneme (fricative or affricate). Mixed effects linear regression models were used under the package *lme4* (REF) in R to attest the significance of the variables. The interaction of L1 and COG is significant, $p < .05$. Within COG, context is significant ($p < .05$), since post-consonant position seems to be conducive to retaining the higher frequencies that characterize Spanish sibilants. The following vowel is significant as well, $p < .01$. Voicing is significant within /tʃ/ ($p < .05$), where the means are 0.43/1 for G1 and 0.22/1 for G2. Finally, intensity and duration are not significant. It is concluded that L2-Basque speakers have not assimilated the realization of Basque /s/ and /tʃ/. Higher COG frequency rates in G2 demonstrate the retention of Spanish sibilant features, showing a contact effect from Spanish.

Hitz-gakoak: L2 Basque, sibilants, acoustic phonetics, contact between Basque and Spanish.

1. Introduction⁴

The study of sibilant sounds in languages has for long been a topic of interest for linguists. Although these sounds are common in most languages, the number of sibilants within each language is not usually large. According to Jurado (2011), in a study that englobed an extensive number of languages, 83% of them had a type of /s/ in their phonemic inventory. These sibilant sounds were, generally, either dental or alveolar. Nevertheless, rarely did they have three contrastive articulation points such as in Basque.

⁴ I am grateful to Dr. Itxaso Rodríguez-Ordóñez for giving me access to her data; to Dr. Chris Eager for writing the script in *Praat*; to Prof. Aurore Mroz for her valuable feedback on earlier versions of the manuscript; to Amalia Reyes for helping me with the statistical analysis; and to Prof. José Ignacio Hualde and Prof. Ryan Shosted for their advice on research procedures.

The two languages in contact considered for this project are Peninsular Spanish and Basque. Peninsular Spanish has two sibilants, which are the voiceless alveolar fricative [s], as in *sal* [sal] (salt) and the voiceless pre-palatal affricate /tʃ/, as in *hacha* [atʃa] (axe) (Hualde, 2014: 32). Comparatively, Standard Basque has six, from which three are fricative and the remaining three are affricate. Table 1 shows sibilants in Standard Basque with their respective features, orthographic representations, and examples within words:

Phoneme	Place of articulation	Manner of articulation	Voicing	Orthographic representation	Examples in Standard Basque
/s/	Apico-alveolar	Fricative	Voiceless	<s>	<i>soka</i> [soka] (rope)
/ʃ/	Lamino-alveolar	Fricative	Voiceless	<z>	<i>zopa</i> [ʃopa] (soup)
/f/	Pre-palatal	Fricative	Voiceless	<x>	<i>xagu</i> [fayu] (mouse)
/tʃ/	Apico-alveolar	Affricate	Voiceless	<ts>	<i>atso</i> [atʃo] (old woman)
/tʃ/	Lamino-alveolar	Affricate	Voiceless	<tz>	<i>atzo</i> [atʃo] (yesterday)
/tʃ/	Pre-palatal	Affricate	Voiceless	<tx>	<i>etxe</i> [etʃe] (house)

Table 1. Sibilant system in Standard Basque (adapted from Hualde, 2003: 16)

Therefore, it can be perceived that the absence of certain sibilant sounds in the phonemic inventory of Peninsular Spanish may lead to difficulties for L2-Basque learners. Particularly, due to acoustic and articulatory features of /ʃ/ and /tʃ/, attaining a native-like pronunciation of them seems to be difficult for L2-Basque learners.

The aim of this project is three-fold, and will attempt to answer the following research questions:

- **RQ #1:** What are the main acoustic differences found in the production of /ʃ/ and /tʃ/ by L1 and L2-Basque speakers?
- **RQ #2:** Within the L2-Basque group, can a possible cross-linguistic influence from Spanish be determined?
- **RQ #3:** What are the “*floor*” and “*ceiling*” performance levels⁵ for L1-Spanish speakers with regards to the production of /ʃ/ and /tʃ/?

2. Literature review

2.1. Theoretical framework

Language production can be regarded from a cognitivist perspective, holding that the brain will be primarily responsible for the realization of specific linguistic functions. These linguistic functions can be understood as either individual or social. For the aim of this project, the individualistic standpoint is adopted.

The existence of a critical period for language acquisition seems to be a controversial topic within the linguistic field, and as MacSwan & Pray (2005) attest, time is relevant in order to acquire language properly. Research has shown that each language aspect shows a different critical period, having phonology the shortest one (Pallier, Bosch, & Sebastian-Gallés, 1997).

⁵ ACTFL (2012: 7) defines the *floor* performance level as a “level which can be handled successfully” and the *ceiling* as the “limits of the speaker’s proficiency, i.e., the patterns of weaknesses”.

Furthermore, Natural Phonology states that speakers are aware of the correlation between the mental representation of produced sounds and their articulatory and acoustic features (Donegan & Stampe, 2009). As Oñederra, Jauregi & Epelde (2014) put, if a speaker is trying to discriminate a specific sound from another language, the previously acquired knowledge on sounds will be utilized. Therefore, it can be hypothesized that the expected Basque sibilant sounds under study will be realized as near representations of existing sounds in Spanish.

This also correlates with Flege's (2007) Speech Learning Model, that explains that the sound inventory of the L1 will influence the production of the L2. Furthermore, if a given sound is not present in the L1 but it appears in the L2 and is, phonetically speaking, dissimilar to that of the L1, the speaker will eventually be more prone to producing such sound (Flege, 2007). Nevertheless, if the distinct L2 and L1 sounds are similar to one another, the speaker might find it more complicated to produce, since it assimilates to that of the L1.

2.2. *Sibilants in research*

Because of the difficulty in the acoustic study of such sounds, research conducted on sibilants has not been proliferous. However, because of the relatively rich system of sibilants that Basque possesses, studies on this type of sounds and their acoustic properties have been thoroughly studied (see Urrutia, Etxebarria, Turrez, & Duque, 1988, 1989, 1991). Some studies point out that Basque has an influence on the Spanish spoken in the Basque Country (Iribar & Isasi, 2008; Iribar, Isasi, Gómez Seibane, & Moral del Hoyo, 2005; Isasi, Iribar, & Moral del Hoyo, 2009). Others have brought to light neutralization phenomena such as *seseo* and *tzetzeo* in Basque sibilants (Hualde, 2010; Beristain, in preparation). For the aim of this project, both the aforementioned phonological processes will be of relevance, since the Basque variety spoken in the area under study has undergone both of them.

2.3. *Neutralization in Basque sibilants*

According to Hualde (2010), Basque sibilants have undergone neutralization processes. On the one hand, regarding fricative sibilants /s/, /ʃ/ and /f/, three possible mechanisms have been found. *Seseo* implies that the sounds /s/ and /ʃ/ have merged into /s/. Furthermore, Hualde (2010) explains that more local neutralization processes have also occurred in localities such as Azpeitia and Azkoitia. In the concerned areas, *zezeo* is the most overspread process among fricatives, i.e., the merging of the sounds /s/ and /ʃ/ into the lamino-alveolar /s/. In addition, although not so prevalent, it is explained that Basque varieties in contact with French have undergone processes of *xexeo*, i.e., the neutralization of /s/ and /f/ into /f/. Regarding affricate sounds, *tzetzeo* is the most overspread process, and it is defined as the merging of the affricates /tʃ/ and /tʃ/ into /tʃ/. Since the area under study in this project is Bizkaia, where *seseo* and *tzezeo* are most prominent, the remaining processes will be overlooked for the aim of this project.

According to Iribar & Isasi (2008) and Iribar et al. (2005), *seseo* appears in a vast geographical area in Bizkaia (Western Basque Country). Consequently, the distinction between alveolar and dental sibilant sounds is non-existent in this area. However, different processes are found in central and north-eastern parts of the Basque Country. As illustrated by Jurado (2011), the complete distinction between the canonical six-

sibilant system is still maintained in several dialects in Gipuzkoa (Central Basque Country).

2.4. Previous studies in Basque sibilants

Research has primarily focused on the production and analysis of Basque sibilants by native speakers of Basque. Larraza (2015) and Jurado (2011) also included productions by Spanish/Basque speakers, but these were balanced bilingual speakers. Nevertheless, the production of such sounds by L2-Basque learners seems to have been overlooked. It is known that having multiple languages in contact at the time of learning a new language can be significant, since the L1 can have major effects on the L2 (Etxebarria, 1997; Oñederra, Jauregi, & Epelde, 2014). That is the factor that is intended to be analyzed in this project.

Iribar et al. (2005) were pioneers in the acoustic study of Basque sibilants and *seseo* specifically. In their study, they intended to analyze the production of [s] in a group of Basque-dominant speakers. They wanted to investigate whether the productions of the [s] sounds in Basque and Spanish were ‘equal’, or at least similar as far as their acoustic features were concerned. Experts debate whether the realizations of the sibilant sound [s] in Spanish and Basque are acoustically the same. According to Iribar et al. (2005), it can be inferred that there are minor differences among the two sounds. They attest that the Spanish [s] has higher values of intensity in general, yet with regard to its place of articulation, the alveolar [s] of the Spanish is realized as an apical sound (like the Basque /s̺/). Thus, it can be deduced that Basque-dominant speakers maintain the articulatory features of the Basque voiceless apical fricative when they speak Spanish. These results were confirmed by Iribar & Isasi (2008), who included more dialects of Basque.

The most relevant study for this project is that of Larraza (2015), who compared the perception and production of the Basque sibilants /s̺/, /s̺̺/, /ts̺/ and /ts̺̺/ between early Spanish-Basque bilinguals and late bilinguals who learned Basque at the age of 15, 20 and 22 respectively. According to her results, late bilinguals were not able to produce the voiceless lamino-alveolar affricate sound /ts̺̺/. Their realizations varied in a continuum of the respectively apico-alveolar and pre-palatal affricates /ts̺/ and /tʃ/. The purpose of this study will be to confirm or disconfirm these results.

As the number of learners of Basque has increased in recent years, materials that contribute to a better acquisition of the language have been created as well (Gaminde, Aurrekoetxea, Etxebarria, Garay, & Romero, 2016; White, 2008). Thus, this project will intend to bridge the research gap that exists in the production of Basque sibilants by L1-Spanish speakers that are not balanced bilinguals, but learners whose domain of Basque is limited. Also, from an educational standpoint, possible *floor* and *ceiling* performance levels will be described as far as L2-Basque sibilants are concerned.

3. Methodology

3.1. Overview of design

The first step of this process was to observe the different perspectives on Basque *seseo* and *tzetzeo*. Based on previous studies on Basque sibilants (Hualde, 2010; Iribar

& Isasi, 2008; Iribar et al., 2005; Isasi et al., 2009; Larraza, 2015; Medina del Moral & Romera, 2016), a decision was made to replicate their methodology, adding the L2-Basque factor.

At stage 2, access to the dataset of Rodríguez-Ordóñez (2016) was obtained. From a total number of 99 participants, only a sub-sample of one L1-Basque speaker and four L2-Basque female speakers was selected. Since previous studies showed that duration, center of gravity (COG), intensity, and amount of voicing were all critical for the study of sibilants, these same factors were used in the analysis phase of this project (Jongman, Wayland, & Wong, 2000). After segmenting and annotating the data via *Praat* (Boersma & Weenink, 2016) and obtaining values from a script, results were observed via *R* (R Core Team, 2016). At stage 3, following upon these results, possible future educational implications were drawn.

3.2. Sampling

The sampling procedure of this project relied on that operated by Rodríguez-Ordóñez (2016) for her dissertation titled *Direct Object Marking in Basque: Grammaticalization, Attitudes and Ideological Representations*. Direct Object Marking in Basque was studied across three groups: Basque-Spanish bilingual speakers, Basque-French bilingual speakers, and late learners of Basque. A total number of 99 participants were selected for her study.

Since the aim of this project is to assess the oral efficacy of L2-Basque production, only a sub-sample from Rodríguez-Ordóñez's (2016) L2-Basque group was used. Thus, as the L1-Basque data belonged to a female participant, only female participants were elected in the L2 group in order to facilitate data normalization. Also, since in Rodríguez-Ordóñez (2016) L2-Basque learners were divided into advanced, intermediate and novice levels, a decision was made to only select participants within the same group. As error production in the advance learners was not salient, yet it was excessively evident in the novice group, the intermediate group was chosen as the optimal sub-sample for this project. Within this group, the decision was made to select four female speakers. Therefore, the new sample included a total number of four L2-Basque speakers and one L1-Basque speaker. These participants belong to two different linguistic groups: those for whom Basque is an L1 (Group 1), serving as control, and those for whom Basque is an L2 (Group 2). Table 2 summarizes the relevant information for each participant.

Participant	Profile	Area	Token N. N=600	N. of years learning Basque	Age
1	L1-Basque	Gernika	200	native	27
2	L2-Basque	The Greater Bilbao	100	5	40-45
3	L2-Basque	The Greater Bilbao	100	5	18
4	L2-Basque	The Greater Bilbao	100	15	21
5	L2-Basque	The Greater Bilbao	100	5	40-45

Table 2. Participants and their backgrounds

Participants were from nearby areas within Bizkaia, Western Basque Country. The only participant in Group 1 is from Gernika, a semi-urban town which hosts 16,000 inhabitants, and is 21 miles away from the capital of the Biscayan Province of Bilbao.

Participants in Group 2 are from different areas within the Greater Bilbao area. This area hosts around 950,000 inhabitants, and it is the metropolitan area that includes the city of Bilbao and its suburbs (Rodríguez-Ordóñez, 2016).⁶

3.3. Data collection

In order to elicit as much spontaneous speech as possible, sociolinguistic oral interviews were used. Rodríguez-Ordóñez (2016) spent between 30 to 45 minutes interacting with the participants both individually and in pairs in some occasions.

Topics employed in the conversation were everyday-like topics. Participants were encouraged to give anecdotes, summer plans or personal views on diverse topics such as Basque language and culture. For Rodríguez-Ordóñez's (2016) own purposes, participants were interviewed both in Spanish and Basque. Nevertheless, this project has only utilized the Basque parts of these audio samples.

Furthermore, as some recordings were longer than others, the decision was made to only code the first 100 tokens of /s/ and /tʃ/ in the interviews for validity and reliability purposes. As the balance between the number of participants in Group 1 and Group 2 was uneven (1 to 4, respectively), 200 tokens were coded in the speech of the only participant in Group 1.

3.4. Data analysis

The realizations of /s/ and /tʃ/ were analyzed via *Praat*. The first step was to segment and annotate the realizations of the sounds by means of the *TextGrid* function in *Praat*. In order to have a clearer visualization of the analysis, three tiers were created. A visual description is provided in Figure 1.

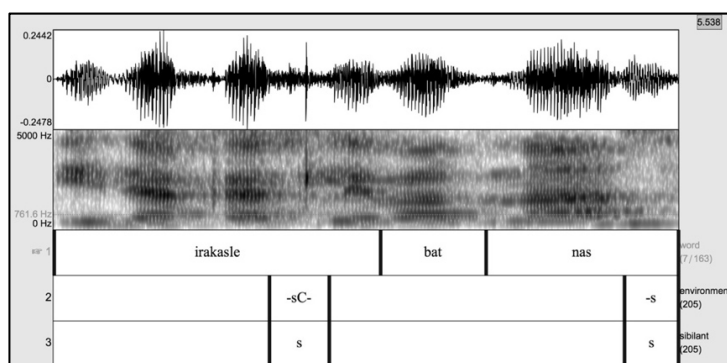


Figure 1. Screenshot of a segmentation and annotation in Praat

The first tier was used to transcribe the words that contained the sibilant sounds under study. The spelling in the first tier does not correspond to Standard Basque, but it represents the spelling that would correspond to the way in which the participant pronounces that specific word. For instance, tier 1 in Figure 1 shows *irakasle bat nas* [I am a teacher], which in Standard Basque is *irakasle bat naiz*. That way, the correlation between the use of <s> in the spelling and *seseo* is apparent.

⁶ Like in Gernika, Basque started to be officially taught in schools after the 1970's. Nevertheless, because of its metropolitan status, Spanish has gained prominence and the use of Basque is less frequent in this area than it is in Gernika.

Tier 2 includes the possible linguistic environments for each sound to appear. Initially, all possible contexts were considered: initial (s-/tz-), intervocalic (-s-/tz-), pre-consonantal (-sC-/tzC-), post-consonantal (-Cs-/Ctz-) and final position (-s-/tz). Nevertheless, since some of these contexts resulted to be scarcely recurrent, they were overlooked with regards to final results. This was the case for -tzC-, as this type of consonant cluster hardly ever appears in Basque. It needs to be mentioned that the intervocalic environment is not limited to word boundaries. If the context for a given sibilant is word-final, but the adjacent word starts with a vowel, this environment is considered intervocalic. As shown in Figure 2, the final [s] in *nais* is considered intervocalic because the initial sound in the proceeding segment *eta* [eta] is [e].

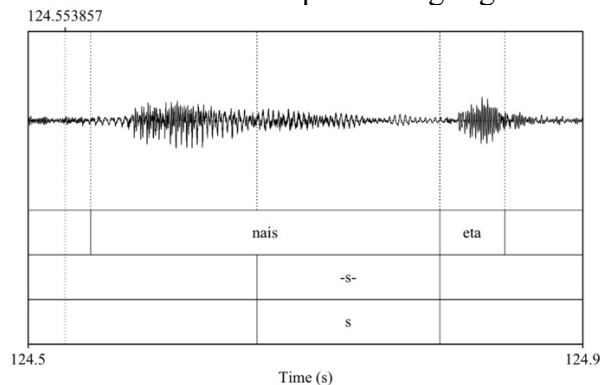


Figure 2. Example of an across-word intervocalic environment in Praat

Tier 3 was used to mark the specific sibilant sounds to be studied in this project, that is, /s/ and /ʃ/ (note that the last sound is not /tʃ/, but only /ʃ/). This is because /tʃ/ is an affricate sound and the values for intensity, duration, COG, and amount of voicing of the initial stop [t] would interfere with the results of the proceeding fricative /ʃ/. Thus, only the sibilant-fricative segments were analyzed.

Next, using a script that was specifically written for this project, values for the COG, duration, intensity, and amount of voicing of fricative were obtained. Table 3 shows an example of such values.

Participant	Phoneme	COG (Hz)	Duration (ms)	Intensity (dB)	% voicing (/1)
1	s	4037.281	91.411	50.111	0
1	s	4437.478	66.886	59.001	0.44

Table 3. Example of distribution of results

Once the data were coded, mixed effects linear regression models were used under the package lme4 (REF) in R (R Core Team, 2016) to attest the significance of the variables.

4. Results

4.1. Center of gravity (COG)

As expected from previous studies, COG had a significant impact on the interaction between the L1 and the context of the two phonemes to be analyzed ($p < .05$). The mean value for Basque /s/ was 3334.961 Hz whereas the Spanish group's was 4398.673 Hz. Moreover, the mean for /ʃ/ in the Basque group was 2625.148 Hz,

whereas it was 4843.631 Hz in the Spanish group. After data normalization, as shown in Figure 3, the mean values for the Basque group were negative, whereas that of the Spanish group were positive. This implies that while the control group (L1-Basque) clearly realizes /s/ and the fricative in /ts/ as different, the L2-Basque group has neutralized the distinction between the two fricatives.

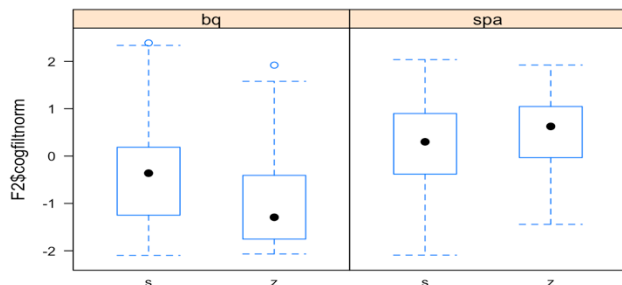


Figure 3. COG values by phonemes within the group (where s=/s/ and z=/tʃ/)

The interaction between context and the phoneme was significant ($p < .05$). As shown in Figure 4, mean values for the contexts within the Basque group are near the threshold of 0, or below it, whereas that of the Spanish group is above 0.

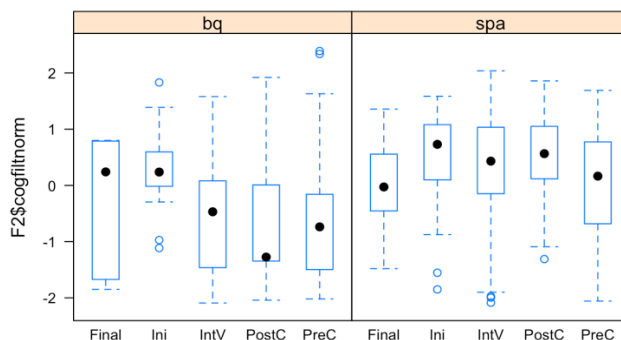


Figure 4. COG values by context within the group (where: Final = final position; Ini = initial position; IntV = intervocalic position; PostC = post-consonantal position; and PreC = pre-consonantal position)

Since previous research showed that the following vowel was significant in COG values (Moral del Hoyo & Romera, 2016; Quilis, 1981), the same procedure was replicated. The three corner vowels, /i, a, u/, were considered and results confirmed previous studies. The significance for the following vowel is $p < 0.01$, which is the same as the interaction between the following vowel and the L1. As shown in Figure 5, the Spanish group produces sibilants in a higher frequency in corner vowels. Values for the frequencies of each vowel and group are shown in Table 4.

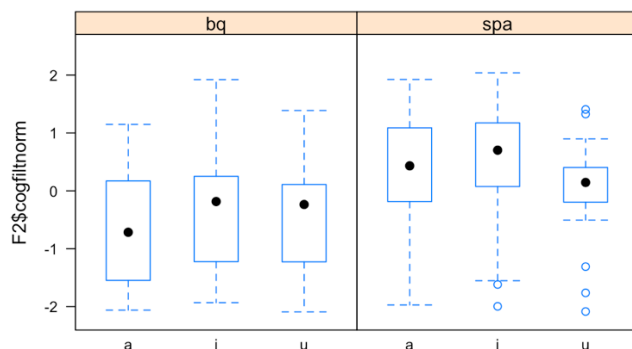


Figure 5. COG values by following vowel within the group

	/a/	/i/	/u/
Basque	2975.294 Hz	3599.601 Hz	3282.213 Hz
Spanish	4582.965 Hz	4975.686 Hz	4239.583 Hz

Table 4. COG frequencies by following vowel within the group

4.2. Voicing of fricative

Voicing of fricative was only found to be significant for /s/ <z> (p<.05). The mean value for /s/ in the Basque group is 0.43/1, whereas that of the Spanish group is 0.22/1. Context was significant as well, with a value of p<.05. As shown in Figure 6, final and post consonantal context seem to favor the voicing of the lamino-alveolar sibilant.

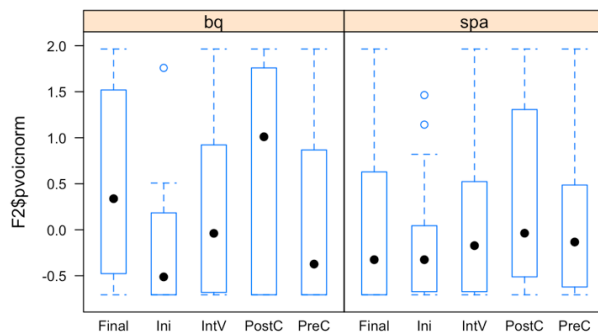


Figure 6. Voicing values by context within the group

4.3. Duration

Duration was not significant in any of the groups, either by phoneme or context. Both L1 and L2-Basque groups showed very similar results, with 79.160 ms and 82.651 ms as mean values for the /s/ in L1 and L2-Basque respectively, and 47.795 ms and 42.333 ms for /z/ in L1 and L2-Basque respectively. Figure 7 shows the normalized values of duration.

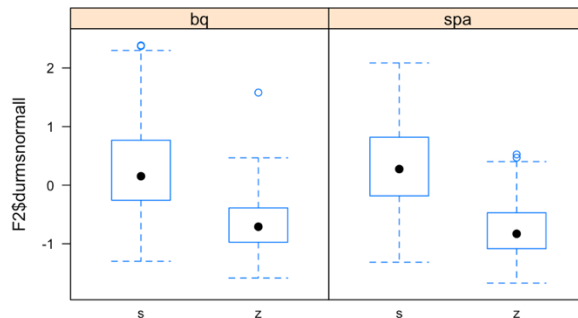


Figure 7. Duration values by phoneme within the group

4.4. Intensity

As shown in Figure 8, values of intensity regarding phoneme within L1 were similar, and were found to be non-significant. Mean values for /s/ were 55.571 dB in the Basque group and 54.991 dB in the Spanish group, whereas /z/ obtained a mean of 56.402 dB in the Basque group and 53.034 dB in the Spanish.

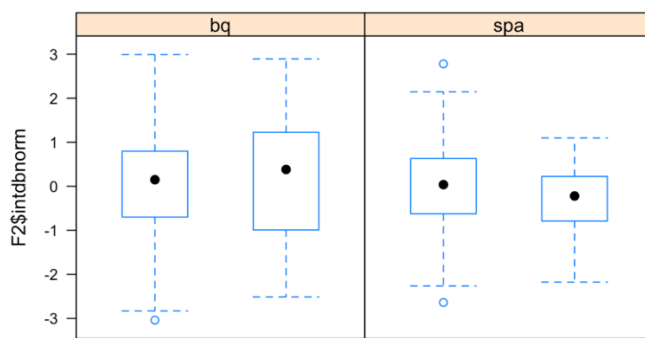


Figure 8. Intensity values by phoneme within the group

However, the interaction between the context and the phoneme was found to be significant ($p < .01$). As illustrated in Figure 9, intensity values seem to correspond between Basque and Spanish groups, yet, the post-consonantal context showed greater variation.

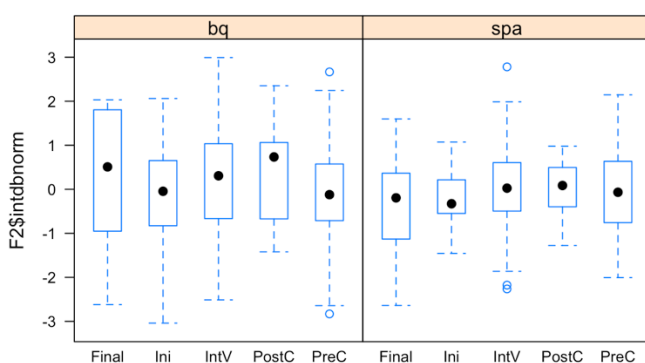


Figure 9. Intensity values by context within the group

5. Discussion

Based on the previous results, it has been proven that L2-Basque speakers have not assimilated the realization of Basque /s/ and /tʃ/, and that contact with Spanish had an impact. As for the apico-alveolar /s/, COG values show that L2-Basque speech is characterized by higher frequency rates, which confirms Iribar et al.'s (2005) results. Furthermore, since /tʃ/ is not present in the Spanish phonological system, it is produced as the closest sound for L1-Spanish speakers, i.e., /tʃ/, which is a highly-stigmatized marker for /tʃ/.

Although COG results do not coincide with the previous research (Medina del Moral & Romera, 2016; Quilis, 1981), it can be explained by the fact that participants in their projects included both males and females, and it is known that male participants' speech frequency is different to females'. Nevertheless, what is significant in this project is that frequency values are higher in all cases of L1-Spanish speakers. This confirms Jurado's (2011) results, namely that Spanish sibilants are characterized by higher frequency rates than those of Basque, proving that the Spanish phonological system of L2-Basque speakers is retained in L2-Basque speech.

Furthermore, although mean values obtained in the Spanish group for /tʃ/ do not coincide with those of Hualde (2010), this should not be unforeseen, since Hualde (2010) dealt with L1-Basque speakers solely. What is significant to this project is that

the mean value of /tʃ/ in the L2-Basque group is closer to that of /tʃ/ (3829.000 Hz) than that of /tʃ/ (7183.700 Hz) in Hualde's (2010) study.

Results show that L2-Basque speakers are far from the ceiling level of Basque sibilant production. The different boxplots show that, although some contexts might be more conducive for a native-like pronunciation, they have not achieved it yet. The acoustic differences are so minimal that L2-Basque speakers might not be able to discriminate between the two sounds. Regarding the floor level, two different ones can be established. That of the fricative sound is higher than that of the affricate sound since the similarities between native/non-native production are more prominent in the case of the fricative. However, the affricate one is generally produced with a different place of articulation, as a pre-palatal instead of lamino-alveolar which is striking to a native's ears.

5.1. Educational implications

It is proposed that pronunciation should have a more prominent role in educational settings, since it is seen that immersion alone is insufficient to help the correct realization of these phonemes. Because of the acoustic and articulatory features of each sound, the procedure to be followed would change from one sound to the other. Regarding /s/, it needs to be said that the acoustic differences between this sound and the Spanish /s/ are minimal (Beristain, in preparation). Therefore, focus should be on perception first, and once students acquire an advanced level, they should proceed to production. As far as /tʃ/ is concerned, perception of such sound is present since the beginning. Therefore, focus should be on production, which could be addressed from an intermediate level on. Gaminde et al. (2016) have opened the way to an innovative instructional way to acquire Basque pronunciation. It will be interesting to see whether explicit instruction helps L2-Basque learners overcome phonological difficulties.

6. Venues for future research

One of the biggest limitations of this project has been the low number of participants (N=5) and tokens (N=620), and the fact that only female participants were included in the sample. For future projects, a bigger sample with a gender-balanced selection of participants should be employed.

Nonetheless, obtained results open up a new research venue: to try to attest the reasons for some linguistic contexts being more conducive to retaining the acoustic features of the L1. It would also be interesting to see whether the retention of the acoustic features of the dominant language in the weak language happens in bilinguals from birth.

7. References

- ACTFL (2012). *ACTFL Proficiency Guidelines*. Retrieved from <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>
- Beristain, A. (in preparation). *Small L1-L2 phonetic contrasts in bilingual societies: Sibilants in three Basque and Basque-Spanish varieties* (Unpublished M.A. thesis). University of Illinois, Urbana-Champaign.

- Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.21, retrieved 26 October 2016 from <http://www.praat.org>
- Donegan, P., & Stampe, D. (2009). Hypotheses on Natural Phonology. *Poznan Studies in Contemporary Linguistics*, 45(1), 1-31.
- Elejabeitia, A., & Bizcarrondo, G. (1992). *La S en el español de Vizcaya*. Bilbao: University of Deusto.
- Etxebarria, M. (1997). Bilingüismo y adquisición temprana del lenguaje: Procesos fonológicos en el contacto vasco/español. In *First International Symposium on Bilingualism: Bilingual Communities and Individuals* (pp. 237-265). Vigo: University of Vigo.
- Flège, J. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-381). New York: Mouton de Gruyter.
- Gaminde, A., Aurrekoetxea, G., Etxebarria, A., Garay, U., & Romero, A. (2016). *Ahoskera lantzeko argibideak eta jarduerak. Laguntzarako materiala: teoria eta praktika*. Bilbao: University of the Basque Country.
- Hualde, J.I. (2003). Segmental phonology. In J.I. Hualde & J. Ortiz de Urbina, *A Grammar of Basque* (pp. 15-31). Berlin: Mouton de Gruyter.
- Hualde, J.I. (2010). Neutralización de sibilantes vascas y seseo en castellano. *Oihenart*, 25, 89-116.
- Hualde, J.I. (2014). *Los sonidos del español*. Cambridge: Cambridge University Press.
- Iribar, A., & Isasi, C. (2008). El seseo vasco: nuevos datos para una vieja cuestión. *Oihenart*, 23, 241-258.
- Iribar, A., Isasi, C., Gómez Seibane, S., & Moral del Hoyo, C. (2005). Notas para la descripción acústica del seseo vizcaíno. In M. González, E. Fernández, & B. González (Eds.), *III Congreso Internacional de Fonética Experimental* (pp. 389-398). Santiago de Compostela: Xunta de Galicia.
- Isasi, C., Iribar, A., & Moral del Hoyo, C. (2009). Una transferencia vasca: el seseo de hablantes vizcaínos y guipuzcoanos. *Oihenart*, 24, 201-235.
- Jurado, M. (2011). Caracterización de sibilantes fricativas vascas y su percepción en el sistema fonético español. *Anuario del Seminario de Filología Vasca Julio de Urquijo*, 45(1), 81-137.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustic Society of America*, 108(3), 1252-1263.
- Larrazá, S. (2015). *Acquisition of Phonology and Spanish-Basque Bilinguals' Phonological Systems*. Vitoria: University of the Basque Country.
- MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition of English among children in a bilingual education program. *Bilingual Research Journal*, 29(3), 687-712.
- Medina del Moral, N., & Romera, M. (2016). Análisis acústico del seseo vasco en los datos del Archivo del Patrimonio Inmaterial de Navarra. *Huarte de San Juan: Filología y Didáctica de la Lengua*, 16, 34-51.
- Oñederra, L., Elordui, A., Epelde, I., Etxebarria, P., Jauregi, O., & Salaberria, J. (2015). Euskaltzaindiaren Ahoskera batzordearen txostena (Ahoskerak axola du), *Euskera*, 60(2), 499-531.
- Oñederra, M.L., Jauregi, O., & Epelde, I. (2014). Hiztunaren fonema inbentarioa eta elebitasun goiztiarra: euskara gehi frantsesa vs. euskara gehi gaztelania. *Lapurdum*, 18, 75-95.
- Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A Limit on Behavioral Plasticity in Speech Perception. *Cognition*, 64, 9-17.

- Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer program]. Version 1.0.136, retrieved 26 October 2016 from <https://www.R-project.org>
- Rodríguez-Ordóñez, I. (2016). *Differential Object Marking in Basque: Grammaticalization, Attitudes and Ideological Representations* (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.
- Urrutia, H., Etxebarria, M., Turrez, I., & Duque, J.C. (1988). *Fonética Vasca I. Las Sibilantes en el Vizcaíno*. Bilbao: University of Deusto.
- Urrutia, H., Etxebarria, M., Turrez, I., & Duque, J.C. (1989). *Fonética Vasca II. Las Sibilantes en el Guipuzcoano*. Bilbao: University of Deusto.
- Urrutia, H., Etxebarria, M., Turrez, I., & Duque, J.C. (1991). *Fonética Vasca III. Las Sibilantes en los dialectos Orientales*. Bilbao: University of Deusto.
- White, L. (2008). *Aurrera! A Textbook for Studying Basque* (Vol. 1). Reno: University of Nevada Press.

IGORREKO AZENTUAREN KORRELATU AKUSTIKOEZ⁷

Aitor Iglesias, Naia Eguskiza, Iñaki Gaminde eta Lorea Unamuno
Euskal Herriko Unibertsitatea

aitor.iglesias@ehu.eus, naia.eguskiza@ehu.eus, inaki.gaminde@ehu.eus,
lorea.unamuno@ehu.eus

Laburpena

Lan honen helburu nagusia Igorreko azentuaren korrelatu akustikoak aztertzea da, Igorreko azentuaren izari akustikoetan zein den garrantzirik handiena daukana zehazteko. Lana burutu ahal izateko 20 berbak osatutako grabazioak jaso dira. Sei izan dira informatzaileak: 1990-1992 urteen artean jaiotako hiru neska eta hiru mutil igorreztar. Inkesta 10 berbaz egon da osatua: 5 silaba bikoak izan dira, lehenengo silaba izanik azentuduna eta beste 5 hiru silabakoak, azentua erdian daukatenak. Denetara, 1.800 datu jaso eta aztertu dira Praaten bidez. Bokal bakoitzeko f_0 , iraupena, energia eta bokalen f_1 , f_2 eta f_3 aztertu dira. Azterketak egiteko kontuan hartu da: bokala azentuduna edo azentu bakoa den, berba silaba bikoak edo hirukoak den, azentu bakoetan azentuaren aurrekoa (aa) edo ostekoa den (ao) eta informatzaileen sexua. Azken sailkapena, multzokatze-azterketa eta dimentsio aniztunen eskala erabilia egin da.

Keywords: hizkuntza-bariazioa, azentua, euskara, Igorreko barietatea

1. Sarrera

Lan hau egiteko ideia Hualderen (2017) Azpeitiko azentu-sistemaren korrelatu akustikoez egindako artikulua bat irakurtzean sortu da. Hasieran berak bere lanean esaten dituenak Igorren berdin betetzen diren edo alderik ote dagoen ikusteko jakingurak bultzatuta ekin genion gure lantxo honi. Igorren erabiltzen den hiztegia eta gramatika ezaugarriak gogotan, lehenengo eta behin aldaketa metodologiko batzuk egitea pentsatu genuen eta abiatu ginen.

Hualdek bere lanean esaten duen moduan, alde handia dago euskal hizkeren artean azentu-sistemei dagokienez eta gerta daiteke korrelatu akustikoetan ere aldeak egotea. Hizkera batzuetan iraupena gailentzen da (Hualde, Lujanbio eta Torreira 2008) eta beste batzuetan f_0 (Gaminde, 1992, 1993, Elordieta eta Hualde, 2003) esate baterako. Bestalde, ez dira asko azentu sistemen korrelatuez datu akustikoak batu dituzten lanak eta honek ere adore eman digu hemengo datuak jasotzeko eta aztertzeko.

Gure lan honen helburua izan da jakitea Igorreko azentuaren izari akustikoetan zein den garrantzirik handiena daukana. Bide batez, inkestak burutzeko aukeratu ditugun hitz guztien bokalak “a” izan direnez gero, jakin gura izan dugu azentuak zenbatean eragiten dien bokalen formakinei.

Lana bost atal nagusitan banatuta aurkezten dugu. Sarrera honen osteko bigarren atalean Igorreko azentu-sistemaren ezaugarri nagusiak aztertzen dira. Hirugarren atalean erabili dugun corpusaren ezaugarriak eta berau jasotzeko metodologia aurkezten dira. Laugarren atalean datuen azterketa ematen da. Azkenik, bosgarren atalean ondorio nagusiak laburbiltzen dira.

⁷ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

2. Igorreko azentuaren ezaugarriak

Hualdek (1997) eta Gamindek (1998) egindako sailkapenetan euskarak lau azentu-mota nagusi ditu. Sailkapen horien arabera Igorreko azentua mendebaldeko zenbait sistema bereziren artean kokatu behar litzateke, zehazki Arratia eta Zeberio azpisailkapenean eta horrela berresten du Iglesiasek (2008 & 2014) bere lanetan emandako datuetan.

Igorreko azentua, sailkapen honetakoak legez, ez-tonala da eta azentua bereizgarria da; izan ere, berbak ez dira berdin azentuatzen. Azentu-sistema honetan hitz mugatu guztiak azentudunak dira, flexiozko morfemek azentuaren muga markatzen dute eta azentua ez da inoiz mugatzailetik eskuinera pasatzen. Azentua ez da finkoa; silaben kokapena eta norabidea kontatzeko kontuan hartzen dituzten erregela orokorren bitartez asignatzen da azentua.

Erregela nagusiak dio Igorreko azentuaren kokalekua 1] dela, baina erregela honek hainbat salbuespen ditu (Iglesias, 2008 & 2014).

Azentuaren kokagunea azaltzeko, bereizi behar dira silaba bat eta biko hitz-erroak hiru silaba edo gehiago dituzten hitz erroez. Hiru silabakoetan singularrean azentua hitz-erroaren azken silaban txertatzen da: 1].

alabéa
alabéari
alabéan
alabéagas

Silaba bateko eta biko hitz-erroetako berbetan ere 1] da, baina horretarako azaldu behar dugu talde prosodikoaren azken silabaren estrametrikalitatea. Hizkera honetan talde prosodikoaren azken silaba estrametrikoa da. Hitzaren azentua azkenaurreko silaban kokatzen da hitza talde prosodikoaren bukaeran dagoenean, baina, ezpabere, azken silaban kokatzen da.

<i>mutíle</i>	baina	<i>mutilé da, mutíl bet...</i>
<i>lúrre</i>	baina	<i>lurréri, lurréko...</i>

Beraz, azentu-unitate barruan sartzen badugu, artikulua edo bokal epentetikoak azentuaren kokagunea 1] da silaba bateko eta biko hitz-erroetako berbetan.

Mugagabeen ere azentuaren kokalekua 1] da eta honetan ere kontuan hartu behar dugu talde prosodikoaren azken silaba estrametrikoa dela.

<i>lau abáde</i>	baina	<i>abadé bat...</i>
------------------	-------	---------------------

Berba plurala denean, azentua silaba bat ezkerrerago txertatzen da.

<i>abadéa</i>	<i>abádeak</i>
<i>abadéari</i>	<i>abádeari</i>
<i>abadéan</i>	<i>abádean</i>

Baina arau honek salbuespen bat du, zeren eta -a-z amaitutako berbetan silaba bat barik bi ezkerrerago txertatzen baita.

<i>alabéa</i>	<i>álabak</i>
<i>alabéari</i>	<i>álabari</i>
<i>alabéan</i>	<i>álaban</i>

Azentu-sistema honetan salbuespen hau ez eze beste salbuespen batzuk ere badaude.

Lehenengoa salbuespena atzizki eta oinarri markatuak dira. Badira berba batzuk azentu finkoa dutenak, bai singularrean bai pluralean eta hauetan kokagunea ez da aldatzen. Iglasiasek (2014: 121) zerrenda luzea ematen du.

<i>olláskoá</i>	<i>olláskoak</i>
<i>olláskoari</i>	<i>olláskoari</i>
<i>olláskoan</i>	<i>olláskoan...</i>

<i>belárrie</i>	<i>belárriek</i>
<i>belárriari</i>	<i>belárriari</i>
<i>belárrien</i>	<i>belárrien...</i>

Bestalde, atzizki batzuk aurreazentuatzaileak dira eta deklinabide-kasu guztietan azentua aurreko silaban ezartzen dute. Hauek dira hizkera honetako atzizki markatuak: *-garren*, *-tar*, *-en*, *-txu*, *-xe*, *-txe*, *-ago* eta *-egi*.

irúgarren, *txikítxue*, *dimóstarra*, *ónena*, *andíegie*, *luséago*, *onéxeri...*

-dun atzizkia ere markatua da, baina honek, aurrekoek ez bezala, azentua bereganatu egiten du.

bixerdúne, *euskaldúne*, *erdaldúne...*

Salbuespenak aipatzen amaitzeko, esan behar da hainbat deklinabide-atzizki ere markatuak direla eta hauek ere aurreazentuatzaileak direla. Hau gertatzen da ablatibo, inesibo eta leku genitibo kasuetan.

errítik, *erríen*, *erríkoa...*

Beraz, laburbilduz, esan dezakegu Igorreko azentua txertatzeko arauak honako hauek direla:

- 1) Azentua singularrean eta mugagabeen 1] da
- 2) Pluralean azentua silaba bat ezkerrerago txertatzen da
- 3) Bi arau hauek salbuespen batzuk dituzte:
 - Pluralean -a-z amaitutako berbak
 - Atzizki eta oinarri markatuak

3. Corpora eta metodologia

Gure lan honen helburua Igorreko azentuaren korrelatu akustikoak aztertzea denez gero, inkestak egiteko itemak helburu hori gogoan aukeratu dira eta kontrolatu nahi izan dira aldagai batzuk. Denetara hamar hitz aukeratu dira eta “_____diñot” ingurunean, hau da, aztertutako hitza galdegai gunean eta esaldiaren hasieran, kokatuta egon dira kasu guztietan. Hamar hitz horietako bost silaba bikoak izan dira eta lehen silaba izan da azentuduna: “*txárra*”, “*látza*”, “*sárra*” (zaharra), “*mátza*” (mahatsa) eta “*gátza*”. Beste bostak hiru silabakoak izan dira eta azentua erdian daukate: “*asála*”, “*sabála*”, “*adárra*”, “*sagárra*” eta “*papárra*” dira aukeratutakoak. Informatzaile bakoitzak berba guztiak birritan esan zituen; beraz, informatzaile bakoitzeko 20 berba jaso dira aztertzeko.

Denetara 300 bokal aztertu dira; 120 azentudunak eta 180 azentu bakoak. Azentu bakoetan 120 silaba azentudunaren ostekoak izan dira (ao) eta 60 silaba azentudunaren aurrekoak (aa). Informatzaileak sei izan dira, hiru neska eta hiru mutil. Seiak Igorrekoak dira eta 1990-1992 urte artean jaiokoak.

Grabazioak egiteko Samsung Galaxy J7 smartphonea erabili da Boya BY-M1 mikrofonoarekin eta Smart Recorder aplikazioa erabili da .wav formatoan grabatu ahal izateko.

Behin grabazio guztiak eginda, datuak “praat” programaren bidez (Boersma eta Weenink, 2016) aztertu dira. Izari guztien datuak automatikoki jaso dira script baten bidez. Bokal bakoitzeko f0 (hz.etan), iraupena (ms.etan), energia (db.etan) eta bokalen f1, f2 eta f3 (hz.etan) jaso dira. Denetara 1800 datu jaso eta aztertu dira. Datuak orokorrean aztertzeko erabili diren aldagaiak lau izan dira:

- (1) Bokala azentuduna edo azentu bakoia izatea.
- (2) Berbaren silaba kopurua, 2 edo 3 izatea.
- (3) Azentu bakoetan azentuaren aurrekoa izatea (aa) edo ostekoa izatea (ao).
- (4) Informatzaileen sexua.

Datuak izaririk izari aztertu dira emaitza orokorretatik abiatuta; gero aldagai bakoitzaren balizko eragina ikertu ahal izateko. Era honetara ikusi ahal izan dugu zeintzuk diren eragin estatistikoa daukaten aldagaiak, berori zenbatekoa den eta zein izariri eragiten dioten.

4. Datuen azterketa

Atal hau azpiatal bitan banatuta aurkezten dugu; lehen azpiatalean azentuaren korrelatu akustiko nagusien izarien emaitzak ematen ditugu eta bigarreanean “a” bokalaren lehen hiru formakinen azterketa aurkezten dugu.

4.1. Azentuaren korrelatuen izari akustikoak

Azpiatal honetan silaba azentudun eta azentu bakoien oinarrizko maiztasuna, iraupena eta energia aztertuko ditugu.

Oinarrizko maiztasuna:

Silaba biko berbetan oinarrizko maiztasunari (\bar{x}) dagozkion batezbestekoak eta desbideratzeak (sd) 1. taulan aurkezten dira informatzaile bakoitzeko. Informatzaile guztiek ez dute berdin egiten; horrela bada M1 eta N1 informatzaileek f0ren batezbestekorik altuena azentu osteko silaban daukate eta gainerakoek, ostera, azentudun silaban.

	+ Azentua		-Azentua	
	\bar{x}	sd	\bar{x}	sd
M1	130,93	2,48	145,75	5,32
M2	130,56	3,61	118,64	3,71
M3	125,42	3,97	110,84	4,95
N1	240,26	14,40	260,15	27,13
N2	214,79	11,10	185,50	9,43
N3	238,16	6,00	226,89	11,99

1. taula: Silaba biko berba azentudunen eta azentu bakoien f0ren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

2. taulan silaba bien arteko batezbestekoen aldeekin batera, Mann-en U probaren emaitzak ematen ditugu, probaren z koefizientearekin batera probabilitatea (p) eta heinaren batezbestekoak silaba azentudunean (+Azentua) eta azentu bakoan (-Azentua). 2. taulan, silaben arteko f0ren batezbestekoen aldeak estatistikoki esanguratsuak direla, informatzaile guztien kasuan, ikusten da.

Informatzailea	Aldea	z	p	+Azentua	-Azentua
M1	-14,82	-3,785	0,000	5,50	15,5
M2	11,92	-3,717	0,000	15,40	5,60
M3	14,58	-3,747	0,000	15,45	5,55
N1	-19,89	-2,501	0,012	7,20	13,80
N2	29,29	-3,554	0,000	15,20	5,80
N3	11,27	-2,271	0,023	13,50	7,50

2. taula: Silaba azentudunen eta azentu bakoien f0ren aldea eta Mannen Uren emaitzak informatzaile bakoitzeko.

Hiru silabako berbetan lortu ditugun batezbestekoak eta desbideratzeak 3. taulan erakusten dira; azentu aurrekoa (aa), azentuduna eta azentuaren ostekoa (ao) bereiz aurkezten dira. Silaba bikoetan bezala, hemen ere M1 eta N1 informatzaileek f0ren batezbestekorik altuena azentuaren osteko silaban daukate eta gainerakoek silaba azentudunean bertan.

	aa		+ Azentua		ao	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
M1	105,81	3,31	134,12	2,84	144,38	5,49
M2	112,92	5,03	130,57	2,73	115,96	4,34
M3	109,52	2,97	125,04	5,39	116,57	6,68
N1	209,88	7,13	235,82	10,11	265,83	9,85
N2	184,78	5,19	213,62	8,48	180,22	8,46
N3	212,93	7,19	249,56	4,32	216,17	8,85

3. taula: Hiru silabako berba azentudunen eta azentu bakoien f0ren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

4. taulan silaba azentudunen eta azentu bakoien batezbestekoen aldeak estatistikoki esanguratsuak diren ala ez ikusteko Mann-en U probaren emaitzak ikus daitezke.

Oraingo honetan M1 eta N1 informatzaileen kasuan aldeak ez dira estatistikoki esanguratsuak eta gainerakoen kasuan, berriz, bai.

Informatzailea	z	p	+Azentua	-Azentua
M1	-0,088	*0,930	15,70	15,40
M2	-4,408	0,000	25,50	10,50
M3	-3,683	0,000	23,85	11,33
N1	-0,088	*0,930	15,70	15,40
N2	-4,403	0,000	25,50	10,50
N3	-4,403	0,000	25,50	10,50

4. taula: Mannen Uren emaitzak informatzaile bakoitzeko silaba azentudunen eta azentu bakoien batezbestekoen aldean esangura estatistikoari dagokionez.

Azentu ostekoak f_0 altuagoa dauka azentu aurrekoak baino kasu guztietan N2 informatzailearenean izan ezik; aldeak estatistikoki esanguratsuak dira M1, M3 eta N1 informatzaileen emaitzekin.

Iraupena:

5. taulan silaba biko berben iraupenaren batezbestekoak eta desbideratzeak ematen ditugu; M2ren kasuan izan ezik gainerako guztietan silaba azentuduna luzeagoa da azentu bakoa baino.

	+ Azentua		-Azentua	
	\bar{x}	sd	\bar{x}	sd
M1	112	10,33	105	12,69
M2	91	11,97	100	11,55
M3	85	11,79	68	9,19
N1	117	9,49	101	9,94
N2	104	10,75	91	5,68
N3	120	18,86	108	20,44

5. taula: Silaba biko berba azentudunen eta azentu bakoien iraupenaren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

Hala ere, 6. taulan agertzen den moduan, Mannen U probaren arabera aldea estatistikoki esanguratsua da M3, N1 eta N2 informatzaileen kasuan eta gainerakoen kasuetan ez.

Informatzailea	Aldea	z	p	+Azentua	-Azentua
M1	7	-1,099	*0,272	11,9	9,10
M2	-9	-1,700	*0,089	8,35	12,65
M3	17	-2,904	0,004	14,20	6,80
N1	16	-2,868	0,004	14,20	6,80
N2	13	-2,817	0,005	14	7,00
N3	12	-1,036	*0,300	11,85	9,15

6. taula: Silaba azentudunen eta azentu bakoien iraupenaren aldea eta Mannen Uren emaitzak informatzaile bakoitzeko.

Hiru silabako berbetan lortu ditugun batezbestekoak eta desbideratzeak 7. taulan ematen dira; azentu aurrekoa (aa), azentuduna eta azentuaren ostekoa (ao) bereiz aurkezten dira. Ikusten den bezala, silabarik luzeena azentuaren aurrekoa da M1, M3, N1 eta N2 informatzaileen emaitzetan, azentuduna N3 informatzailearen kasuan eta azentu ostekoa M2ren kasuan.

	aa		+ Azentua		ao	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
M1	104	6,99	100	14,14	94	8,43
M2	89	17,29	87	15,67	101	16,63
M3	84	13,50	81	15,95	64	6,99
N1	127	19,47	98	9,19	103	13,37
N2	100	10,54	98	13,17	91	11,97
N3	111	14,49	115	10,80	93	9,49

7. taula: Hiru silabako berba azentudunen eta azentu bakoien iraupenaren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

Silaba azentudunen eta azentu bakoien batezbestekoen aldeak txikiak dira eta N1 eta N3 informatzaileen kasuetan baino ez dira estatistikoki esanguratsuak (8. taula).

Informatzailea	z	p	+Azentua	-Azentua
M1	-0,046	*0,964	15,60	15,45
M2	-1,345	*0,179	12,50	17,00
M3	-1,183	*0,237	18,10	14,20
N1	-2,323	0,020	10,35	18,08
N2	-0,204	*0,838	15,95	15,28
N3	-2,609	0,009	21,30	12,60

8. taula: Mannen Uren emaitzak informatzaile bakoitzeko silaba azentudunen eta azentu bakoien batezbestekoen aldean esangura estatistikoari dagokionez.

Silaba azentu bakoetan azentu aurrekoak iraupen luzeagoa dauka ostekoak baino M2 informatzailearen kasuan izan ezik; aldeak estatistikoki esanguratsuak dira M1, M3, N1 eta N3 informatzaileen emaitzekin.

Energia:

9. taulan silaba biko berben energiaren batezbestekoak eta desbideratzeak ematen ditugu; M1en kasuan izan ezik gainerako guztietan silaba azentudunak energia altuagoa dauka azentu bakoak baino.

	+ Azentua		-Azentua	
	\bar{x}	sd	\bar{x}	sd
M1	82,64	0,45	83,31	0,44
M2	83,62	0,72	83,14	0,70
M3	82,63	0,50	81,02	1,16
N1	83,82	0,46	83,75	0,72
N2	83,98	0,81	82,31	0,65
N3	79,75	0,63	77,50	1,92

9. taula: Silaba biko berba azentudunen eta azentu bakoien energiaren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

10. taulan batezbestekoen aldeekin batera Mannen U probaren emaitzak ere ematen ditugu; batezbestekoen aldeak estatistikoki esanguratsuak dira informatzaile guztietan M2 eta N1 informatzaileetan izan ezik.

Informatzailea	Aldea	z	p	+Azentua	-Azentua
M1	-0,67	-2,873	0,004	6,70	14,30
M2	0,48	-1,932	*0,053	12,85	8,15
M3	1,61	-2,948	0,003	14,40	6,60
N1	0,07	-0,189	*0,850	10,75	10,25
N2	1,67	-3,402	0,001	15,00	6,00
N3	2,25	-2,343	0,019	13,60	7,40

10. taula: Silaba azentudunen eta azentu bakoien energiaren aldea eta Mannen Uren emaitzak informatzaile bakoitzeko.

Hiru silabako berbetan lortu ditugun batezbestekoak eta desbideratzeak 11. taulan ematen dira; azentu aurrekoa (aa), azentuduna eta azentuaren ostekoa (ao) bereiz aurkezten dira. Oraingo honetan batezbestekorik altuena daukan silaba azentuduna da N1 informatzailearen emaitzetan izan ezik.

	aa		+ Azentua		ao	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
M1	80,87	1,03	82,35	0,49	83,23	0,57
M2	81,85	1,38	84,26	0,48	82,73	0,83
M3	80,73	1,55	82,49	0,77	81,65	0,75
N1	82,19	0,80	83,62	0,41	84,02	0,43
N2	81,91	1,37	84,89	0,65	83,06	1,03
N3	75,99	1,41	79,85	1,10	77,44	1,28

11. taula: Hiru silabako berba azentudunen eta azentu bakoien energiaren batezbestekoak eta desbideratzeak informatzaile bakoitzeko.

Silaba azentudunaren eta azentu bakoien arteko aldeak estatistikoki esanguratsuak dira M1 eta N1 informatzaileen kasuetan izan ezik (12. taula).

Informatzailea	z	p	+Azentua	-Azentua
M1	-0.176	*0,860	15,90	15,30
M2	-3,739	0,000	23,80	11,35
M3	-2,794	0,005	21,85	12,33
N1	-0.880	*0,379	17,50	14,50
N2	-4,092	0,000	24,80	10,85
N3	-4,003	0,000	24,60	10,95

12. taula: Mannen Uren emaitzak informatzaile bakoitzeko silaba azentudunen eta azentu bakoien batezbestekoen aldean esangura estatistikoari dagokionez.

Azentu bako silabetan azentu ostekoak energia altuagoa dauka azentu aurrekoak baino, aldea estatistikoki esanguratsua da M1, N1 eta N3 informatzaileen kasuetan.

Silaba biko berben emaitzekin azentuduna izateari edo ez izateari eragiten dioten hiru izarien eragina eta pisua aztertzeko erregresio logistikoa deritzon estatistikoa erabili dugu. 13. taulan izari bakoitzaren eraginaren esanguraren p-ren emaitza ematen dugu (kontuan har bedi estatistikoki esanguratsua izateko p 0,05 baino txikiagoa izan behar dela). Taularen datuei erreparatuta ikusten dugu f_0 estatistikoki esanguratsua dela bost informatzailearen emaitzetan, energia lau informatzailearen kasuetan eta iraupena hiru kasutan. Beraz, esan dezakegu azentua bereiztean eraginik handiena f_0 k daukala, gero energiak eta azkenik iraupenak. Bestalde, f_0 rekin batera beste izariren bat estatistikoki esanguratsua denean ere, beronen p txikiagoa da. Kasu bakar batean (N1) da iraupena izari esanguratsua.

Informatzailea	Iraupena	F0	Energia
M1	*0,174	0,000	0,005
M2	*0,094	0,000	*0,068
M3	0,004	0,000	0,002
N1	0,003	*0,051	*0,764
N2	0,005	0,000	0,000
N3	*0,171	0,017	0,004

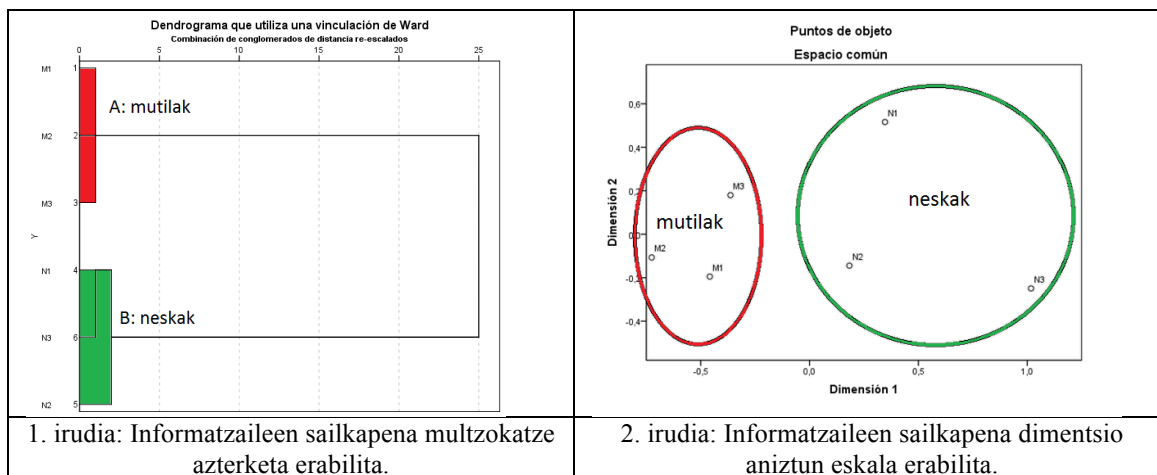
13. taula: Silaba biko izari bakoitzaren esangura estatistikoaren p-ren emaitza informatzaile bakoitzeko.

Hiru silabako berben emaitzak 14. taulan erakusten dira; f0 estatistikoki esanguratsua da lau informatzailearen emaitzetan, energia beste lau informatzailearen kasuetan eta iraupena kasu bitan. Kasu bakar batean (N1) iraupena esanguratsua den izari bakarra da eta M1 informatzailearen kasuan izari batek ere ez dauka eragin esanguratsurik.

Informatzailea	Iraupena	F0	Energia
M1	*0,808	*0,164	*0,518
M2	*0,221	0,000	0,000
M3	*0,226	0,000	0,008
N1	0,019	*0,828	*0,161
N2	*0,590	0,000	0,000
N3	0,023	0,000	0,000

14. taula: Hiru silabako izari bakoitzaren esangura estatistikoaren p-ren emaitza informatzaile bakoitzeko.

Informatzaileek erabiltzen dituzten izari akustikoen batezbestekoak erabilia sailkatzen ahalegintzen garenean dela multzokatze azterketa eginda (1. irudia), dela dimentsio aniztun eskala (2. irudia) eginda, multzo nagusi bi agertzen zaizkigu non eta informatzaileen sexua erabatekoa den.

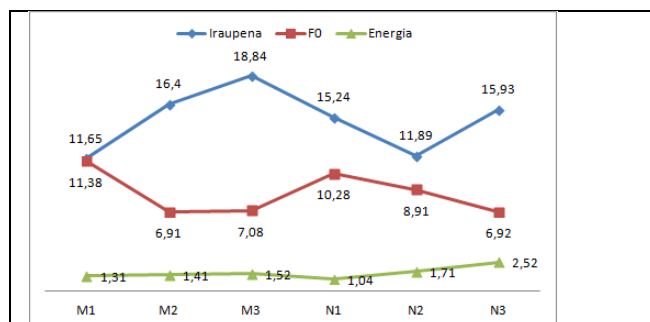


Atal hau mukurutzeko izarien aldakuntza koefizienteak eta izarien arteko korrelazioak emango ditugu informatzaile bakoitzeko.

15. taulan informatzaileen izari bakoitzaren aldakuntza koefizienteak ikus daitezke. Kasu guztietan koefizienterik altuena iraupenak dauka, gero F0k eta azkenik energiak (3. irudia).

Informatzailea	Iraupena	F0	Energia
M1	11,65	11,38	1,31
M2	16,40	6,91	1,41
M3	18,84	7,08	1,52
N1	15,24	10,28	1,04
N2	11,89	8,91	1,71
N3	15,93	6,92	2,52

15. taula: Izari bakoitzaren aldakuntza koefizienteak informatzaile bakoitzeko.



3. irudia: Izarien aldakuntza koefizienteak informatzaile bakoitzeko

16. taulan izarien arteko korrelazioak ematen ditugu informatzaile bakoitzeko eta baita korrelazio esanguratsuak ere. F0ren eta energiaren arteko korrelazio positiboa ikus daiteke; berau esanguratsua da kasu guztietan eta M1 eta M2 informatzaileen kasuan nahikoa korrelazio handia agertzen da. F0ren eta iraupenaren arteko korrelazioak esanguratsuak dira N2 eta N3 informatzaileen datuekin, hala ere, korrelazioak nahikoa txikiak dira. Azkenik, iraupenaren eta energiaren arteko korrelazioak ez dira esanguratsuak.

	f0_iraupena		f0_Energia		Iraupena_Energia	
	Korre	p	Korre	p	Korre	p
M1	0,171	*0,135	0,822	0,000	0,183	*0,110
M2	-0,149	*0,301	0,831	0,000	-0,075	*0,603
M3	0,101	*0,377	0,546	0,000	0,079	*0,494
N1	-0,008	*0,947	0,651	0,000	0,086	*0,455
N2	0,237	0,037	0,634	0,000	0,041	*0,232
N3	0,341	0,002	0,666	0,000	0,249	*0,028

16. taula: Izarien arteko korrelazioak eta euren esangura estatistikoa informatzaile bakoitzeko.

4.2. Bokalaren formakinak

Bokalen formakinak orokorrean aztertuta lortzen ditugun batezbestekoak honako hauek dira: f1 716,18 Hz. (sd.: 115,38), f2 1492, 63 (sd.:138,24) eta f3 2614,28 (sd.: 258,25).

Silaba azentudunen eta azentu bakoien arabeko emaitzak 17. taulan erakusten direnak dira; aldeak estatistikoki esanguratsuak dira flaren kasuan (T = (a.m.: 298) - 3,418; p = 0,001), gainerako formakinen kasuetan, oster, aldeak ez dira estatistikoki esanguratsuak.

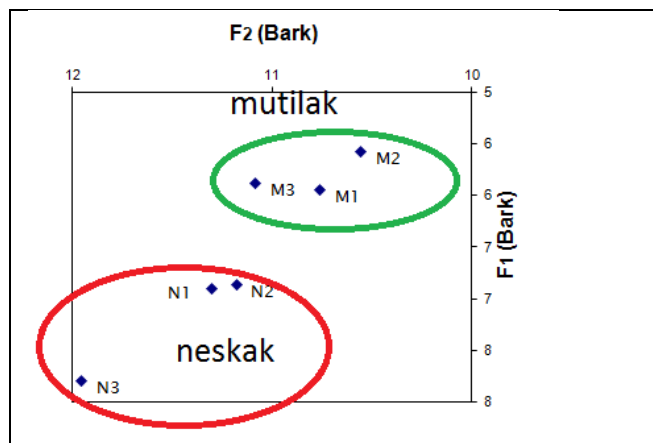
	f1		f2		f3	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
+ Azentua	743,58	128,92	1503,77	142,65	2601,42	233,64
- Azentua	697,91	101,73	1485,21	135,11	2622,86	273,73

17. taula: "a" bokalaren formakinen batezbestekoak eta desbideratzeak silaba azentudunen eta azentu bakoen arabera.

Informatzaile bakoitzaren emaitza orokorrak 18. taulan erakusten direnak dira. Lehen bi formakinen arabera dispersioak aztertuta (4. irudia), ikusten dugu nesken bokalak irekiagoak eta aurreratuagoak direla mutilenak baino.

	f1		f2		f3	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
M1	643,46	40,42	1407,34	109,84	2579,30	105,95
M2	598,34	29,49	1366,12	78,34	2519,52	176,93
M3	636,22	40,16	1477,52	70,35	2788,94	106,56
N1	766,18	61,78	1525,92	134,02	2685,84	305,49
N2	761,50	48,30	1496,88	82,08	2303,08	256,30
N3	891,38	94,39	1682,02	83,84	2809,02	98,90

18. taula: "a" formakinen batezbestekoak eta desbideratzea informatzaile bakoitzeko.

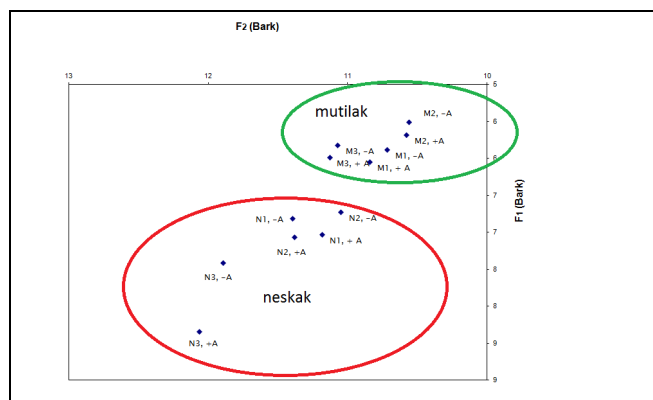


4. irudia: "a" bokalaren lehen bi formakinen arabera dispersioa.

Azentudunen eta azentu bakoen emaitzak bereiz aztertzen baditugu, 19. taulako datuak lortzen dira. Azentudunak kasu guztietan irekiagoak dira azentu bakoak baino eta aurreratuagoak N1 informatzailearen kasuan izan ezik (5. irudia).

		f1		f2		f3	
		\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
M1	+ Azentua	655,70	38,80	1422,80	116,52	2599,85	114,51
	- Azentua	635,30	40,03	1397,03	105,91	2565,60	99,46
M2	+ Azentua	611,10	28,20	1368,75	75,24	2539,75	133,43
	- Azentua	589,83	27,60	1364,37	81,56	2506,03	201,90
M3	+ Azentua	648,75	22,23	1484,95	63,75	2754,20	110,07
	- Azentua	627,87	47,12	1472,57	75,07	2812,10	99,31
N1	+ Azentua	783,75	53,37	1496,85	138,01	2504,20	332,97
	- Azentua	754,47	65,03	1545,30	129,99	2806,93	217,08
N2	+ Azentua	789,10	29,03	1541,60	71,22	2391,35	213,35
	- Azentua	743,10	50,14	1467,07	75,90	2244,23	268,72
N3	+ Azentua	973,10	47,06	1707,65	90,85	2819,15	87,07
	- Azentua	836,90	76,99	1664,93	75,60	2802,27	106,96

19. taula: "a" formakinen batezbestekoak eta desbideratzea informatzaile bakoitzeko silaba azentudunen eta azentu bakoen arabera.



5. irudia: “a” bokalaren lehen bi formakinen araberako dispersioa, silaba azentudunetan (+A) eta azentu bakoetan (-A).

Bokalen dispersioan azentuduna edo azentu bakoia izatea baino garrantzitsuagoa da informatzaileen sexua. 20. taulan formakin bakoitzaren batezbestekoak eta desbideratzeak ematen ditugu. Bertan agertzen diren aldeak estatistikoki esanguratsuak dira f1-i ($T = (a.m.: 298) 21,719; p = 0,000$) eta f2-ri ($T = (a.m.: 298) 11,311; p = 0,000$) dagokienez.

	f1		f2		f3	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
Neskak	806,35	92,71	1568,27	130,72	2599,31	319,73
Mutilak	626,01	41,81	1416,99	98,70	2629,25	176,51

20. taula: “a” bokalaren formakinen batezbestekoak eta desbideratzeak informatzaileen sexuaren arabera.

Azentu- eta sexu-aldagaiak ANOVA faktorialaren bidez batera aztertzen ditugunean, agerian jar dezakegu aldagai bakoitzaren eragina; era berean, formakin bakoitzaren azterketan gainerako formakinen eragina kontrolatzen dugunean aldagai bakoitzaren eta bien eragina zehatzago aztertzea lortzen da. Jarraian formakin bakoitzaren emaitzak aurkeztuko ditugu, koaldagaiak kontuan hartu barik eta kontuan hartuta.

F1i dagokionez, azentuduna izatea edo ez izatea estatistikoki esanguratsua da ($F = (a.m.: 1) 33,015; p = 0,000$) eta sexua banan ere bai ($F = (a.m.: 1) 543,510; p = 0,000$); era berean, biak batera aztertuta, estatistikoki esanguratsuak dira ($F = (a.m.: 1) 9,752; p = 0,002$). Bien artean bariantzaren % 65,8 azaltzen da (R^2 zuzendua = 0,658). Azterketan f2 eta f3 koaldagai gisa sartzten ditugunean euren eragina kontrolatzeko orduan bariantzaren % 73,4 azaltzen da (R^2 zuzendua = 0,734).

F2ren kasuan, azentuduna izatea edo ez izatea ez da estatistikoki esanguratsua. Informatzaileen sexua, oster, bai ($F = (a.m.: 1) 124,234; p = 0,000$). Aldagai biak batera aztertzen direnean bien eragina ez da estatistikoki esanguratsua; bestalde bien artean bariantzaren % 29,8 baino ez da azaltzen (R^2 zuzendua = 0,298). Azterketan f1 eta f3 koaldagai gisa sartzten ditugunean euren eragina kontrolatzeko orduan bariantzaren % 53,2 azaltzen da (R^2 zuzendua = 0,532); baina gainerako emaitzak ez dira aldatzen.

Azkenik, f3ri dagokionez, ez azentuak ez informatzaileen sexuak ez daukate estatistikoki eragin esanguratsurik aldeak azaltzean. Hala ere, f1 eta f2 koaldagai gisa sartzten ditugunean, orduan emaitzak aldatzen dira eta azentua estatistikoki esanguratsua

da ($F = (a.m.: 1) 13,155$; $p = 0,000$) eta sexua ere bai ($F = (a.m.: 1) 97,169$; $p = 0,000$); bien eragina batera ere estatistikoki esanguratsua da ($F = (a.m.: 1) 5,494$; $p = 0,020$) eta bariantzaren % 38,9 azaltzen da (R^2 zuzendua = 0,389).

5. Ondorioak

Azken atal honetan datuen azterketaren bidez aurreko atalean lortu ditugun ondoriorik nabarmenenak laburbilduko ditugu. Lehenengo eta behin esan behar da azentuaren korrelatu akustikoen erabilerari dagokionez, informatzaile guztiek ez dutela jokabide bat bera erakusten. Nabarmenetako bat f0rik altuena azentu osteko silaban informatzaile bik edukitzea izan daiteke. Baina, aldi berean izarien gailentze maila eta esangura estatistikoa ere ez datoz bat informatzaile guztietan.

Ikusi dugun moduan, f0 da korrelatu nagusia azentudun silaba bereiztean, baina beronekin batera energiak ere garrantzia dauka; iraupena, berriz, eragin txikiena daukana da oro har. Bestalde f0ren eta energiaren arteko korrelazioa nahiko azkarra eta positiboa dela ikusi dugu. Alta, iraupena ez da korrelazionatzen besteekin hain azkarki.

Hiru silabako berba azentu bakoetan ikusi dugu silaba azentudunaren ostekoak f0 eta energia altuagoa daukala silaba azentudunaren aurrekoak baino. Iraupenari dagokionez, azentuaren aurrekoa luzeagoa da azentuaren ostekoa baino.

Informatzaileen sailkapena izari akustikoen emaitzen arabera egiten dugunean, informatzaileen sexuak eragin handia daukala ikusi dugu, izan ere, multzo bi egitean neskek bat osatzen dute eta mutilek bestea.

“a” bokalaren formakinak aztertzean ikusi dugun bezala, azentuduna izateak lehen formakinean baino ez dauka esangura estatistikoa. Hala ere, informatzaile guztiek ikusi dugun bezala azentudunak irekiagoak dira eta aurreratuxeago agertzen dira azentu bakoak baino. Hemen ere informatzaileen sexuak, arrunta den bezala (Fernández Planas, 2013), eragin handia dauka bokalen lehen bi formakinetan.

6. Bibliografia

- Boersma, P. & Weenink, D. (2009). *Praat: doing phonetics by computer, Version 5.1*. [Computer program]. <<http://www.praat.org>>
- Elordieta, G. & Hualde, J.I. (2003). Tonal and durational correlates of accent in context of downstep in Lekeitio Basque. *Journal of the International Phonetic Association*, 33, 195-209.
- Fernández Planas, A. M. (2013). Fonetica acustica y experimental. Las vocales. Las glides y las consonantes sonantes en la cadena hablada. In M.A. Penas (red.), *Panorama de la fonética española actual* (291-320 orr.). Madrid: Arco Libros.
- Gaminde, I. (1992). *Urduliz eta Gatikako Herri hizkeren Azterketa Linguistikoa*. Doktorego Tesia. Bilbo: Deustuko Unibertsitatea.
- Gaminde, I. (1993). Bermeoko azentu-ereduaz”, *Uztaro*, 8.
- Gaminde, I. (1998). *Euskaldunen azentuak*. Bilbo: Labayru Ikastegia.
- Hualde, J. I. & Beristain, A. (2017). Azpeitiko azentuaren gauzatze fonetikoaz. (Agertzeko FLVn)
- Hualde, J.I., Lujanbio, O., & Torreira, F. (2008). Lexical tone and stress in Goizueta Basque. *Journal of the International Phonetic Association*, 38(1).

- Hualde, J.I. (1997). *Euskararen azentuerak*. Bilbo: ASJUren gehigarriak 42.
- Iglesias, A. (2008). Igorreko azentuaz. *Euskalingua*, 13, 16-26.
- Iglesias, A. (2014). *Igorreko hizkeraren azterketa dialektologikoa*. Doktoretza-tesia, Gasteiz: EHU.

COMPRESS-EUS: I(RA)KASLEEN LABURPENAK LORTZEKO TRESNA

Unai Atutxa, Mikel Iruskietza, Olatz Ansa eta Alejandro Molina
Euskal Herriko Unibertsitatea eta Consejo Nacional de Ciencia y Tecnología
(CONACyT, México)

uatutxa002@ikasle.ehu.es, mikel.iruskietza@ehu.es, olatz.ansa@ehu.es,
amolina@conabio.gob.mx

Laburpena

Euskaldun zaharrak eta berriak barnebiltzen dituen ikerketa bat egin 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.

Testuak laburtzea hainbat arlotan da garrantzitsua, besteak beste hezkuntzan eta hizkuntzaren i(ra)kaskuntzan. Izan ere, laburtze gaitasuna hezkuntza prozesuan ikasleek garatu beharreko kompetentzia da. Eskola batean ikasle-taldearen laburtze gaitasuna garatu ahal izateko, ezinbestekoak dira baliabide didaktiko egokiak izatea eta ikasle-multzoaren laburpenak kudeatzeko tresna egokiak izatea. Alde batetik, irakasleak ikasle-taldearen zenbait azaleko datu modu automatikoan jasotzen baditu (laburpenen tamaina egokia den eta ideia nagusia mantendu duten, esaterako), bestelako eta sakonagoak diren trebetasun linguistikoak lantzeko: kohesioa eta koherentzia, adibidez. Lan honetan COMPRESS-EUS sistema aurkezten dugu, erabiltzaile-talde baten estrakziozko eta abstrakziozko laburpenak biltzeko tresna eta erabiltzaileek egindako laburpenaren azaleko ezaugarriak deskribatzeko erabil daitekeena, emaitzak kalkulu-orrian ematen dituelako.

Gako-hitzak: Laburpena, abstrakzioa, estrakzioa, COMPRESS-EUS, hezkuntza

1. Sarrera

Laburpena arlo askotan da garrantzitsua, batik-bat, hezkuntzan. Ikasleak informazio asko izan dezake gaur egun, egokia den informazioa, baina, horrek ez du bermatzen ikasleak informazioa ulertzen eta barneratzen duenik. Arazo horren aurrean laburpen gaitasuna garatu beharreko kompetentzia da, informazioa ulertzeko, interpretatzeko, erabiltzeko, berrantolatzeko eta egokiro idazteko gai izan dadin ikaslea. Horretaz gain, laburpena jarrera kritikoa eta testuaren ulermena ebaluatzeko tresna baliagarria izan daiteke, berari esker irakasleak jakin ahal izango baitu ea ikasleak curriculumeko eduki ezberdinak ulertu eta barneratu dituen.

Oinarrizko Hezkuntzaren Curriculumean, 236/2015eko Dekretuan, Oinarrizko zehar-kompetentziak (orokorrak) eta diziplina barneko/arteko oinarrizko kompetentziak (espezifikokoak) ageri dira. Oinarrizko zehar kompetentziei⁸ dagokienez, 5 kompetentzia ditugu, denak garrantzi berekoak, eta horien artean dugu “Hitzezko, hitzik gabeko komunikaziorako eta komunikazio digitalerako kompetentzia”. Diziplina barneko/arteko

⁸ Oinarrizko zehar kompetentziak hauek dira: 1) Hitzezko, hitzik gabeko komunikaziorako eta komunikazio digitalerako kompetentzia. 2) Ikasten eta pentsatzen ikasteko kompetentzia. 3) Elkarbizitzarako kompetentzia. 4) Ekimenerako eta espiritu ekintzailerako kompetentzia. 5) Norbera izaten ikasteko kompetentzia.

oinarrizko konpetentziei⁹ erreparatuta 7 konpetentzia daude, horietako bat “Hizkuntza-eta literatura-komunikaziorako konpetentzia”.¹⁰

Oinarrizko zehar konpetentziei eta diziplina barneko/arteko oinarrizko konpetentzien arabera, argi ikusten da hezkuntzan laburpena lantzeak izan dezakeen garrantzia. Horren erakusgarri dira ondorengo adierak agertzea curriculumean: “testuak ulertzea eta jarrera kritikoz balioestea”, “testuak sortzea”, “testuak sortu eta ulertzea” eta “testuak interpretatzea”.

Laburpenak hezkuntzan izan dezakeen zeresana ikusita, kontuan izan behar dugu zenbateko neurrian eta nola garatzen den laburtze gaitasuna, baita hori egiteko baliabideen egoera zein den ere. Esaterako, sarritan eskoletan ez dira laburpenak lantzen tresnarik ez dagoelako. Usu izaten den beste arazo bat denbora da, 20 ikasleren laburpenak zuzentzeak eta horiekin lan egiteak denbora luzea eskatzen baitu. Eskolako materialei erreparatuz gero, laburpena bera gutxitan lantzen da, egia da laburpen gaitasuna lantzeko beharrezkoak diren hierarkiak eta erlazioak lantzen direla, esaterako buru mapak erabiliz; baina ariketa horiekin ez dira idatzizko gaitasunak lantzen. Azken batean, horrek guztiak dakarrena da gramatika, progresio tematikoa, konexioa, kohesioa eta koherentzia bezalako baliabideak modu eskasagoan lantzea.

Aipatutako beharizan horiek asebetetzeko, baliabide didaktikoak eta laburpenak kudeatzeko tresna berriak sortu behar dira, laburtze gaitasuna ahalik eta ondoen garatu ahal izateko. Kasu honetan laburtze gaitasuna eta berarekin bat idatzeko konpetentziak garatzen laguntzen duten baliabide edo tresnez ari bagara ere, nabarmendu nahi dugu baliabide eta tresna horiek disziplinarrekotasunean oinarritutako hezkuntzan aplikatzeko modukoak izatea hobesten dugula.

Horiek horrela, lan honen helburua da irakasleentzat baliagarri izango den laburpenak jasotzeko eta aztertzeko Compress-Eus sistema aurkeztea.¹¹ Tresna honek bi laburpen mota jasotzea ahalbidetzen digu, estrakziozkoak eta abstrakziozkoak. Estrakziozkoetan testuko esaldiak ezabatzen dira, baina ez da inolako aldaketarik egiten

⁹ Diziplina barneko/arteko oinarrizko konpetentziak hauek dira: 1) Hizkuntza-eta literatura-komunikaziorako konpetentzia. 2) Matematikarako konpetentzia. 3) Zientziarako konpetentzia. 4) Teknologiarako konpetentzia. 5) Konpetentzia sozial eta zibikoa. 6) Artearen konpetentzia. 7) konpetentzia motorra.

¹⁰ Konpetentzia horrek honako osagai hauek ditu:

- a) Hainbat esparrutako hitzezko, idatzizko eta ikus-entzunezko testuak, analogikoak zein digitalak, ulertzea eta jarrera kritikoz balioestea, xede edo helburu pertsonalak, sozialak edo akademikoak erdiesteko.
- b) Esparru pertsonal, sozial edo akademikoen berezko hitzezko, idatzizko eta ikusentzunezko testuak sortzea, eraginkortasunez erantzuteko askotariko komunikazio-beharrizanei.
- c) Hizkuntzen erabilera-arauei eta hizkuntzen sistemari buruzko jakintza erabiltzea hitzezko eta idatzizko testuak sortu eta ulertzean, komunikazio egoki eta eraginkorrak egiteko.
- d) Literatura-testuak interpretatzea eta balioestea, testuen esanahia modu partekatuan eraikita; eta, horren bidez, mundua ulertzea, literatura-ondarea banako eta taldeko esperientzia sinbolizatze modu gisa balioestea, norberaren kultura-nortasuna eraikitzea eta sentsibilitate estetikoa garatzea.
- e) Gure gizartearen hizkuntza- eta kultura-aniztasuna aitortzea eta balioestea, aniztasun horrekiko jarrera ona izateko.

http://www.hezkuntza.ejgv.euskadi.eus/contenidos/informacion/heziberri_2020/eu_2_proyec/adjuntos/O_H_curriculum_osa.pdf -tik berreskuratua.

¹¹ COMPRES-EUS tresna CLARIN-ERIC europar azpiegituraren barruko K-zentru batean, IXA-CLARIN K-zentruan garatu da (Bel et al., 2016). IXA-CLARIN K-zentruak hizkuntzaren prozesamenduko tresnak erabilerrazak izateko helburua du, ikerketarako eta hezkuntzarako. Tresna hori hemen proba daiteke: <http://ixa2.si.ehu.es/clarink/tresnak/compress-eus>.

testu horietan. Abstrakziozkoetan, ostera, laburpena ematen saiatzen gara, jatorrizko testua oinarri izanda, baina, testuko esaldiek ez dute zertan bere horretan agertu, kasu honetan testua berregin egiten da.

1. Marko teorikoa

1.1. Testuen laburpena

Laburtzea idatzitakoa ulertzeko teknika da eta laburpeneko informazioak adieraziko digu testua zenbateko neurrian ulertu den (Sanz, 2005). Sanzen (2005) arabera, laburpena egitea ulertzea baino konplexuagoa da. Ulertutako testuaren barne-irudikapena (proposizioen erlazio-egitura) egiteaz gain, irakurleak erabaki beharko du zein diren barne-irudikapen horretako osagai garrantzitsu eta nagusiak (gaiaren eta azpigaiaren arteko koherentzia-erlazioak). Beraz, laburtzeak bi osagai ditu bere baitan, testua ulertzea eta ideiak hierarkizatzea.

Ulermenari dagokionez, irakurleak alde zuzenetik duen informazioa eta testuak dakarrena uztartu behar dira. Horako prozesu horrek jakintza garatzea dakar, baina baita hizkuntza-gaitasuna bera ere. Hori horrela izanda, kontuan hartu beharrekoa da esanahia ez dagoela soilik testuan, baizik eta irakurleak eraikitzen duela bere ezagutzak eskemak aktibatuz eta horrek, testuaren behearazko prozesaketaz baliatuz, edukiei buruzko iragarpen kognitiboak egitea ahalbidetzen baitio. Horretaz gain, irakurleak testuak dioena aztertzen du, bere iragarpen kognitiboak baieztatuz edo berrikusiz, horri testuaren goranzko prozesaketa esaten zaio. Testua irakurri ahala informazioak bere eskemekin bat egiten badu, hasieran zituen iragarpenak baieztatuko ditu, informazioa ulertuz eta bere eskemetan barneratuz, edo informazioa guztiz bat ez badator, norbere eskemak aldatu eta aberastuko dira. Hala ere, gerta liteke iragarpenok ez betetzea eta orduko eskemekin ezin konparatzea, kasu horretan ezin izango du bere informazioa egoki kodifikatu eta horrek bi ondorio ekar litzake; testua ez ulertzea edo informazioa distortsionatzea.

2.2 Laburpen automatikoa

Testuen laburpen automatikoan egindako ikerketetan, besteak beste, eta gure lanarekin lotuta, bi alderdi bereizteko joera ageri zaigu. Batetik, estrakzio bidezko laburpenak eta, bestetik, abstrakzio bidezkoak. Estrakziozko laburpena egitean testuak ez du inolako aldaketarik izaten, hots, testuko zati garrantzitsuenak hartzerara mugatzen gara. Adibidez, testu bat irakurri ostean zatirik garrantzitsuenak azpimarratzea. Ostera, abstrakziozko laburpenetan testua berregin egiten da, testuaren egitura eta erabilitako hitzak aldatuz. Laburpen automatikoan abstrakziozko testuekin aritzea oso zaila da; izan ere, ideia garrantzitsu berberak modu ezberdinean idatzita baitaude (Molina et al., 2011) eta zaila da mota horretako laburpen baten kalitatea ondo neurtzea. Horregatik, ez da harrizkoa estrakziozko laburpenen arloan gehiago aurreratu izana egundaino.

Gure lanaren oinarri garrantzitsuenetako bat Molinak (2013) eginiko lana da. Haren esanetan laburpen automatikoen sorkuntza erronka handia dakarren gaia da, eta bi familia metodologiko sortu ditu, aipaturiko estrakzioa eta abstrakzioa, hain zuzen ere. Esaldien konpresioak bi familien arteko zubia eraikitzen du Molinaren hitzetan. Bere lanean esaldien konpresioari buruzko ikerketa aurkezten du eta laburpenean aplikagarria den eredu lineal bat proposatu, esaldi-arteak segmentuak ezabatzea aurreikusten duena.

Eredu honek urre-patroizko testuetatik zuzenean erauzten ditu ezaugarriak eta diskurtso-segmentuak mantentzeko edo ezabatzeko urre-patroiko corpuseko erabakiekin dago entrenatua testuinguru guztia eta sortutako testua kontuan hartuz. Analisi estatistikoa baliatuz zein segmentu ezabatu aurreikusten duten ezaugarri esanguratsuenak ateratzen dira. Horren ostean, eredurik onena erabiltzen da esaldiak konprimatzeko.

3. Erlazionaturiko lanak

COMPRESS-EUS sistema egiterakoan Molinak (2013) eginiko lana oinarri izan dugun arren, gure helburuetara iristeko moldaketa batzuk egin dizkiogu. Molinak (2013) dokumentuen laburpen automatikoak lortzea du helburu, horretarako perpausen konpresioan oinarritzen da eta estrakzio teknikak erabiltzen ditu. Guk ostera, abstrakziozko laburpenak ere biltzeko sistema eratu dugu, estrakziozko laburpen batean egon daitezkeen akats gramatikal eta progresio tematiko desegokiak zuzentzeko. Horretaz gain, esana dugun moduan, biltzen ditugun laburpenak erlaziozko diskurtso-egituran (Iruskieta 2014) aztertu nahi ditugunez, diskurtso-unitateak eta testuko ideia nagusiak zehaztu ditugu gure sisteman.¹²

DBH 4ko ikasleen laburpenak deskribatu ahal izateko, COMPRESS-EUS tresna erabili du Garinek (2017). Programa horri esker estrakziozko eta abstrakziozko laburpenak bildu ditu. Deskribapena egiteko, bi testuak (estrakziozkoa eta abstrakziozkoa) hartu eta alderatu ditu, testu batetik besterako garapena kontuan hartuta. Ikerketan COMPRESS-EUS tresnari buruzko zenbait ondorio interesgarri atera ditu. Abstrakziozko laburpenak egiteko tresna erabilgarria dela badio ere, ikasleak tankera honetako tresnak erabiltzera ez daudela ohituak nabarmendu du. Ohitura falta hori estrakziozko laburpenetan nabari da gehien Garinen esanetan, diskurtso-unitate gutxi ezabatzeko joera baitute ikasleek. Estrakziozko laburpenak egiteko gaitasuna gehiago garatu beharko litzatekela azpimarratzen du.

4. Compress-eus sistema

Lehendabizi, COMPRESS-EUS sistemaren ezaugarri nagusiak emango ditugu aditzera eta, ondotik, tresnaren erabilera azalduko dugu.

4.1. Ezaugarriak

Irakasleen eta ikasleen laburpenak biltzeko tresna da COMPRESS-EUS sistema. COMPRESS-EUS sisteman bi testu mota sartuko ditugu, ikasleen eta irakasleen laburpenak bildu ahal izateko, kasu honetan estrakziozkoak eta abstrakziozkoak: i) Alde batetik, jatorrizko testu laua, hots, laburtu beharko den testua. ii) Bestetik, RSTko¹³ erlazio-egituraz etiketatua¹⁴ dagoen testua, teoria horretan adituak diren irakasleek eginikoa. Erlazio-egitura horrekin zehazten baitu COMPRESS-EUS sistemak ideiarik garrantzitsuena (unitate zentrala).

¹² Etorkizunean koherentzia erlazioak ere inplementatu nahi dizkiogu sistemari.

¹³ *Rhetorical Structure Theory* (RST) Mannek eta Thompsonek (1987) proposatu zuen koherentzia hizkuntzalaritza konputazionalan koherentzia deskribatzeko.

¹⁴ Adituen laburpenak dira Compress-Eus sisteman urre-patroizko (Gold Standard) corpusa. Hau da, corpus zuzena, ontzat hartuko duguna eta ikasleen laburpenak zuzentzeko erabiliko dugun corpusa.

Jarraian, aipatu berri ditugun bi testu motak eta unitate zentralaren (UZ) adibide bana azalduko ditugu.

Jatorrizko testua da laburtuko den testua. Aipatzekoa da edozein testu sar daitekeela COMPRESS-EUS sisteman, horrek ikasleen beharrianen arabera testu ezberdinak laburtzeko aukera ematen digu. Jatorrizko testua oinarrizko diskurtso-unitateetan (EDU) banatzen da. Sistemari esker, laburtu nahi den testuko diskurtso-unitate guztiak jatorrizko testuaren parrafoen arabera sailkatzen dira. Ondorengo taulan jatorrizko testu baten adibidea ageri da, “Erbia” izeneko testua.¹⁵

<p>1.paragrafoa</p> <p>[ERBIA]</p> <p>2.paragrafoa</p> <p>[Erbiaren lagunik ez dugu ezagutzen animalien munduan.] [Arerio franko ditu, ordea,] [horiek guztiak haragijale: erbinude txikia edo erdi barazkijalea den azkonar alferra, kasu.]</p> <p>3.paragrafoa</p> <p>[Jazarpen horren ondorioz, erbiak ikasi du dena mesfidantzaz ikusten, zuhur-zuhur jokatzeko eta ihesari emateko gaitasunak ontzen.] [Arriskupean dagoen erbi batek ez du lerro zuzenean jotzen, beste animaliek egiten duten bezala,] [sigi-sagan baizik, eta jauzi xebreak ematen ditu,] [horrela etsaiari nahasmendua eragiten dio.] [Zoriak laguntzen badio eta gordelekura iristea lortzen badu,] [zuzenean barrura sartu beharrean,] [harrapatzailea desorientatzeko asmoz,] [aztarnak nahasten ditu.]</p> <p>4.paragrafoa</p> <p>[Erne eta arretaz ibilita ere,] [makina bat erbi harrapatzen dituzte.] [Naturak horrela aurreikusita,] [eta espeziaren iraungipena zailtzeko,] [oso ugalkorra egin zuen animalia hau.] [Hamabost hilabeteko eme batek urtean lau umaldi izan ditzake,] [guztira 9 ume erdituz.] [Eman dezagun 9 horietatik 4 emeak direla,] [eta ugalketa bide horretatik jarraituz gero,] [9 urtean erbi batek 65.501 ume izan ditzake.]</p>
--

Taula 1. *Erbia* testua paragrafoetan bereizita eta diskurtso-unitateetan banatuta

4.1.1. Unitate zentrala (UZ)

COMPRESS-EUS sistemak testuko unitate zentrala, hau da, unitaterik garrantzitsuena zein den zehaztuta izango du. Automatikoki jakin ahal izateko ea erabiltzaileak eginiko laburpenetan testuko unitaterik garrantzitsuena dagoen. Ondorengo taulan “Erbia” testuko unitate zentrala ageri da.

¹⁵ Testu hau Sanzek (2005) irakurmena lantzeko jarduerak prestatzeko proposatzen du, Lehen Hezkuntzako 3. ziklorako eta Bigarren Hezkuntzako 1. ziklorako. Atutxak eta Iruskietak (2017) Lehen Hezkuntzako ikasleen laburtze gaitasuna aztertzeko, lehendabizi ikasleekin jarduera batzuk egin zituzten eta hauek ere jarduera horietarako testu bera erabili zuten. “Erbia” testua I Taulan ageri da paragrafoetan bereizita eta diskurtso-unitateetan banatuta (parentesien bidez).

[Arerio franko ditu, ordea,]

Taula 2. Erbia testuaren unitate zentrala (UZ)

4.2 Erabilera

Jarraian COMPRESS-EUS sistemaren erabilera azalduko dugu. Gure asmoa ikasleen eta irakasleen laburpenak biltzea da; hala ere, webgunera sartzeko aukera duen edozein pertsonak erabil dezake tresna hau. Webgunera heltzeko honako helbidera jo behar du erabiltzaileak: <http://ixa2.si.ehu.es/clarink/tresnak/compress-eus>.

Hurrengo urratsa erregistratzea izango da, horretarako 1 Irudian ageri den lekuan klik egin beharko du erabiltzaileak eta ondoren 2 Irudian ageri diren datuak bete.

Hasi saioa

Erabiltzaile-izena

Pasahitza

Hasi saioa

[Dokumentuak laburtzeko erregistratu.](#) [Pasahitza ahaztu al?](#)

ERABILTZAILEREN ERREGISTROA

ERREGISTRATU AHAL EZATERO BEHARRI ESKATZEN JALDIZKIZUN DATUAK EMAN

Erabiltzaile-izena: [input field]

Lehen abizena: [input field]

Bigarren abizena: [input field]

NAN zenbakia: [input field]

Lanbidea (Irakaslea, ikaslea, ...): [input field]

Herria: GALDARAO

Ikasketako maila: [input field]

Posta elektronikoa: [input field]

Pasahitza: [input field]

Ereplikatu pasahitza: [input field]

Erregistratu

Irudia 1 Saioa hasteko gunea

Irudia 2 Erabiltzailea erregistratzeko gunea

4.2.1 Gida azkarra

Eskatzen zaizkion datuak sartu ostean, erabiltzaileak bere erabiltzaile izena eta pasahitza izango ditu eta horri esker saioa hasteko moduan izango da. Saioa hasten duenean gida azkarra aurkituko du erabiltzaileak, bertan sistemarekin aritzeko behar dituen jarraibide eta arau guztiak ditu. Jarraibide horiei jarraituz egingo ditu bi laburpen motak, lehendabizi estrakziozkoa eta ondoren abstrakziozkoa.

Prozesuko edozein unetan, testua berriz egiten hasteko, klikatu TESTUA BERRASI botoian.

TESTU KONPRIMATUA lortzeko, segmentuak aldeztu mugatu dira eta ezin dira moldatu.

Segmentuaren ertz borobilduak soilik adierazten du beronen amaiera, ertza zuzena bada, segmentuak hurrengo lerroan jarraituko du.

Gutxienez segmentu bat mantendu edo kendu beharko duzu testu bakoitzean, hurrengo testua egin ahal izateko.

Garrantzitsua soilik mantendu behar duzu.

Konprimatutako esaldi bat letra xehe batekin has daiteke, gero testua editatu eta horrelako zein bestelako akatsak konpondu ahalko dituzu.

Konprimitutako esaldi baten amaiera puntua ez den beste ikurren bat izan daiteke. Ahalik eta gehien konprimatu behar da.

Segmentuen gainean klik eginez gaitzen edo ezabatzen dira segmentuak.

TESTU KONPRIMATUA bukatuta, LABURPENA IKUSI ETA EDITATU botoian klikatu eta LABURPENA ESKUZ ZUZENTZEKO atala gaituko duzu

Nahi duzun laburpena osatzeko, edozelako zuzenketak egin ditzakezu, baita testu osoa moldatu ere

TESTUA BUKATU botoia sakatu ostean, ezin daiteke testu bera berriz egin eta hurrengo testua agertuko zaizu.

Testu guztiak amaitzean, amaiera-mezua agertuko zaizu.

Taula 3. COMPRESS-EUS erabiltzeko gida azkarra

4.2.2 Estrakziozko laburpena

Behin gida azkarra irakurrita, laburpenak egiteari ekin diezaioke erabiltzaileak; esan bezala, estrakziozko laburpena eginez hasiko da. Estrakzio lanaren nondik norakoak argiago ikusteko, honako irudi honetan erabiltzaileek ordenagailuaren pantailan aurkituko dutena ikusten dugu:

Irudia 3 Estrakziozko laburpena egiteko beharrezkoak diren jatorrizko testua eta testu konprimitua

3 Irudian ikus daitekeen moduan, erabiltzaileak bi testu ikusiko ditu. Ezkerreko testua laburtu beharreko testua da. Testu hori diskurtso-unitateetan ageri da segmentatuta (EDUetan) eta guk hango unitaterik garrantzitsuenak aukeratu ditugu estrakziozko laburpena egiteko.¹⁶ Horretarako, erabiltzaileak beharrezkoak ez diren segmentuak ezabatu beharko ditu, segmentuon gainean klik eginez. Eskuineko testuan, testu konprimatuan, erabiltzaileak klikatutako segmentuak ezabatu egingo dira, estrakziozko laburpena eratuz.

Gerta liteke ezabatu dugun unitate bat berreskuratu nahi izatea, laburtu beharreko testuan bere gainean klik eginda berreskuratuko genuke. Laburpena behin eginda, bi aukera ditugu, bukatu dugula esatea sistemari “Laburpena ikusi eta editatu” jartzen duen lekuan klik eginez edo baliteke gustura geratu ez izana egindakoarekin eta berriro hasi nahi izatea, kasu horretan “Berrasi” jartzen duen lekuan egingo dugu klik (baina orduan egin den guztia galduko da, sistemak ez baitu informazio hori gordeko).

4.2.3 Abstrakziozko laburpena

“Laburpena ikusi eta editatu” klikatuta estrakziozko laburpenarekin bukatuko dugu eta abstrakziozkoa egiteari ekin. Abstrakziozko laburpena testu konprimatuaren azpian ageri den leihotxoan egingo dugu “Laburpenak eskuz zuzentzeko” jartzen duen tokian. 4 Irudian ageri den moduan, leiho horretan aurretik egindako estrakziozko

¹⁶ Akats gehiago sortzen badute ere, automatikoki egin daitezke bi ataza hauek, diskurtso-unitateetan testua zatitu eta ideia nagusien detektatu, honako bi tresnekin: i) EusEduSeg tresnarekin testuak zatitu. Hemen proba daiteke: <http://ixa2.si.ehu.es/EusEduSeg/EusEduSeg.pl> (Iruskieta eta Zapirain 2015). ii) Ideia nagusien detektatzailearekin idaeia garrantzitsuena detektatu. Laburpen zientifikoetan ideia nagusia detektatzen duen tresna hemen proba daiteke: <http://ixa2.si.ehu.es/CU-detector/> (Bengoetxea et al. 2017).

laburpena dago eta han guk geuk egin nahi ditugun aldaketak egin diezazkiokegu testuari. Horri esker, estrakziozko laburpenak izan ditzakeen akats gramatikalak edo progresio tematiko desegokiak zuzentzeko aukera izango du erabiltzaileak, baita bere hitzez baliatzekoa ere.

Berregin duen testuarekin konforme gaudenean, “Testua bukatu” jartzen duen lekuan klik egin eta sistemak erabiltzaileak egin duena gordeko du.

Tresna hau ikasleek eta irakasleek erabil dezakete eta batzuek eta besteek egin dituzten laburpen guztiak jaso. Oro har, honako dokumentu hauek jasoko ditu: i) estrakziozko laburpenak, ii) abstrakziozko laburpenak eta iii) erabiltzaile guztien eragiketa nagusiak jasotzen dituen kalkulu-orria.

The screenshot shows the COMPRESS-EUS interface. At the top, there are buttons for 'Testua berrasi', 'Laburpena ikusi eta editatu', and 'Testua bukatu'. A counter indicates 'Berrikusi duzu 0 orainarte eginiko testuak'. The main content is divided into two columns: 'ERBIA' (original text) and 'Testu konprimatua' (compressed text). The 'ERBIA' column contains several paragraphs of text, some with highlighted words. The 'Testu konprimatua' column shows the compressed version of the text. Below the compressed text, there is a section titled 'Laburpena eskuz zuzentzeko' (Manual editing of the summary) with a text area for corrections.

Irudia 4 COMPRESS-EUS sistema abstrakziozko laburpena egiterakoan

4.3. COMPRESS-EUS tresnarekin datu multzo handiak kudeatzen

Erabiltzaile-multzo handi baten emaitzak automatikoki aztertzeko aukera eskaintzen du COMPRESS-EUS tresnak; izan ere, eragiketa nagusiak kalkulu-orri batean eskaintzen ditu. Tresnak ematen duen informazioaren laburpena egingo dugu 4 Taula.

Testu originala						Egile11k kendutako informazioa					
Dokumentua	Paragr.	Esaldia	EDU	Hitzak	UZ	Egile11	UZ	Paragr.	Esal.	EDU	Hitz kop
1	1	1	1	14	1	11	0				
		2	2	8							
		3	3	9						3	9
	2	3	4	12			2	3	4	12	

Taula 2. Adituek eginiko Erbia testuaren estrakzio laburpena

Taula horretan testu originalaren zenbait datu daude: dokumentu horrek 2 paragrafo ditu. Lehenengo paragrafoak 2 esaldi ditu eta testuko bigarren esaldiak 2 EDU ditu. Eskuman Egile11k kendutako informazioa deskribatzen da. Taulan dagoenaren arabera, egile horrek UZa mantendu du eta ez du kendu (bere balioa 0 baita), bigarren paragrafoan hirugarren eta laugarren EDUak kendu ditu. Estrakziozko laburpena denez, EDUak dituen beste hitz kendu ditu (hitz kop.), guztira: 21 hitz gutxiago.

Modu horretan erabiltzaile-multzo handi batean ideiarik garrantzitsuena (UZa) zenbatak kendu duten jakin dezakegu, baita zein paragrafo, esaldi edo EDU kendu duten, edo zenbat hitz kendu duten.

Horrez gain, erabiltzaileek egiten duten laburpena adituen laburpenarekin alderatzen da, hots, erabiltzaileak egindako laburpenaren eta adituaren arteko berdintasun eta desberdintasunak aztertzen dira, adituarena zuzentzat edo ontzat hartuz. Jarraian, aditu batek eginiko estrakziozko laburpena ageri da 5 Taulan, jatorrizko testuko diskurtso-unitate garrantzitsuenak aukeratuta dago eginda.

1.paragrafoa
2.paragrafoa
[Arerio franko ditu, ordea,]
3.paragrafoa
[Jazarpen horren ondorioz, erbiak ikasi du dena mesfidantzaz ikusten, zuhur-zuhur jokatzeko eta ihesari emateko gaitasunak ontzen.]
4.paragrafoa
[makina bat erbi harrapatzen dituzte.] [oso ugalkorra egin zuen animalia hau.]

Taula 3. Adituek eginiko Erbia testuaren estrakzio laburpena

Arerio asko ditu Erbiak. Hori dela eta, animalia mesfidatia da, zuhurtziaz jokatzeko duena eta, ihes egiteko ahalmen handiduna. Erbi asko harrapatzen dituztenez, naturak oso ugalkorak egin ditu.

Taula 4. Adituek eginiko Erbia testuaren abstrakzio laburpena

Abstrakziozko laburpena egitean (6 Taula), estrakziozkoan egon daitezkeen akatsak zuzendu ditugu. Kasu honetan, testuaren hasieran “Erbia” hitza jarri dugu, estrakziozko laburpenean ez baitago jakiterik zeri egiten dion erreferentzia. Egin den beste aldaketa bat “ordea” lokailua kentzea izan da, ez baita zuzena hurrengo esaldiarekin lotzeko. Azkeneko esaldian ere erreferentzia arazo bat dugu, estrakziozkoan ez dakigu nork edo zerk egin duen ugalkor erbia, honako honetan “naturak” egin duela zehaztu behar izan da. Azkenik, lexikoan dagoen aldea hartuko dugu kontutan; izan ere, laburpena egin duenak bere hitzekin berregin baitu laburpena.

5. Ondorioak

Lan honetan COMPRESS-EUS izeneko sistema aurkeztu dugu.¹⁷ Tresna honek ikasleek eta irakasleek eginiko laburpenak biltzeko eta erabiltzaile-taldearen laburpenak aztertzeko balio du; gainera, bi motatako laburpenak bildu ahal izango ditugu, alde batetik estrakziozko laburpenak, kasu honetan testuak ez du inolako aldaketarik, eta beste aldetik abstrakzioak, azken laburpen hauetan testua berreginda dago.

Testuak bildu ostean, gure nahia da laburpen horiek aztertu ahal izatea, zehazki esanda, diskurtso-egituran oinarritutako azterketa da bihar-etzi egin nahi duguna. Hori horrela izaki, gure tresnak ezaugarri jakin batzuk ditu, guk egin nahi dugun azterketa mota egitea erraztuko dizkigunak. Diskurtso-egituran oinarritutako azterketa errazteko, COMPRESS-EUS sistemak automatikoki adieraziko digu erabiltzaileentzat zeintzuk diren testu zati garrantzitsuenak. Horretaz gain, unitate nagusia erabiltzailearen laburpenean dagoen esango digu, hau ere automatikoki. Eta, bukatzeko, estrakziozko eta abstrakziozko laburpenen arteko alderaketa egiteko aukera ere ematen digu, oraingoz eskuz bada ere.

COMPRESS-EUS sistema eskolako ikasleen laburpenak aztertzeko baliabidea da, hots, gure asmoa da jakitea ikasle eta irakasleek zer-nola laburtzen duten eta tresna honek bide horretan lagunduko digu. Etorkizunari begira, gure asmoa laburtze gaitasuna garatzea ahalbidetuko duten baliabide didaktikoak sortzea da eta horiek garatzeko eta ebaluatzeko egokia da COMPRESS-EUS sistema, gure ustez. Sistemari aukera gehiago inplementatu nahi dizkiogu, esaterako erabiltzaileari *feedback* automatikoa ematea egindako laburpenaren arabera. *Feedbacka* ezberdina izan daiteke, gomendio hutsetatik hasita, laburpena garatzeko ariketak proposatzeraino. Ikasle zein irakasleei emango diegun *feedbacka* zeinahi izanda ere, diziplinartekotasunean oinarrituta egitea da gure asmoa; izan ere, uste dugu gaur egungo hezkuntzaren ardatzetako batek diziplinartekotasuna behar duela izan. Horregatik, ez da gure nahia laburtze gaitasuna modu isolatuan lantzea, baizik eta curriculumeko beste gaitasun edo edukiekin batera, denen arteko oreka eta kalitatea bermatuz.

6. Bibliografia

- Atutxa, U., & Iruskieta, M. *Abstrakziozko laburpen-testuak ebaluatzeko metodoa. Lehen Hezkuntzan: testu-luzera, koherentzia-erlazioak, gai nagusia. Tantak*, 29(1), 53-80 or.
- Bel, N., González-Blanco, E., & Iruskieta, M. (2016). CLARIN Centro-K-español. *Procesamiento del Lenguaje Natural*, 57, 151-154.
- Bengoetxea, K. Atutxa, A., & Iruskieta, M. (2017). Un detector de la unidad central basado en técnicas de aprendizaje automático en textos científicos para el euskera. *Procesamiento del Lenguaje Natural*, 58, 37-44
- Iruskieta, M. (2014). *Pragmatika erlaziozko diskurtso-egitura: deskribapena eta bere ebaluazioa hizkuntzalaritza konputazionalean*. Donostia: Euskal Herriko Unibertsitatea.
- Iruskieta, M., & Zapiroain, B. (2015). Euseduseg: A dependency-based edu segmentation for basque. *Procesamiento del Lenguaje Natural*, 55, 41-48.

¹⁷ COMPRESS-EUS sistema hemen probatu daiteke: <http://ixa2.si.ehu.es/clarink/tresnak/compress-eus>.

- Mann, W.C., & Thompson, S. A. 1987. Rhetorical Structure Theory: A theory of text organization. *Text*, 8(3), 243-281.
- Molina, A., da Cunha, I., Torres-Moreno, J. M., & Velazquez-Morales, P. (2011). La comprensión de frases: un recurso para la optimización de resumen automático de documentos. *Linguamática*, 2(3), 13-27.
- Molina, A. (2013). Compresión automática de frases: un estudio hacia la generación de resúmenes en español. *Inteligencia Artificial*, 16(51), 41-62.
- Sanz, A. 2005. *Irakurmena lantzeko jarduerak nola prestatu: Lehen Hezkuntzako 3. zikloa eta DBHko 1. zikloa*. Nafarroako Gobernua
- Garin, I. (2017). *DBH 4ko ikasleen testu laburpenak deskribatzen, IKTak erabiliz. UPV/EHU-UEU HEZikt Berezko Tituluko proiektua*. Eibar.

ZUBERERAZKO DARDARKARIEN AZTERKETA SOZIO-FONETIKOA¹⁸**Xantiana Etchebest****Euskal Herriko Unibertsitatea eta Université de Pau et des Pays de l'Adour**xetchebest@hotmail.fr**Laburpena**

Euskaraz eta hemen zubereraz, bi dardarkari mota desberdin badira: /r/ bakunak eta /r/ anizkunak. Bakunaren kasuan, edo ahoskatzen da edo desagertzen da bi bokal artean /hari/ edo /hai/ kasuetan bezala. Anizkunaren kasuan aldiz, bi motatakoak badira: apikariak /r/ edo ubularrak /ʁ/. Horretarako, informateek ahoskatze hautu desberdinak badituzte. Lan horretan ikusiko da adina den faktore sozialen bidez hautuak eginak direla, hala nola gazteek egiten dituzten hautuak ez dira baitezpada zaharrek egiten dituztenak. Dardarkari horien guztien aztertzeo, inkestak egin ditugu. Inkesta hori bideratua izan da itzulpenen bidez eramana izan baita. Honetan, hitz zenbaitzuen itzultzea galdegina izan zaie. Aldi bakoitzean ber hitzak izan dira ondotik konparaketa baten eramateko. Hiztunak haien adinaren arabera hautatuak izan dira (gazte, heldu, zahar). Ondotik, nagusiki bi programarekin lan egin da: Praat eta SPSS programak. Eta programa horiei esker bildu diren datuekin, azterketa estatistikoak eraman dira eta gaur egungo dardarkarien argazki bat atera ahal izan da.

Gako-hitzak: fonetika, bariazioa, dardarkariak, euskara, adina

1. Sarrera

Gauza jakina da hizkuntzak beti aldatzen eta garatzen direla. Trask eta Milroy-en errana oinarritzat hartzen baldin bada: «One of the fundamental things you need to understand about languages is that they are always changing.» (1996: 1) zubererak ez du salbuespen bat sortzen. Gaur, Zuberoan baina orokorki, Euskal Herrian ere, euskaldun gazteek eta adineko euskaldunek ez dute berdin hitz egiten. Eta hori puntu desberdinen arabera semantikan edo gramatikan bezala. Baina hemen fonetikaren arabera zentratua izanen da lana. Beraz, lan honetan frogatu nahi izan dena da aldaketak badirela. Eta aldaketa horiek belarri hutsez entzuten baitira, fonetikaren lantzeko nahia piztu zait. Baina aldaketa guztien aztertzeo, fenomeno edo kasu honetan fonema sail bat hautatu behar zen. Eta aski emankorra den /r/ dardarkariei buruz interesatu naiz lan honetan. Euskal Herrian, baina beste hizkuntza zenbaitzuetan ere, dardarkariak badira. Ladefoged eta Maddieson-ek (1996) diote munduan, hizkuntzen % 76,7k dardarkariak badituztela. Euskararen eta hemengo kasuan, zuberera kopuru honetan sartzen da. Gainera Txillardegik dioen bezala (1978), euskarak bi dardarkari desberdin baditu: dardarkari ahulak eta dardarkari azkarrak. Lan honetan, dardarkari bakunak eta anizkunak deituko direnak. Hala ere, hemen onartuko da azkarretan, edo anizkunetan bi desberdin badirela. Alde batetik, Txillardegik azkarrak deitzen dituen apikariak izanen dira eta bestalde, ubularrak izanen dira. Hauek, Txillardegik (idem) adibidez ez ditu onartzen gaurko euskaran.

¹⁸ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

2. Metodologia

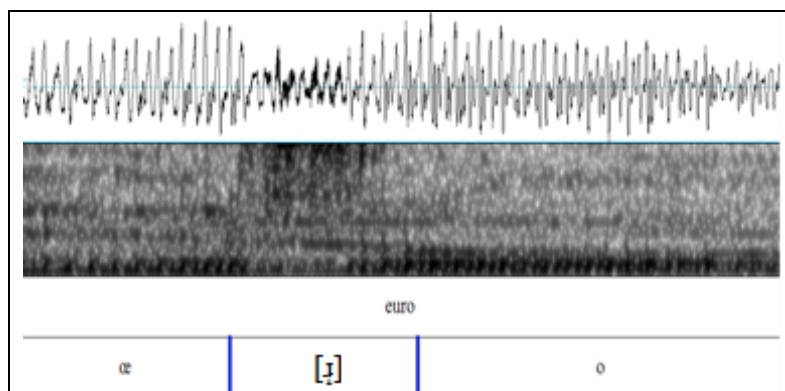
Lan honen helburua zuberotar dardarkarien argazki baten ateratzea izan baita, pauso nagusia jendeen inkestatzea izan da. Horretarako, jendeak hautatuak izan dira eta bereziki adinaren arabeko hautaketa bat egin da. Jendeak hiru adin multzotan bildu dira. Lehenik gazteen adin multzoan 15-20 urte arteko gazteak hautatuak izan dira haiek D eredia burutik buru ezagutzen baitute eta besteek ukan ez duten posibilitate bat izan baita. Haatik, inkestatutakoetan ez dira denak D ereditik pasa. Bestalde, 30-40 urteko helduak hautatuak izan dira, haiek D ereduaren lehen gizaldia markatzen baitute. Eta azkenik, 60 urte baino gehiagokoak hautatuak izan dira D eredia ez baitute ezagutu eta bortxaz etxean ikasitako euskara bat ematen dutelako. Beraz, eraman izan den lan honetan, euskararen jatorria eta adina hurbiletik lotuak dira. Gazteetan 11 hitzun bildu dira, helduetan 9 eta zaharretan 10. Jende horiei frantsesetik euskarara 64 hitzen itzultzea galdegina izan zaie. Hitz horietan guztietan, hitzak dardarkariak testuinguru desberdinetan lortzea entseatu da: hitz bukaeran, bokalarlean, kontsonante desberdinen aitzinean. Horren ondotik, 10 hitzekin lan egitea deliberatua izan da. 10 hitz hautatuak izan dira (herri, hari, ürxaix, armada, erdi, orkatx, euro, hor, hurra eta oro).

3. Datuen analisia

Datuen analisiaren egiteko, gorago errana izan den bezala, Praat programa baliatua izan da. Programa honek ematen dituen espektrogramei esker, dardarkari desberdinak lortu dira. Lan honetan, espektrogramak landu direnean, 9 dardarkari desberdin lortu dira. Lau dardarkari bakun, bi dardarkari anizkun apikari eta 3 dardarkari anizkun ubular. Horretarako, azpiatal hau hiru sekziotan banatuko da.

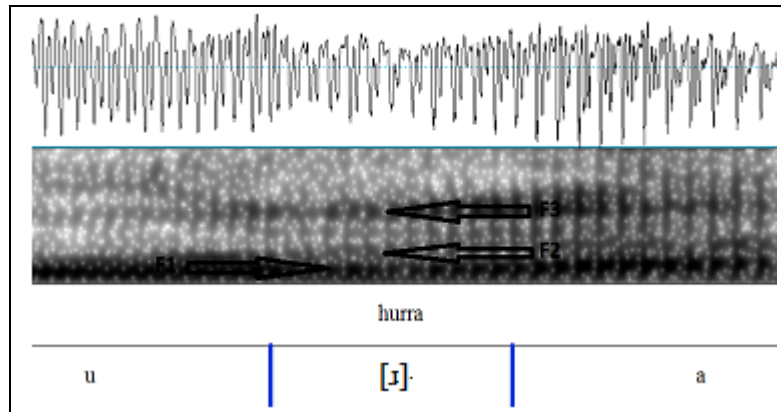
3.1. Dardarkari bakunak

1. irudian agertzen den bezala, lortutako lehen dardarkari bakuna [ɹ] izan da. Dardarkari honek, 3 formakinak aurkezteaz gain, hirugarren formakinetik gora, zarata ere markatzen du. Alofono hori, behin baizik ez da agertu. *Euro* hitzean agertu da, eta emazte zahar batek eman du.



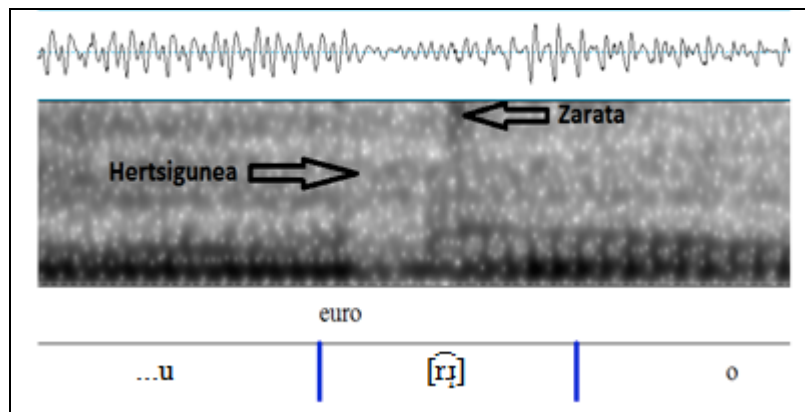
1. irudia: [ɹ] alofonoaren espektrograma

2. irudian, bigarren dardarkari bakun bat agertzen da eta hau da [ɹ] alofonoa. Honek, aurrekoak bezala, hiru formakinak pasarazten ditu. Aldiz, zaratarik ez du markatzen. 1. irudian eta 2. irudian agertzen diren alofonoak, hurbilkariak deitzen dira ez baitute hutsunerik markatzen. [ɹ] alofonoaren kasua 12 aldiz emana izan da eta beti bokal artean zelarik: *oro*, *euro*, *hurra* eta *erri*, hitzetan agertzen baita.

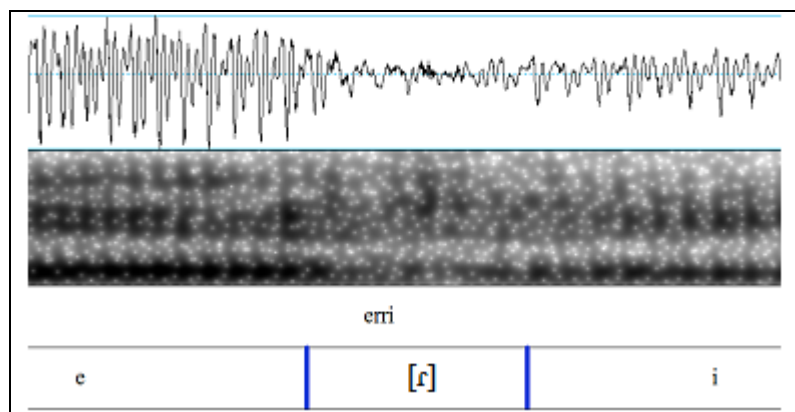


2. irudia: [ɹ] alofonoaren espektrograma

[$\widehat{\text{r}}_i$] alofonoaren kasua 3. irudian agertzen da. Ikusten da alofono honek hastapenean hertsigune bat markatzen duela eta ondotik, zarata markatzen duela. Alofono hau, 4 aldiz agertu da, eta *euro* maileguan edo *hurra* hitzean agertu da.

3. irudia: [$\widehat{\text{r}}_i$] alofonoaren espektrograma

Azken dardarkari bakuna aztertzen bada 4. irudian, erran nahi baitu [ɹ] alofona, ikusten da honek hutsune bat markatzen duela dardarkari osoan zehar, eta beraz, formakinik ez dela agertzen. Soinu hori, 24 aldiz agertu da. Bakun erabiliena da beraz. Eta hori, bokal artean eta hitz bukaeran agertu da *hor* hitzean bezala.

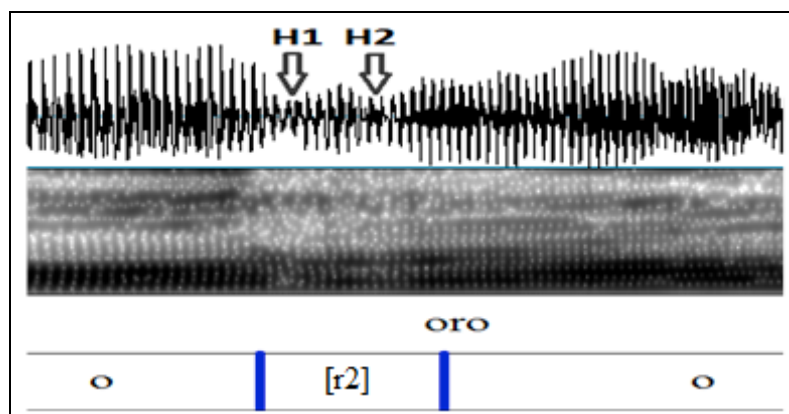


4. irudia: [ɹ] alofonoaren espektrograma

Ikusi da beraz, bakunak orotara 41 aldiz erabiliak izan direla. Baina erran behar da dardarkari multzo hau zaharrek gehiengoan erabiltzen duten bat dela. 41 ahoskatze horietan, 4 aldiz gazte edo heldu batek eman du. Erran nahi du beraz, bakunak zaharrek erabiltzen dituzten dardarkariak direla.

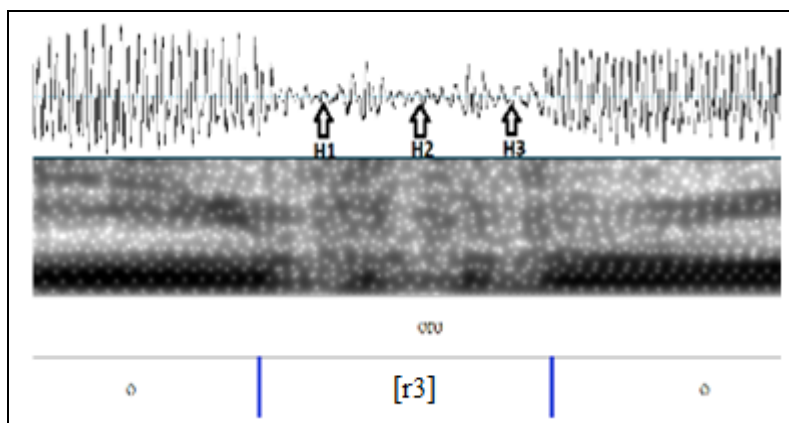
3.2. Dardarkari anizkun apikariak

Dardarkari anizkun apikarietan, bi alofono desberdin lortu dira : [r2] eta [r3]. 5.irudiaren kasuan, agertzen da bi hertsigune badirela, beraz [r2] baten kasuan gara. Soinu hori, 39 aldiz emana izan da. 8 hitz ezberdinetan markatua izan da erran nahi baitu: *euro*, *erri*, *hurra*, *armada*, *erdi*, *orkatz*, *ürxaintx* eta *hor* hitzetan. Beraz, bokal artean, kontsonante aitzinean edo hitz bukaeran agertu den soinu bat da.



5. irudia: [r2] alofonoaren espektrograma

Aitzineko taulan 2 hertsigune markatzen baziren, 6. irudiaren kasuan 3 hertsigune agertzen dira. Kasu honetan beraz, [r3] baten aitzinean gara. Hau lau aldiz agertu da eta *oro* hitzean baizen ez da agertu.



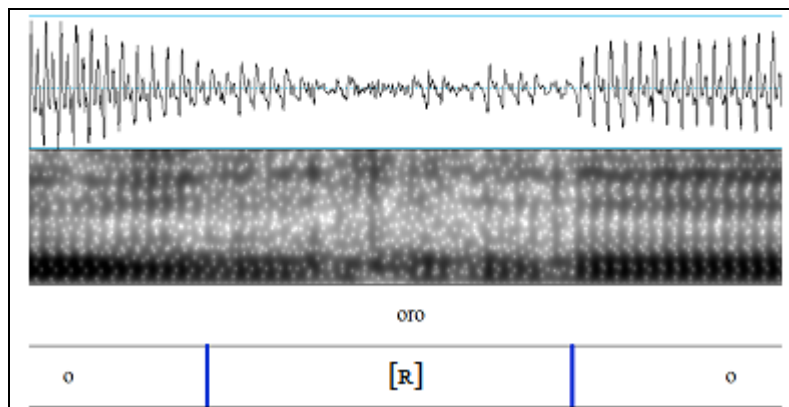
6.irudia: [r3] alofonoaren espektrograma

Dardarkari anizkun apikariak, salbuespenik gabe, zaharrek ahoskatu duten soinu multzo bat da. Hori gazte eta helduen gizaldien artean, desagertua den soinu bat da.

3.3. Dardarkari anizkun ubularrak

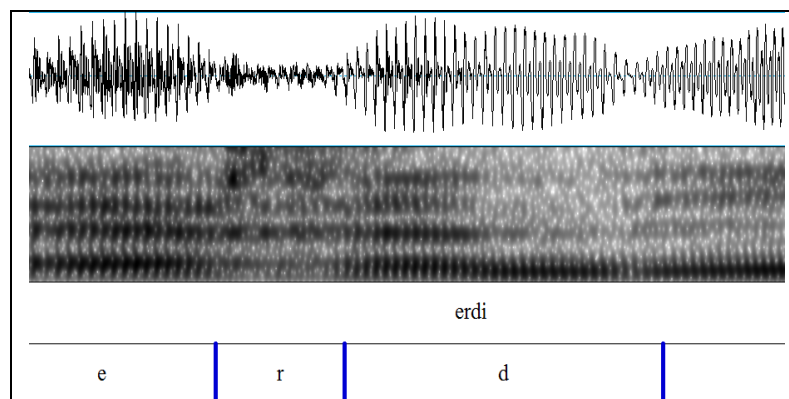
Dardarkari anizkun ubularretan hiru alofono desberdin lortu dira. 7.irudian [R] alofona agertzen da. Irudi honetan agertzen da hertsigune anitz markatuak direla

dardarkari osoan zehar. 25 aldiz baliatua izan da. Adin talde guztiek erabiltzen duten soinu bat da. Soinu hau, bi bokalen artean ahoskatua izan da *oro*, *hurra* edo *euro*, baina ere kontsonante aitzinean *orkatz*, *ürtxaintx* eta *erdi* hitzekin edo hitz bukaeran *hor* hitzarekin.



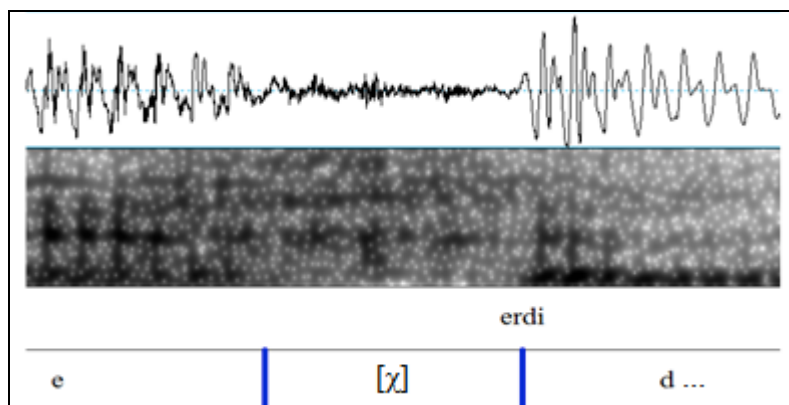
7.irudia: [ʀ] alofonoaren espektrograma

8. irudian agertzen den alofona gehienik erabilia izan dena da. [ʀ] alofona orotara 96 aldiz emana izan da. Soinu honek, [ʀ]-k bezala, hertsiguneak ezagutzen ditu, baina gainera, hirugarren formakinetik gora, zarata badu. Soinu hau, Bi bokalen artean *oro*, *hurra*, *erri* edo *euro* hitzekin ikusten da, baita kontsonante aitzinean ere, *orkatz*, *ürtxaintx*, *erdi* eta *armada* hitzekin. Hiru gizaldiek bederen behin ematen duten soinu bat da.



8. irudia: [ʀ] alofonoaren espektrograma

Lortutako azken alofona, 9. irudian agertzen den [χ] alofona da. Honek berezitasun bat badu, ahoskabetua den dardarkari bat baita. Hori ikusten da F0 formakin desagertzen baita dardarkari osoan zehar. Soinu hau, 53 aldiz erabilia izan da eta zaharrek baliatzen ez duten soinua da. Soinu hau, testuinguru guztietan atzemaia izan da, eta gehienik hitz bukaeran lortutakoa izan da.

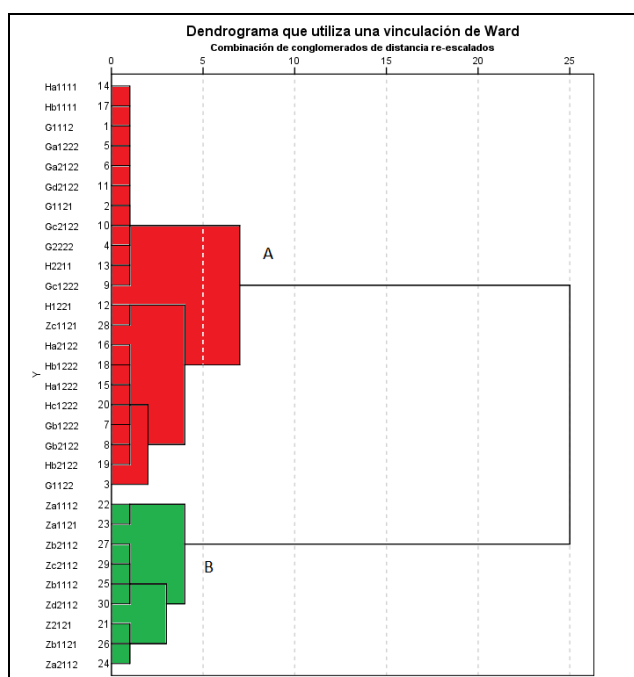


9. irudia: [χ] alofonoaren espektrograma

Dardarkari anizkun ubularrak, gazteek eta helduek gehienik erabiltzen duten soinu multzo bat da. Zaharrek orotara, 5 aldiz erabiltzen dute. Beraz erraiten ahal da gaur egun nagusitzen diren dardarkariak gazteek eta helduek erabiltzen dituzten dardarkari anizkun ubularrak direla.

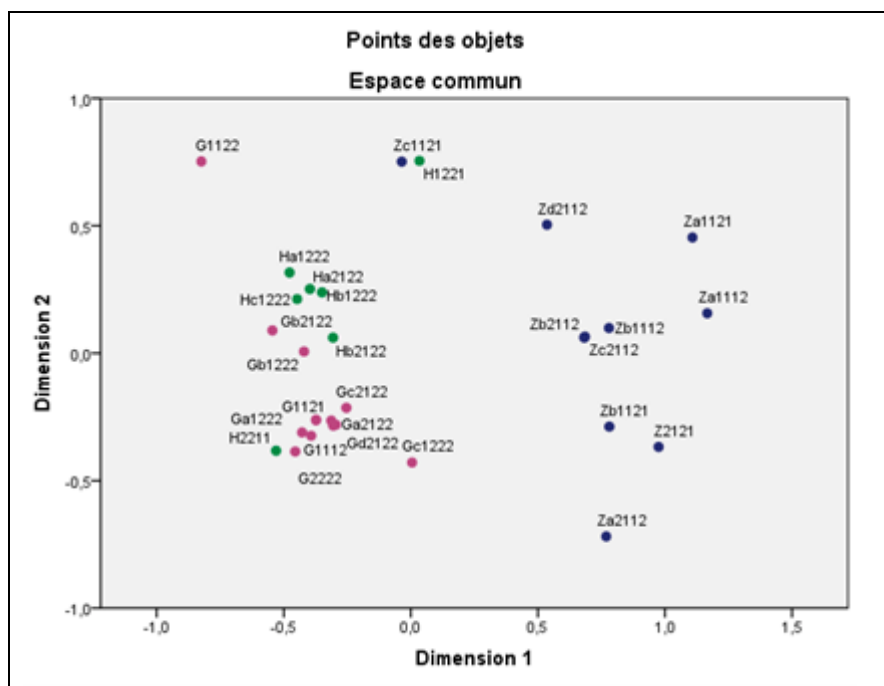
4. Emaitzak

Espektrogramak aztertzerakoan, agertu da adinaren arabera hausturak bazirela dardarkari hautuaren egiterakoan. Hala nola, zaharrek bakun edo apikarietara jotzen dute; aldiz, gazte eta helduen ubularrak nahiago dituzte. Lan honetan, iraupena, energia eta F2 formakina aztertzerakoan adinaren araberako esanguratasun estatistikoak lortu dira. Eta SPSS programaren bidez lortutako 1. dendogramak erakusten du argiki zaharrek gizaldi desberdin bat markatzen dutela. Ikusten da gorritz markatuak diren informanteek talde bat osatzen dutela, eta bestalde, berdez markatzen dutenek beste bat osatzen dutela. Gorritz markatuak direnak gazte eta helduak dira. Hala ere, talde honetan, agertzen da gazteek talde bat osatzen dutela eta helduek beste bat.



1. dendograma: adinaren araberako informanteen sailkapena

10. irudi honetan, agertzen da hiru gizaldiek hiru talde osatzen dituztela. Ikusten da irudiaren eskuin aldean, urdinez markatuak diren zaharrek talde bat osatzen dutela nahiz eta aski barreiatuak diren. Bestalde, ezker aldean, arrosez diren gazteek eta berdez diren helduek beste talde bat osatzen dute. Baina gorago errana izan den bezala, agertzen da ezker aldeko goiko partean berdea dela nagusi, eta azpiko partean aldiz, arrosa dela nagusi. Agertzen da nahiz eta zaharren desberdintasuna bezala ez hain markatua izan, beraz, heldu eta gazteen artean desberdintasuna badirela ere.



10. irudia: informanteen eskala dimentsio anizkuna

5. Ondorioak

Ondorio gisa, gorago aipatu den bezala, erranen da zaharren eta besteen artean haustura bat badela. Erran nahi baitu, zaharrek dardarkari desberdinak erabiltzen dituztela eta beraz, 10. irudian agertzen den bezala, maparen bestaldean agertzen direla. Beraz, lortutako argazkian agertzen den bezala, zaharrek desberdin egiten dute, eta gazte eta helduen artean nahiz eta desberdintasun txipi bat ukan, bateratu egiten dira. Aurreuposatzen da honekin, dardarkari anizkun apikariak, baita dardarkari bakunak ere emeki-emeki ahozkoan desagertuko direla. Hala ere, kantuetan oraindik atxikiak diren soinuak baitira, ezin da erran testuinguru guztietan desagertuko diren soinuak direla.

6. Bibliografia

- Coyos, J.B. (2002). Zubereraren idazkeraz: ortografia, fonetika eta fonologia. *Lapurdum*, 7, 201-219. doi: [10.4000/lapurdum.1088](https://doi.org/10.4000/lapurdum.1088)
- Gaminde, I. (2006). Dardakarien ezaugarri akustikoez. *Euskalingua*, 8, 75-82.
- Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*, Blackwell, Oxford. doi: [10.2307/417875](https://doi.org/10.2307/417875)
- Trask, R. L. & Millar, R. M. (1996). *Trask's Historical Linguistics*. Hodder, Arnold
- Txillardeg. (1978). Fonetika baturuntz. *Euskera*, 24, 683-690.

LARRABETZUKO EUSKALDUN BERRIEN SEME-ALABA EUSKALDUN ZAHARREN EZAUGARRI PROSODIKOEZ ¹⁹

Naia Eguskiza, Iñaki Gaminde, Ander Olalde, Aintzane Etxebarria eta Ursua Gaminde

Euskal Herriko Unibertsitatea

naia.eguskiza@ehu.eus, inaki.gaminde@ehu.eus, ander.olalde@ehu.eus,
aintzane.etxebarria@ehu.eus, ursua.gaminde@gmail.com

Laburpena

Euskaldun berriak ugarituz etorri diren heinean, seme-alabei euskararen barietate berria transmititu diete eta gazte horiek egiten duten barietatea da hemen aztertu nahi dena. Bestalde, Larrabetzu aukeratu da herri euskalduna delako eta barietate tradizionalekin batera barietate berri hauek nahasten direlako. Beraz, alde batetik, talde honetako hiztunek euskararen barietate berria daukate ama hizkuntzatzat, baina, beste alde batetik, bizi direneko ingurune sozialaren eraginez, bertako barietateen eragina daukate, txikitatik bertako gazte euskaldun zaharrekin batera bizi direlako. Gure aurreko lan batean (Gaminde et al, 2016b) azentu ereduaren eta oinarritzko esaldien intonazioaren aldakortasuna hartu genuen aztergaitzat informatzaileen belaunaldien eta generoaren arabera. Oraingo honetan talde honetako informatzaileen azentu ereduaren eta oinarritzko esaldien intonazioaren ezaugarriak aztertuko ditugu eta adin bereko barietate tradizionalen hiztunen gauzapenekin erkatuko ditugu. Hortaz, lan honen helburu nagusia da deskribatzea talde honetako hiztunen azentu eta intonazio ereduak eta erkatzea bertako barietate tradizionalekin, jakiteko bateratzen diren edo talde homogeneo eta bereiz bi osatzen duten. Lana egin ahal izateko, hiztun talde bakoitzeko launa informatzaileen datuak erabili dira.

Gako-hitzak: hizkuntza barietazioa, prosodia, euskara, Larrabetzuko barietatea

1. Sarrera

Euskal hiztunen tipologia aspaldiko urteekin aldatuz doana gauza jakina da. Orain dela 50 bat urte euskal hiztun mota bi baino ez zeuden; euskaldun zaharrak izenekoak eta euskaldun berriren bat edo beste. Lehenengoek familia transmisioaren bidez ikasten zuten barietate tradizionalen bat eta gero bakarren bat alfabetatu ere egiten zen euskaraz. Euskaldun berriek euskaldun zaharrekin zituzten erlazioen bidez ikasten zuten euskara, ez gaur egun ezagutzen ditugun eredu formalen bidez; ikasten zutena barietate tradizionala zen.

Gerora euskararen berpizkundearen hasieran euskara ikasteko eredu formalak ezarri ziren eta euskaldun berriak izenekoak euskararekin jabetzen ziren hasieran gau eskoletan, euskaltegietan geroago eta antzeko beste ikastegi batzuetan eta eskoletan oraintsuago. Euskaldun hauetako gehienek, salbuespenak salbuespen, ikasi duten euskararen barietatea edo aldaera berritatzat jo behar da. Talde hau, jakina, ez da homogeneoa ez hizkuntzaren ezagutzaren ikuspuntutik ez eta hizkuntzaren erabilerarenetik ere. Gainera, ezin uka daiteke euron inguruneen egoera soziolinguistikoa baldintzatzen dituzten ikasketa, jabetze prozesua eta jarduna.

Edozelan ere, urteak joan urteak etorri, talde honetako euskaldun berri batzuek seme-alabak eduki dituztenean, sarri askotan izugarritzko ahalegina bitarteko, euskara transmititu diete eta orain gazte horiek, gehienak gazteak direlako, euskaldun zaharrak

¹⁹ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

dira, familia transmisioaren bidez euskara lehen hizkuntzat jaso dutelako; hala ere, jaso duten barietatea ez da barietate tradizionala izan, beste bat baino.

Euskara batuaren eta estandarren sorreran eta ondorengo urteetan zehar izan duten bilakaera ere kontuan hartu beharko litzateke prozesu honetan (Maia, 2001; Amorrortu, 2003; Zuazo, 2005); nola ulertzen zen euskara batua abian paratu zenean, geroko bilakaeran eta gaur zer den. Ez dakigu orduko sasoian estandarizazio prozesua martxan ipini zutenek aurre ikusten zuten barietate hori hitz egiteko erabiliko zenik eta are gutxiago lehen hizkuntza gisa transmitituko zenik familia giroan.

Baina gatozen darabilgun gaiaren harira eta zilegi izan bekigu galdera bat plazaratzea: Euskaldun berrien seme-alaba euskaldun zaharrek ba ote daukate aldaera homogeneorik, Euskaltzaindiak eman eta ematen dituen arau guztiengatik ere?

Lan xume honen bidez ezin erantzun diezaiokegu galdera honi zehaztasun osoz, baina bai ohartarazi ikerketa lerro hau abiarazteko garrantziaz; izan ere, gerora talde honek soziologikoki pisua hartuko du gizartean ezberrik gabe.

Bestalde, eta hipotesi gisa besterik ez bada ere, gure intuizioetan oinarrituta, plazaratu dezakegu erratzeko beldur handi barik ezen erantzuna ezezkoa izango dela. Talde honetan ere, aldakortasuna egongo da ezinbestez, zeren, euskaldun berriekin irudikatu dugun bezala, hala euren inguruneen egoera soziolinguistikoak oso ñabarrak eta heterogeneoak baitira. Aldez aurretik, pentsa daiteke ezin ber ere linguistikoak erabiliko dituztela Larrabetzu bezalako herri euskaldun batean bizi direnek edo ingurunean erdara baizen ez daukatenek. Guk honako aukeratu dugun kasu honetan hitzun hauek barietate tradizionala jaso duten hitzuneekin batera bizi izan dira euren denbora osoan; talde bietako hitzunik elkarrekin egon dira txikitatik, elkarrekin olgatu dute, elkarrekin eskola berean egon dira eta kasu batzuetan lagun talde bereko kideak dira. Gauzak honela, pentsa daiteke elkarren ezaugarri linguistiko komunak agertuko direla eta taldeen banaketa lausotu eta ez dela agerikoa izango, etxean jasotako barietateak batzuk izanagatik ere. Nolanahi ere den, datuak eta ikerketa etorri beharko dira gure hau ezesten edo baiesten.

Behin honezkero badauzkagu Larrabetzuko euskara tradizionalaren bilakaeraren gainean egineko lan batzuk (Etxebarria et al., 2016; Gaminde et al. 2015, 2016a), baita honetarako aukeratu ditugun gaien gaineko beste bat ere (Gaminde et al., 2016b). Bertan adierazi bezala, Larrabetzuko azentuan eta oinarritzko esaldien intonazio ereduetan belaunaldien arteko aldeak nabarmenak dira, zein, bestalde, gertakari arruntzat hartu behar baita, hizkuntzen dinamikan.

Honenbeste, lan honen helburua nagusia da, Larrabetzuko euskaldun berrien seme-alaba euskaldun zaharren azentu-ereduaren eta oinarritzko esaldien intonazioaren ezaugarri nagusiak deskribatzea eta bertako adin bereko barietate tradizionala ama hizkuntzat jaso duten hitzunen ezaugarriekin erkatzea. Egiten dihardugun lan sakon baten behin-behineko emaitzen arabera (Gaminde et al., 2017), arlo batzuetan ez dago talde hauen arteko alde nabarmenik. Hemen aurkezten ditugun gaiok aukeratu ditugu aurreko lanean ezarritako lan ildo berean kokatzen direlako.

2. Corpora eta metodologia

Aurkezten dugun azterketa burutzeko erabili dugun corpora osatzeko zortzi informatzaile izan ditugu; lau euskaldun berrien seme-alaba euskaldun zaharren taldekoak (C taldea hemendik gora) eta beste lau barietate tradizionala jaso dutenen taldekoak (A taldea hemendik aurrera). Informatzaile guztiak 1991-2000 urte artean jaiokoak izan dira eta talde bakoitzeko bi neska izan dira eta bi mutilak.

Datuak jasotzeko galdekizun zehatza egin da 22 galderarekin; 15 azentuari dagozkio eta 7 oinarrizko esaldien intonazioari.

Galdekizunak Marantz PMD610 izeneko grabagailuarekin eta kanpo mikrofono batekin grabatu dira. Grabaketak “praat” (Boersma eta Weenink, 2016) izeneko programaren bidez transkribatu eta etiketatu 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.

3. Azentu eredua

Larrabetzuko azentu-sistema eta beronen bilakaera aurreko lan batean aztertu genuen (Gaminde et al., 2016b) eta azentu-sistemaren arau orokorrak ezartzeko irizpideak ere eman ziren bertan (Gaminde, 1998, 2010, 2011). Irizpide horiek erabilita Larrabetzuko zaharren euskararen azentu sistemaren tasunak ondoko era honetara laburbil daitezke:

- (1) Azentu bereizgarria
- (2) Eremua talde klitikoa
- (3) Arau orokora: Azken silaba estrametrikala
Arau nagusia [2]

ó-o: súrre, txárra, úre, óna, e.a.

o-ó-o: adárra, adúrre, andíe, aldátza, e.a.

o-ó-o-o: abádea, abératza, afárie, alábia, e.a.

Berba markatuak azentu sistema honetatik kanpo geratzen direnak dira; hiru silabako hitzen kasuan lehen silaban azentuatzen direnak: “*ábie*” (habia), “*átzia*” (atzea), “*lória*” (lorea), “*órmia*” (horma), “*sútia*” (sukaldea), e.a. Lau silabako berben kasuan markatuek azentua lehen silaban eduki dezakete: “*békokie*” (bekokia), “*béskarie*” (bazkaria), “*dénporea*” (denbora), “*léngusue*” (lehengusua), e.a. Baina badira hirugarren silaban azentua txertatzen dutenak ere: “*biribile*” (biribila), “*errosóya*” (arrazoia), “*martitzéna*” (asteartea), “*txakoline*” (txakolina), “*karakóla*” (barraskiloa), e.a.

Azentu bereizgarriaren erabilera aztertzeko, aurreko lanean legez, singularrak eta pluralak bereizten diren ikusi dugu, galdera multzo biren bidez; batean sintagma baten singularra eta plurala bereizten diren ikusiko da eta bestean erakusleen singularra eta

plurala. Aukeratu den sintagma “*lagunena*” izan da (berau “*lagunarena*” eta “*lagunena*” izan daiteke azentuatutako silabaren arabera informatzaile batzuen erabileran); erakusleak “*oneri*” (honi) eta “*oneri*” (hauei) izan dira.

Larrabetzun bereizkuntza egiten duten edadeko informatzaileek tonuaren bidez egiten dute, izan ere, bai singularrean bai pluralean azentua silaba berean txertatu arren, singularrean L*+H azentu tonuduna txertatzen da eta pluralean H*+L. Honek esan nahi du L*+H txertatzen denean sintagmaren gailurra azentuaren osteko silaban kokatzen dela eta H*+L txertatzen denean, berriz, silaba azentudunean (Gaminde et al. 2016b).

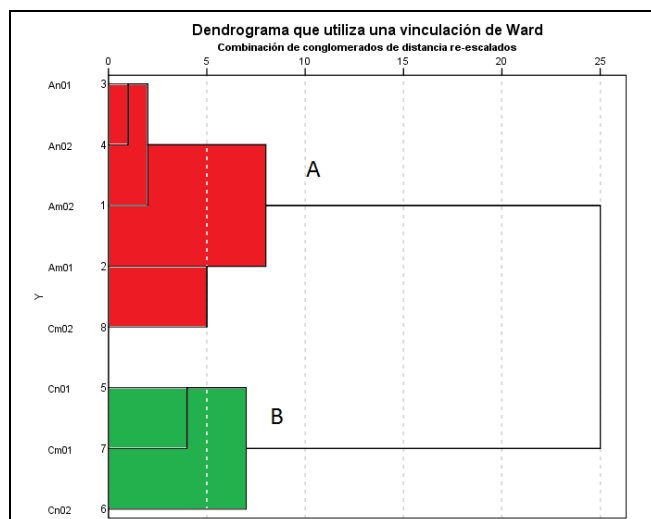
Gure oraingo informatzaile guztiek, A taldekoek zein C taldekoek, sintagman bereizkuntza galdu dute. Erakusleen bereizkuntza A taldekoek baino ez dute egiten.

Aurreko lan batean (Gaminde, 1998) azentuaren eremutat talde klitikoa jotzen zen; honen arabera isolatuki edo talde klitikoan silaba bakarreko erroak desberdin azentuatzen dira (1). Arazo hau ikertzeko galdekizunean galdera multzo bat sartu genuen aurreko lanean eta gazteen artean azentu eremua berba dela probatutzat utzi genuen arren, gure oraingo lanean ere eutsi diogu galdera multzo honi. Emaitzak aztertuta, esan behar da berba hartzen dela eremutat eta “*da*” adizki klitikoa gehituta azentuaren egongunea ez dela aldatzen; hau da “*txárra*” eta “*óna*” berbei “*da*” gehituta azentua ez da aldatzen: “*txárra da*” eta “*óna da*”. Honetan informatzaile guztiak bat datoz.

Berba markatuen jokabidea aztertzeko bederatzi berba aukeratu ditugu: “*sútia*” (sukaldea), “*lékue*” (lekua), “*sémia*” (semea), “*txístue*” (txistua), “*déndia*” (denda), “*intxurre*” (intxaurra), “*léngusue*” (lehengusua), “*ormia*” (horma) eta “*béskarie*” (bazkaria).

Hitz markatuetan dauden 72 aukeretatik markatu gisa egin dira 27 (% 37,5), markatuak galtzeko joera nabarmena da (Pascual, 2013, Gaminde et al., 2016b). “*beskarie*” eta “*lengusue*” berbak behin ere ez dira markatu gisa erabili. Gainerakoak kopuru batzuekin agertzen dira: “*txístue*” (6, % 75), “*sutia*” (6, % 75), “*ormia*” (4, % 50), “*lekue*”, “*dendia*” eta “*intxurre*” (3, % 37,5) eta “*semia*” (2, % 25). “*dendia*” eta “*ormia*” berben kasuan kontuan hartu behar da informatzaile batzuek “a” artikulua eranstea “*dénda*” eta “*órma*” egiten dutela eta orduan hitzak ez dira markatuak, ezpadaze arau orokorraren arabera azentuatuek.

Ikusi berri ditugun tasun hauen guztien probabilitateak kalkulatuta sortu dugu distantzia matrizea informatzaileak sailkatzeko. Sailkapena multzokatze azterketaren bidez egin dugu distantzia euklidear karratua eta Ward metodoa erabilita. 1. irudiko dendograman ikusten den bezala, nahikoa homogeenok diren multzo bi sortzen dira. A multzoan A taldeko informatzaile guztiak kokatzen dira C taldeko informatzaile batekin eta B multzoan gainerako C taldeko informatzaileak kokatzen dira.

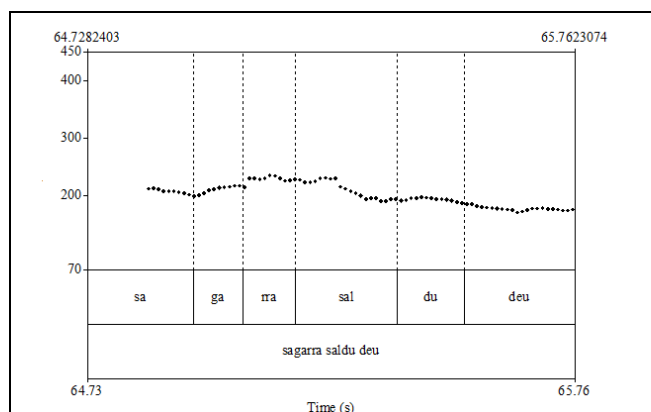


1. irudia: Informatzaileen sailkapena azentu-sistemaren ezaugarrien arabera.

4. Oinarrizko esaldien intonazioa

Esaldien intonazio ereduak aztertzeko eta erkatzeko aukeratu ditugun esaldiak zazpi izan dira: adierazpen esaldiak (sintagma bat eta birekin), bai/ez galderak, n-z galderak, hautazko galderak, oihartzun-galderak eta zalantzako galderak. Hemen aurkezten dugun azterketarako esaldien f_0 -ren azterketa fonetikoan oinarritzen den eredu erabili dugu (Gaminde eta Gandarias, 2015b). Atalaren amaieran informatzaileen sailkapena egingo dugu esaldi moten emaitzen arabera.

Sintagma bakarreko adierazpen-esaldietan mota bi baino ez zaizkigu agertu. Bietan esaldiaren gailurra sintagman kokatzen da, aldea lehen sintagmaren gailurra kokatzen deneko silaban gertatzen da, kasu gehienetan hirugarren silaban agertzen da gailurra (% 88,89), hau da, silaba azentudunaren ostekoan eta kasu bakar batean baino ez zaigu silaba azentudunean bertan agertu (% 11,11). Azken erabilera hau C taldeko mutil batena izan da. Aditzaren gailurra bigarren silaban txertatzen da kasu gehienetan (% 88,89), salbuespen bakarra C taldeko mutil batena da, beronek hirugarren silaban txertatzen du eta. 2. irudian f_0 -ren kurbaren adibide bat erakusten da.



2. irudia: Adierazpenezko esaldi baten f_0 -ren kurba.

Adierazpen-esaldietan sintagma bi daudenean esaldi osoaren gailurra lehen sintagman kokatzen da kasu guztietan; esaldien arteko aldeak sintagmen eta aditzaren gailurren kokaguneetan agertzen dira. Lehen sintagmaren gailurra (1. sin. gai.) kasu guztietan agertzen da hirugarren silaban, bigarren sintagmaren gailurraren (2. sin. gai.)

kokaguneari dagokionez, aukera bi daude, bigarren silaban (5, % 62,5) eta hirugarren silaban (3, % 37,5). Azkenik, aditzaren gailurra (Adi. gai.) bigarren silaban kokatzen da kasu gehienetan (6, % 75) eta hirugarren silaban kasu bitan (% 25). 6. irudiko sonogramaren f0ren kurbaren adibide bat erakusten da. Ezaugarri hauekin matrize bat osatzen badugu, 1. taulan erakusten diren bektoreak lortzen dira, bertan bakoitzaren agerpen kopuruak eta euren portzentajeak adierazten dira.

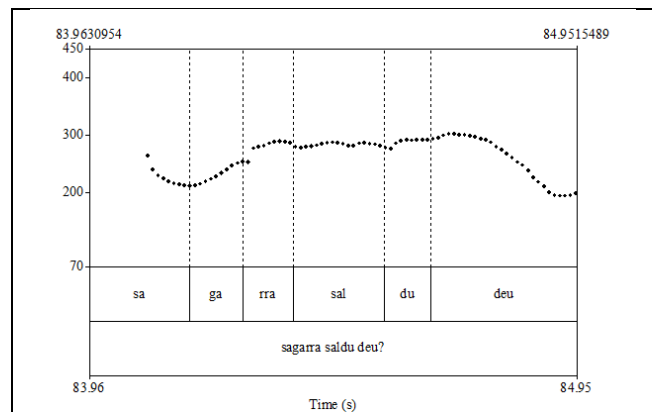
Bektorea	1. sin. gai.	2. sin. gai.	Adi. gai.	kopurua	%
1	3	2	2	4	50
2	3	2	3	1	12,5
3	3	3	2	2	25
4	3	3	3	1	12,5

1. taula: Ezaugarrien bidez lortutako bektoreak eta euren kopuruak eta portzentajeak.

Bektoreen erabilera informatzaileen ama-hizkuntzaren arabera aztertuz gero, ikusten dugu A taldea homogeneoagoa dela besteak baino, horrela bada, A taldeko informatzaile guztiek 1. bektorea erabiltzen dute eta C taldekoek, berriz, beste hirurak.

Bai/ez galderetan kontuan hartu ditugun ezaugarriak sintagmaren gailurra txertatzen deneko silaba eta esaldiaren muga-tonua izan dira.

Gure informatzaile guztiek sintagmaren gailurra beronen azken silaban kokatzen dute. Muga-tonu bi baino ez zaizkigu agertu: HL (7, % 87,5) eta L (1, % 9,09). Azken muga-tonu hau A taldeko neska batek erabili du. Gehien erabiltzen den HL muga-tonua 3. irudian erakusten da.

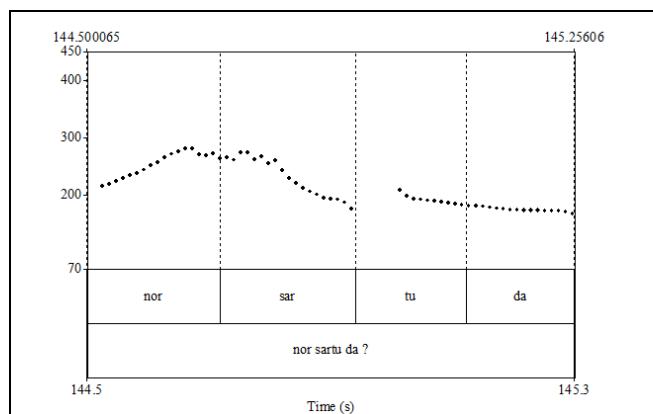


3. irudia: Bai/ez galdera baten f0ren kurba, amaieran HL muga-tonua daukala.

N/Z galderetan muga-tonuak hartu ditugu aztergaitzat. Gure informatzaileetan hiru muga-tonu agertu zaizkigu: L (5, % 62,5), LH (2, % 25) eta H (1, % 12,5). Muga-tonu kopuruak erakusten dira 2. taulan informatzaileen ama hizkuntzaren arabera sailkatuta; era berean, 3. irudian gehien erabiltzen den L muga-tonua ikus daiteke.

AH.	L	LH	H
A	2	2	0
C	3	0	1

2. taula: Muga-tonuen kopuruak informatzaileen ama-hizkuntzaren arabera.

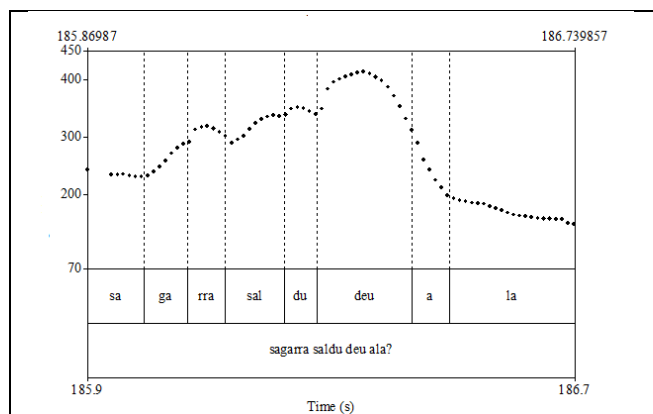


4. irudia: N/Z galdera baten f0ren kurba, amaieran L muga-tonua daukala.

Hautazko galderak egiteko informatzaileek “ala” (4, % 50) erabiltzen dute eta “edo ser” (4, % 50). Sintagmaren gailurra hirugarren silaban agertzen da kasu guztietan; aditzaren gailurra, oster, hirugarrenean egon daiteke (4, % 50) edo bigarreanean (4, % 50). Esaldi osoaren gailurra kasu guztietan aditzean kokatzen da (5. irudian “ala”rekin egineko esaldi baten f0ren kurba erakusten da). Azkenik, muga-tonu bi agertu zaizkigu: L (7, % 87,5) eta LH (1, % 12,5). Datu hauen bidez matrize bat sortzen badugu, 3. taulako bektoreak lortzen dira.

	Morfema	Adi. gai.	M-T	kopurua	%
1	ala	2	L	3	37,5
2	ala	3	LH	1	12,5
3	edo ser	3	L	3	37,5
4	edo ser	2	L	1	12,5

3. taula: Ezaugarrien bidez lortutako bektoreak eta euron kopuruak eta portzentajeak.



5. irudia: Hautazko galdera baten f0ren kurba, amaieran L muga-tonua daukala.

Bektoreen erabilera informatzaileen ama hizkuntzaren arabera aztertuz gero (4. taula), ikusten dugu C taldeko informatzaileek erabiltzen dutela bektore kopuru handiagoa A taldekoek baino.

AH.	1	2	3	4
A	3	0	1	0
C	0	1	2	1

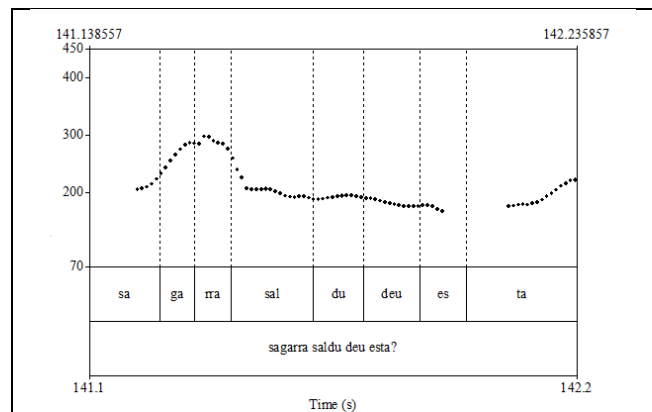
4. taula: Bektoreen kopuruak informatzaileen ama-hizkuntzaren arabera.

Oihartzun galderak “*esta*” edo “*es*” erabilia egin daitezke. Galdera mota hauetan kontuan hartuko ditugun ezaugarriak hauek dira: “*esta*” ala “*es*” erabiltzen den; sintagmaren eta aditzaren gailurraren silaba; esaldiaren gailurra sintagman ala aditzean kokatzen den; eta azkenik, esaldiaren muga-tonuak.

Informatzaileen artean “*esta*” da gehien erabiltzen dena (7, % 87,5) eta “*es*” gutxien (1, % 12,5). Esaldiaren gailurra kasu guztietan sintagman kokatzen da eta berau beti izaten da sintagmaren azken silaba (6. irudian f0ren kurba bat erakusten da). Aditzaren gailurra bigarren silaba izan daiteke (7, % 87,5) edo hirugarrena (1, % 12,25). Muga-tonua H izan daiteke (6, % 75) edo LH (2, % 25). Aldatzen diren ezaugarrien datuak matrizean sartuta 5. taulan erakusten diren bektoreak lortzen dira.

Bektoreak	Morfema	Adi. gai.	M-T	kopurua	%
1	esta	2	H	4	50
2	esta	2	LH	2	25
3	esta	3	H	1	12,5
6	es	2	H	1	12,5

5. taula: Ezaugarrien bidez lortutako bektoreak eta euron kopuruak eta portzentajeak.



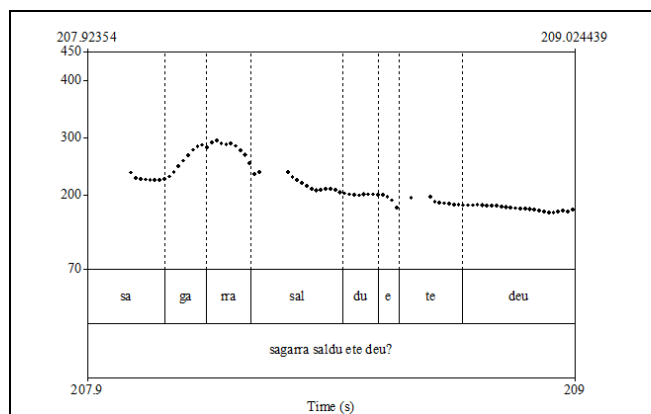
6. irudia: Oihartzun galdera baten f0ren kurba, amaieran H muga-tonua daukala.

Bektoreen erabilera informatzaileen ama hizkuntzaren arabera aztertuz gero (6. taula), ikusten dugu A taldeko informatzaileek erabiltzen dutela bektore kopuru handiagoa C taldekoek baino.

AH	1	2	3	4
A	1	2	0	1
C	3	0	1	0

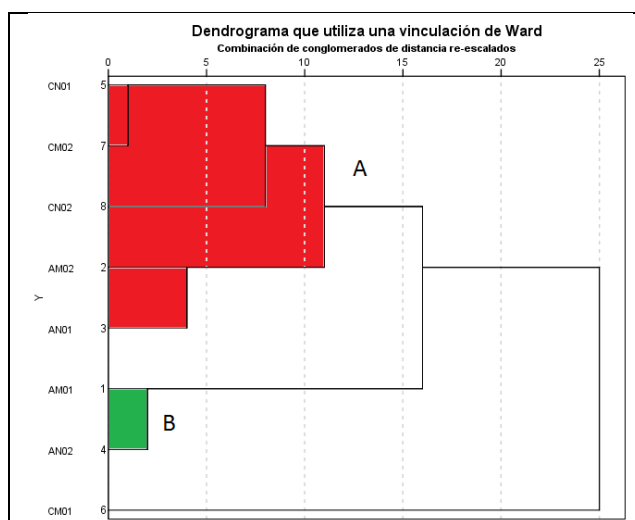
6. taula: Bektoreen kopuruak informatzaileen ama-hizkuntzaren arabera.

Zalantzako galderak “*ete*” (ote) partikularen bidez egiten direnak dira. Gure informatzaileetan hiruk baino ez dute erabili (% 37,5) mota hau; hirurak A taldekoak izan dira eta bi neskak eta bat mutila. Argi dago kasu honetan informatzaileen ama-hizkuntzak eragin zuzena daukana beronen erabileran. Bestalde, informatzaile horietako bik HL muga-tonua erabili dute eta batek L. L muga-tonua erabili duen informatzaileak aditzaren gailurra lehen silaban erabili du (7. irudiko sonograma) eta beste biek bigarrenean.



7. irudia: Zalantzako galdera baten f_0 ren kurba, amaieran L muga-tonua daukala.

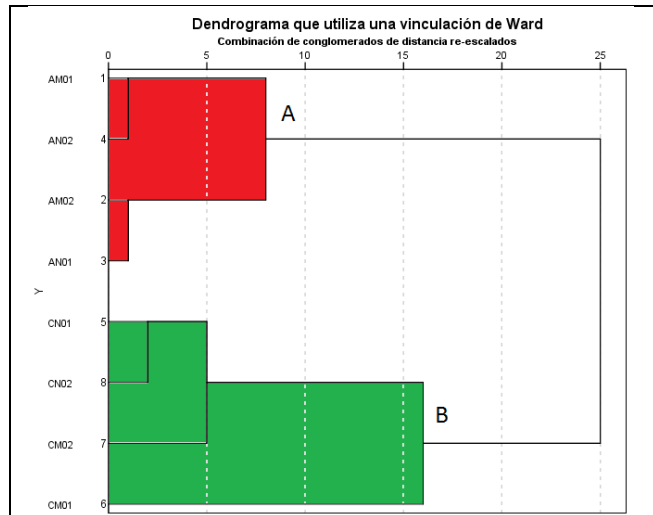
Azkenik, esaldi mota bakoitzeko aztertu ditugun ezaugarri guztien emaitzen arabera multzokatze azterketa egiten badugu, 8. irudiko dendograma lortzen dugu. Bertan agertzen den bezala, A multzoan 5 (% 62,5) informatzaile kokatzen dira eta B multzoan 2 (% 25), hortik kanpo informatzaile bat kokatzen da (% 12,5). Dendogramaren multzoak informatzaileen ama-hizkuntzaren arabera aztertuta, multzo homogeneorik ez dagoela esan dezakegu..



8. irudia: Informatzaileen sailkapena intonazioaren tasunen arabera.

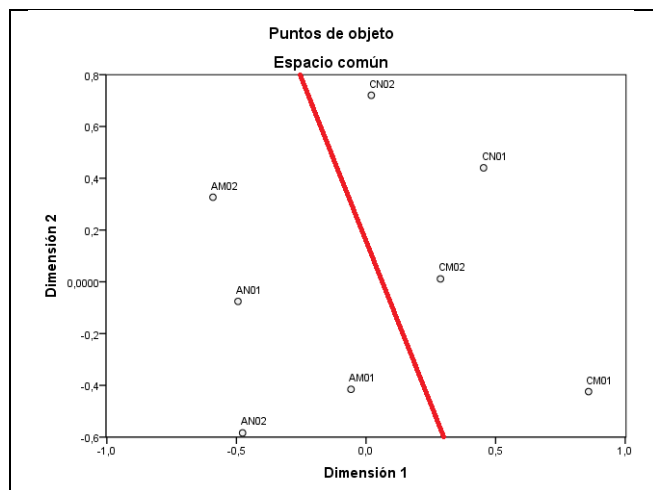
5. Informatzaileen sailkapena

Atal honetan informatzaileen sailkapena egingo dugu teknika bi erabilita. Alde batetik, multzokatze-azterketa erabiliko dugu eta horretarako distantziak kalkulatu ahal izateko aurreko atal bietan lortu ditugun emaitzekin matrizea osatu dugu lehenengo eta behin. Multzokatze azterketa eginda 9. irudian erakusten den dendograma lortzen dugu. Bertan ikus daitekeen bezala, ezaugarri prosodiko guztien arabera sailkapena egiten dugunean, A taldeko informatzaileak multzo batean kokatzen dira eta C multzokoak, berriz, bestean.



9. irudia: Informatzaileen sailkapena azentu sistemaren eta oinarritzko esaldien tasun guztien arabera.

Gauza bera egiten badugu, sailkatzeko erabiltzen den dimentsio aniztun eskala izeneko teknikarekin, 10. irudian ikus daitekeen irudia lortzen da. Bertan argi ikus daitekeen moduan, talde bietako informatzaileak ez dira nahastatzen espazioan.



10. irudia: Informatzaileen sailkapena tasun guztien arabera, dimentsio aniztun eskala probaren arabera.

6. Ondorioak

Aurreko atal bietan ikusi dugun bezala, hemen aztergaitzat hartu ditugun prosodiaren arlo biotako emaitzak, informatzaileen sailkapenari dagokionez, banan aztertuta oso desberdinak izan dira; hau da, azentu-sistemaren ezaugarrien arabera sailkapena egiten dugunean, A taldeko informatzaileak bereiz agertzen dira C taldeko informatzaile batekin eta, sailkapena oinarritzko esaldien ezaugarrien arabera egiten dugunean, informatzaile guztiak nahiko nahasirik agertzen dira multzo argirik agerian ez dagoela.

Alta, arlo biotako ezaugarri guztiak kontuan hartzen ditugunean, eta informatzaileen sailkapena denon arabera egiten denean, banaka agertzen ez diren emaitzak azalartzen dira eta orduan multzo nagusi bi bereiz agertzen dira ondo zedarriturik. Batean A taldeko informatzaileak agertzen dira eta bestean C taldekoak.

Hemendik ondorio nagusi bi atera ditzakegu. Batetik, konplexutasunaren teoriaren ikuspuntutik (Beckner et al., 2011; Ellis, 2011; López Rivera, 2013; Moreno Cabrera, 2008; Léonard et al., 2015), hemen argi agertzen zaigu sistema gehiago dela bere osagaien batura baino eta modelizazio linguistikoaren eta matematikoaren ondorioz azpian dautzan egiturak azaleratzen direla prozesuaren bidez, hots, sistema honek balio diezaguke begi bistan ez dagoena bistaratzeko.

Bestetik, datu hauen argitara esan dezakegu ezen euskaldun berrien seme-alaba euskaldun zaharrak Larrabetzun gainerako euskaldunekin bizi diren arren, oraindik ez direla jabetu prosodiaren arlo biotako A taldeko gainerako euskaldunen ezaugarriekin edo, gurago bada, talde biak ondo mugatuta agertzen direla.

7. Bibliografia

- Amorrortu, E. (2003). *Basque Sociolinguistics: Language, Society, and Culture*. Center for Basque Studies, University of Nevada. Reno.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2011). *La lengua es un sistema adaptativo complejo Lingüística en la Red*. http://www.linred.es/articulos_pdf/LR_articulo_04092011.pdf
- Boersma, P., & Weenink, D. (2016) *Praat: doing phonetics by computer, Version 5.1*. [Computer program]. <<http://www.praat.org>>
- Ellis, N. C. (2011). The Emergence of Language as a Complex Adaptive System. In James Simpson (Ed.), *Routledge Handbook of Applied Linguistics*. London: Routledge/Taylor Francis.
- Etxebarria, A., Gaminde, I., Olalde, A., & Gaminde, U. (2016). Hizkuntza aldakortasuna Larrabetzuko aditz morfologian. In Iglesias, A.; Romero, A. eta Ensunza, A. *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. Bilbo: UPV/EHU.
- Gaminde, I. (1998). *Euskaldunen azentuak*. Bilbo: Labayru.
- Gaminde, I. (2010). *Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I. (2011). Noraka Euskal Azentuak?. In Aurrekoetxea, G. eta Gaminde, I. (edi.) *Prosodiaz eta Hezkuntzaz I. Jardunaldiak/I. Jornadas sobre Prosodia y Educación*. Leioa: UPV-EHU
- Gaminde, I., & Gandarias, L. (2015a). Irizpide batzuk bariazioa aztertzeke intonazioan. In Aurrekoetxea, G.; Romero, A. eta Etxebarria, A. *Linguistic Variation in the Basque an Education-I/Euskararen Bariazioa eta Bariazioaren Irakaskuntza-I* (183-205 or.) Leioa: UPV/EHU
- Gaminde, I., Gaminde, U., Olalde, A., & Etxebarria, A. (2015b). Larrabetzuko testuen prosodiaz. *Euskalingua*, 27, 6-12
- Gaminde, I., Olalde, A., Gaminde, U., & Etxebarria, A. (2016a). Populazioaren sailkapena lexikoaren arabera. *IKASTORRATZA. e-Revista de Didáctica*, 16, 68-84. Retrieved from http://www.ehu.es/ikastorratza/16_alea/4.pdf
- Gaminde, I., Olalde, A., Etxebarria, A., & Gaminde, U. (2016b). Hizkuntza aldakortasuna Larrabetzuko azentuan eta intonazioan. *Euskalingua*, 27, 13-21.
- Gaminde, I., Olalde, A., Etxebarria, A., Eguskiza, N., & Gaminde, U. (2017). *Hizkuntza Aldakortasuna Larrabetzun* [agertzeko].

- Léonard, J. L., Heinsalu, E., Patriarca, M., & 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.) Leioa. 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.
- Maia, J. (2001). Eskola eta tokian tokiko hizkuntz aldaera. In Mendebalde Kultura Alkartea: *Euskalkia eta hezkuntza*. Bilbo.
- Moreno Cabrera, J. C. (2008). *El nacionalismo lingüístico: Una ideología destructiva*. Barcelona: Ediciones Península.
- Pascual, M. (2013). *Mendebaldeko Azentua: Larrabetzuko Hitz Markatuak*. (lan argitaragabea)
- Zuazo, K. (2005). *Euskara Batua: Ezina ekinez egina*. Donostia: Elkar.

PROZEDURA METODOLOGIKOEN KONPARAKETA HAUR ELEBIDUNEN HASIERAKO GARAPEN LEXIKOAREN INFORMAZIOAREN LORPENERAKO²⁰

Irati de Pablo, Asier Romero, Aintzane Etxebarria eta Iñaki Gaminde
Euskal Herriko Unibertsitatea

irati.depablo@ehu.eus, a.romero@ehu.eus, aintzane.etxebarria@ehu.eus,
inaki.gaminde@ehu.eus

Laburpena

Ikerketa honen helburua da haurren garapen lexiko goiztiarra aztertzeke erabiltzen diren prozedura metodologiko ezberdinak arakatzea. Horretarako, lagin berarekin bi prozedura metodologiko desberdin aplikatu ziren: alde batetik, MacArthur-Bates inbentarioak bai euskararako bai gaztelaniarako sortutako bertsiokak eta, bestetik, luzetarako behaketa. Lagina 16 eta 30 hilabete arteko 8 haur euskaldunek sortu dute. Haur kopuru horren erdiak euskara du ama-hizkuntza eta beste erdiak, berriz, gaztelania. Emaitzek inbentarioren balioa erakusten dute lexikoaren inguruko informazioa lortzeko, baina bere mugak ere badituzte, euren seme-alaben gehiegizko balorazioan oinarritzen direnak.

Gako-hitzak: Garapen linguistiko goiztiarra, MacArthur-Bates inbentarioa, lexikoa, hizkuntzaren ebaluazioa

1. Sarrera

Linguistika arloan diziplina zientifiko berrien agerpenak ikerkuntzaren perspektiba metodologikoan aldaketa inportantea eragin du. Kontzeptualizazio enpiriko honetan, linguistika azterlanak oinarritu behar dira datu errealetan, eta horrez gain, kontuan izan behar dute bariazio soziala, geografikoa, estilistikoa eta hitz egiteko era (Garfinkel, 1967; Goffman, 1981; Halliday & Hasan, 1985; Hymes & Gumperz, 1964; Labov, 2001). Datu errealak izateko beharizan honekin planteatzen da zein den informazioa eskuratzeko erarik eraginkorra eta hizkuntzaren garapena aztertu duten ikertzaileek erabili dituzten eretatik nabarmendu daitezke zeharkako eta luzetarako diseinuak. Baina autore batzuek proposatu izan dute bien konbinazioa, bakoitzaren puntu ahulak indartzeko (Baltes, Reese y Nesselroade, 1977), eta horri diseinu sekuentziala deitu zaio. Edozein diseinu izanda, pertsonaren portaera linguistiko espontaneo aztertzeke behaketa-metodoa erabil daiteke edo, aldagaiak manipulatu eta kontrolatzen badira, metodo esperimental jarri daiteke praktikan.

Behaketa metodoak eskaintzen digun tasunik onena da ikertzaileak ez duela zertan esku hartu, hau da, behatutako subjektua ezerk ez du mugatzen bere produkzio linguistikoan eta, horretaz gain, subjektuaren egunerokotasuneko egoera erreala eta testuinguru naturala aztertzen da. Baina, bestetik, ikertzailea egoerak ematen dionarekin konformatu behar da eta, ikertu nahi duen sistemaren lagin partziala eskuratzen du. Behaketaren barruan, testuinguruan kokaturiko behaketa nabarmendu daiteke (Romero, De Pablo, Etxebarria, & Romero, 2017a; Romero, Etxebarria, De Pablo, & Romero, 2017b). Honek bi oinarritzko elementu ditu haurren hizkuntzaren jabeakuntzaren ikerketan: (1) haurrekin segurtasun eta konfiantzazko giroa lortu behar da, hori dela eta, gomendagarria da ikertzailea eta haren instrumentazioa haurrentzat ezagun izatea naturaltasun orekatu batean integrazteke; eta (2) testuinguru naturalean barneratzeko

²⁰ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektuen barnean egin da.

estrategia garatu behar da. Horretarako, datuen konfidentzialtasuna mantendu behar da, ezarritako printzipio etikoak errespetatuz, indarrean dauden araudiak betez eta alde zuzeneko batzorde etikoaren aldeko txostena atxikiz. Bestalde, hizketakideak daude (familia kideak, tutoreak eta irakasleak), zeintzuk bitartekari funtzioa izango duten ikertzailea kanpoko pertsona arraro bezala ez ikusteko eta, horretaz gain, haurren inguruko informazioa emango digutenak behaketaren ikus-entzunezko erregistroa esanguratsuenaz izateko.

Bestetik, metodo esperimentalak jokaera konkretu bat eginarazi nahi du subjektuarengan eta, modu honetan, gaitasun linguistikoaren aspektu baten informazio zehatzagoa eskaintzen du. Ez da ahaztu behar metodo honek, datuak eskuratzeko moduari dagokionez, analisiaren egoeraren kontrol handia erabiltzen duela, planteatutako ikerketa hipotesiak egiaztatzeko. Hizkuntzaren jabeak metodo esperimentalak aplikatzeko zailtasuna da adin goiztiarreko subjektuen kolaborazioa lortzea; izan ere, motibazioa eta atenezioa beharrezkoak dira. Gainera, aurreko zailtasunaz gain, beste muga garrantzitsu bat ere gaineratzen da: testuinguruak planteatzen duen naturaltasun eskasa, horregatik ikertzaile askok zalantzan jartzen dute metodo horren balioa metodologikoa (Kotiliarenco & Mendez, 1988). Metodo esperimentalaren barruan, ikertzaileek erabiltzen dituzten prozeduretatik, MacArthur-Bates inbentarioak edo CDI eskalak daude (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick y Reilly, 1994). Hau gurasoek betetzen duten txosten egituratua da eta euren seme-alaben portaera linguistikoaren eta komunikatiboaren inguruko informazioa ematen du. Beraz, CDI eskalak beste instrumentu batzuekin konparatuz gero, zenbait abantaila dituzte: prozeduraren gastu txikia, lagin handien garapen komunikatiboaren ebaluazioa egin daiteke eta, hogeita hamar egokitzapen baino gehiago izateak konparaketa inter-linguistikoak egitea ahalbidetzen du (Pérez-Pereira y García Soto, 2003). Baina badira aspektu negatiboak ere, konkretuki informazioa lortzeko prozedurari dagokionez, askotan euren seme-alaben garapen linguistikoaren gain balorazioa ematen baita.

Orduan, ikerketa honen helburu nagusia da gurasoek betetako galdetegiaren datuak konparatzea komunikazio espontaneoan egindako grabaketan emaitzekin. Horrela, garapen linguistikoaren datuak erregistratzeko teknika metodologikoekin lortutako emaitzak konparatuko dira eta datuak interpretatuko dira. Gure hipotesia da lortutako puntuazioen korrelazioan bariazioa egongo dela MacArthur-Bates inbentarioen eta testuinguruan kokatutako behaketaren artean.

2.1. Metodoa

Atal honen edukiak hiru azpiataletan banatuta aurkeztuko dira. Lehenengo azpiatalean ikerketaren corpusa aurkeztuko da; bigarrenan, erabilitako materialak eta prozedura eta, hirugarrenan, analisirako erabilitako kodea azalduko da.

2.2. Parte-hartzaileak

Ikerketaren corpusa Bermeoko zortzi haur (4 neska eta 4 mutil) euskaldunen egintza komunikatiboek osatzen dute 17 hilabete dituztenetik 30 hilabete arte. Gurasoek grabaketak egiteko idatzizko baimena eman zuten. Ama-hizkuntzari dagokionez, zortzi haur horietatik lauk dute L1 euskara eta beste laurek L1 gaztelania eta guztiak daude euskararen murgiltze-programa baten barruan.

2.3. Materialak eta prozedura

Hiru grabaketa egin ziren hurrek 17, 25 eta 30 hilabete bete zituzten hilabetean. Egindako hiru sesioak bidez grabatu ziren subjektuen etxeetan behaketa metodologiaren bidez, konkretuki ikertzaileak behaketa ez-partehartzailea erabili zuen. Grabaketa bakoitzak 20-30 minutu inguru zituen, grabaketaren iraupena moldatzen zelarik egun bakoitzean hurrak zuen egoera emozionala eta komunikatiboari. Sesioetan ez zen ariketen antolaketa patroi zehatzik jarraitu, elkarrekintza dinamikoa eta, modu batean, librea bilatuz (Bosh, 2004; Fernández López, 2009). Grabaketak ZOOM Q4HD kamerarekin egin ziren Rode SmartLav kanpoko mikrofonoarekin, haurraren arropan jarrita.

Hurrak bidez grabatzeaz gain, MacArthur-Bates inbentarioen bertsio estandarizatuak edo CDI eskalak (Fenson et al., 1993; gaztelaniazko bertsioa, López-Ornat, Gallego, Gallo, Karousou, Mariscal eta Martínez, 2005; euskarazko bertsioa, García, Arratibel, Barreña eta Ezeizabarrena, 2008) bai gaztelaniazko bai euskarazko egokitzapenak erabili ziren. CDIaren gaztelaniazko egokitzapenak bi galdetegiz osatuta dago (1. CDIa 8 hilabetetik 15 hilabetetara; eta 2. CDIa 16 hilabetetik 30 hilabetera arte) eta euskarazko egokitzapenak hiru galdetegiz osatuta dago (1. CDIa 8 hilabetetik 15 hilabetetara; 2. CDIa 16 hilabetetik 30 hilabetera arte; eta 3. CDIa 30 hilabetetik 50 hilabetera arte). Konkretuki, ikerketa honetan 2. CDIaren ‘hiztegi zerrenda’ atala erabili zen eta gurasoei eman zitzaien betetzeko euren seme-alabek 30 hilabete bete zituztenean. Logikoki, egunerokotasunean gaztelania erabiltzen zuten gurasoei gaztelaniazko egokitzapena eman zitzaien eta euskara erabiltzen zutenei euskarazkoa. 15 egun pasa ondoren, betetako inbentarioak batu ziren eta puntuatu ziren emaitzen interpretazio eta analisisirako. López-Ornat et al. (2005) eta García et al. (2008) autoreei jarraiki, galdetegien emaitzak puntuazio bihurtu ziren, horrela emaitza egoki bakoitzari puntu bat egokitzen zitzaion. Puntuazio guztiak zenbatu ziren, minimoa gaztelaniazko bertsioan 0 puntu izanik eta 588 maximoa eta, euskarako bertsioan, berriz, minimoa 0 puntu eta maximoa 654 puntu izanik.

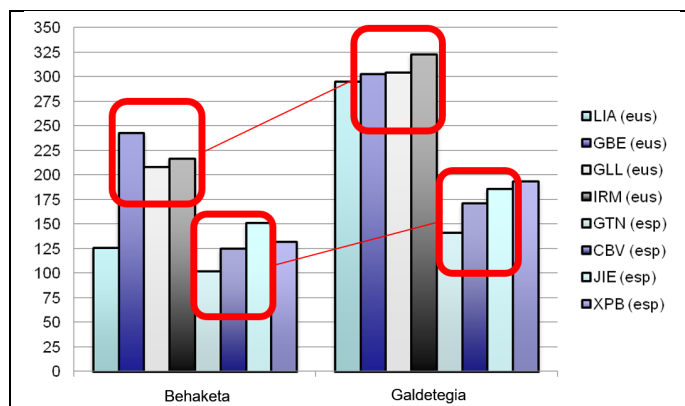
2.4. Kodifikazioa

Egintza komunikatibo guztiak identifikatu eta lokalizatu ziren grabaketetan, ELAN softwarea erabiliz (Lausberg eta Sloetjes, 2009). Egindako kodifikazioaren fidagarritasuna konprobatzeko, aurretik trebatutako bi ebaluatzailek grabaketen % 20 kodifikatu zuten eta horrekin Kappa indizea egin zen (Cohen, 1960), % 89ko akordioa lortuz ($k = 0,872$).

3. Emaitzak

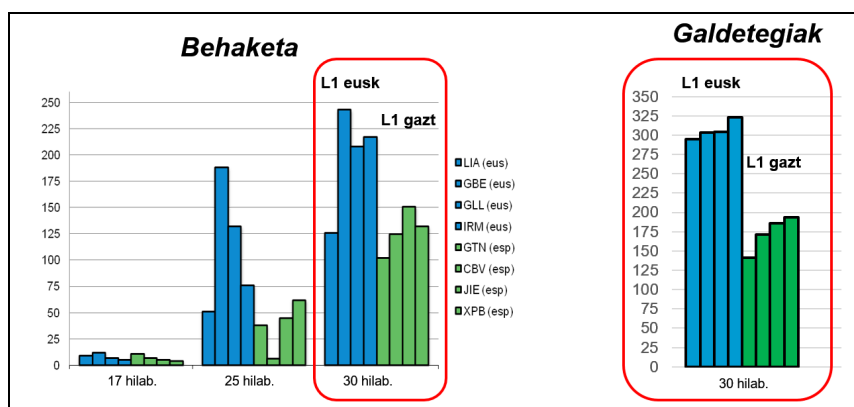
Ikerketa honetan haurren garapen lexikoa analizatu zen. Hasteko, haur bakoitzaren produkzioak bereizita aztertzen badira, 1. grafikoak erakusten duen moduan, GBE, GLL eta IRM hurrek 30 hilabete dituztenean, beste hurrek baino hiztegi zabalagoa daukate. Metodologia batean eta bestean lortutako datuak bat etorri ez arren, antzekotasun interesgarriak azaltzen dira produkzio lexikoari dagokionez, hau da, galdetegietan bai goitik zein azpitik nabarmentzen diren hurrak testuinguruan kokaturiko behaketetan ere berdin kokatzen dira, hau da, datu hauek proportzionalki errepikatzen dira behaketa eta galdetegietan lortutako puntuazioetan (estatistikoki ez da esanguratsua izan).

Prozedura metodologikoren konparaketa
 haur elebidunen hasierako garapen
 lexikoaren informazioaren lorpenerako



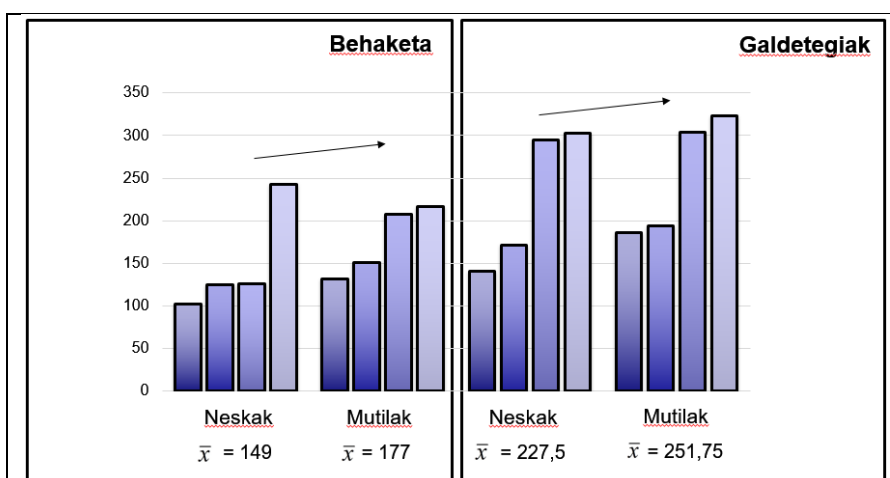
1. grafikoa. Haurren garapen lexikoa.

Baina datuak ama-hizkuntzaren arabera analizatzen badira (2. grafikoa), badira bariazioak: L1 euskara duten haurrek hitz gehiago produzitzen dituzte L1 gaztelania dutenekin alderatuz. Ezberdintasun hauek agerikoak dira, baina ez dira estatistikoki esanguratsuak.



2. grafikoa. Garapen lexikoa ama hizkuntzaren arabera.

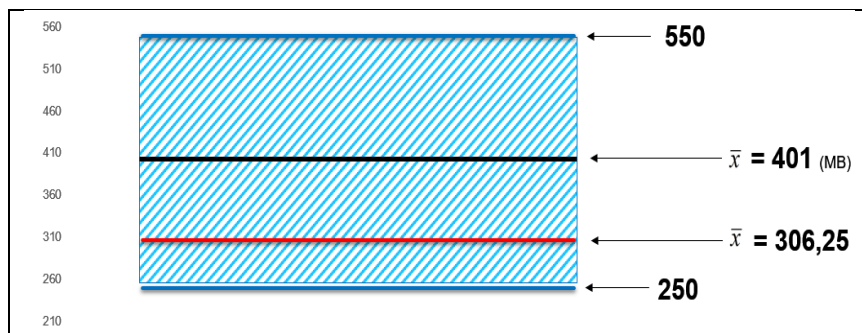
Sexuaren arabera analizatzen badira datuak, ikus daitezke mutilek neskek baino puntuazio altuagoak lortzen dituztela, kasu batean izan ezik.



3. grafikoa. Garapen lexikoa sexuaren arabera.

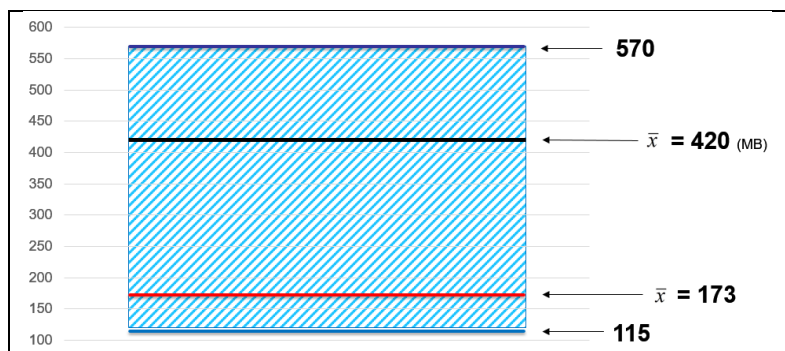
Ondoren, konparaketa bat burutu zen ikerketa honetan euskara inbentarioekin lortutako emaitzak Garcíak et al. (2008) emandako datuekin. Horrela, autore hauek

“normaltasunerako” proposatzen dituzten marjinak 250 eta 550 hitzen artekoak dira eta batezbestekoa 401 hitzetan da. Ikerketa honen kasuan, lortutako batezbestekoa 306,25 hitz dira, eta beraz, euren proposaturiko marjinen barruan sartuko litzateke, baina batezbestekoa askoz baxuagoa da.



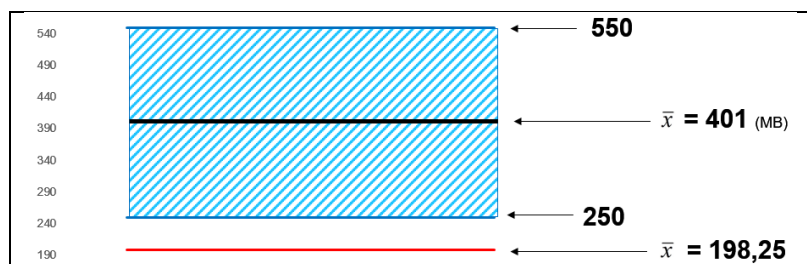
4. grafikoa. Ikerketa honetako euskarazko inbentarioekin lortutako emaitzen konparaketa García et al. (2008) autoreek hornitutako datuekin.

5. grafikoa ikerketa honetako gaztelaniazko inbentarioetako datuekin lortutako emaitzak, López-Ornat eta lagunek (2005) diotenarekin konparatu dira. Azken hauek ezarritako “normaltasun” marjinak 115 eta 570 dira, 420 batezbestekoarekin. Ikerketa honetan lortutako emaitzak parametro horien barruan sartzen dira, baina badago ezberdintasun handia bi batezbestekoen artean, ikerketa honetakoa askoz baxuagoa da eta.



5. grafikoa. Ikerketa honetako gaztelaniazko inbentarioekin lortutako emaitzen konparaketa López-Ornat et al. (2005) autoreek hornitutako datuekin.

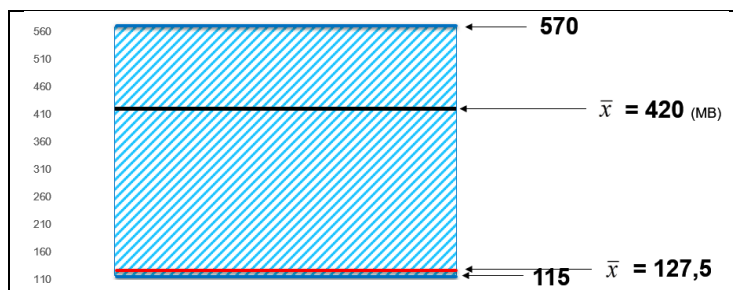
Konparatu zen beste aspektuetako bat izan zen MacArthur-Bates-en euskararako egokitzapenen emaitzak (García et al., 2008) eta ikerketa honetako behaketarekin eskuratutakoak. Kasu honetan, 6. grafikoa aztertu daitekeen bezala, produkzio lexikoaren batezbestekoa ikerketa honetan ez da sartzen García et al. (2008) ezarritako “normaltasun” parametro horietan, eta horrez gain, ezberdintasun esanguratsua dago minimoaren (250) eta horren artean (198,25).



6. grafikoa. MacArthur-Bates inbentarioen egokitzapena

euskarara (García et al., 2008) eta ikerkuntza honen behaketaren emaitzen arteko konparaketa.

Aurreko egoera bera gertatu zen gaztelanian. Horrela, estatu espainoleko gaztelania bertsioaren emaitzak (López-Ornat et al. 2005) konparatu ziren ikerlan honen emaitzekin, konkretuki behaketa-prozeduraren bitartez lortutakoekin. Kasu honetan, batezbestekoa López-Ornatek et al. (2005) proposatutako marjina maximo eta minimoaren artean dago, baina mugan.



7. grafikoa. Estatu espainoleko gaztelania bertsioaren (López-Ornat et al. 2005) eta ikerkuntza honetako behaketaren emaitzen konparaketa.

4. Eztabaida eta ondorioak

Metodologia ezberdin hauetan lortutako emaitzek konfirmatzen dute bai behaketen bai MacArthur-Bates galdetegiaren balioa produkzio lexikoaren azterketarako. Batean zein bestean lortutako datuak bat etorri ez arren, badira antzekotasun interesgarriak produkzio lexikoaren konputoan; hau da, behaketan produkzio komunikatiboan puntuazio altuena zein baxuena zuten hurrek galdetegietan ere berdin gertatzen zen proportzionalki. Logikoki, ezin da ukatu MacArthur-Bates inbentarioen fidagarritasuna, literatura zientifikoan prozedura eraginkor eta baliodun bezala ezaguna baita, baina lortutako datuak kontuan izanik, nolabaiteko gainbalorazioa ikusten da gurasoen beren seme-alaben emisioetan.

Datu horiek ama-hizkuntzaren eta sexuaren arabera aztertzen badira, desberdintasunak aurkitzen dira. Hasteko, euskara ama-hizkuntzat duten hurrek produkzio lexiko ugariagoa dute ama-hizkuntza gaztelania dutenek baino. Bigarrenik, mutilek hitz gehiago produzitzen zituzten neskek baino.

Bestetik, badira elementu batzuk bereizten direnak beste ikerketetan aurkitutakoekin. Horrela, produkzio lexikoaren datuak 16 eta 30 hilabeteko adinean, bai galdetegietakoak bai behaketakoak, ez datoz bat estatu espainoleko MacArthur-Bates gaztelaniazko bertsioarekin (López-Ornat et al., 2005) ez eta MacArthur-Bates euskarazko bertsioarekin (García et al., 2008). Ezberdintasun hauek izan daitezke laginen ezberdintasunengatik, galdetegiei erentzuterakoan egar daitezkeen desberdintasun kulturalengatik, grabaketetan izandako jokabide aldaketengatik edo haur elebidun bakoitzaren hizkuntzaren jabeakuntzan eragina duten prozesuengatik. Euskara-gaztelania elebitasunaren kasurako, zenbait ikerlarik bai hizkuntzaren esposizio maila eta bai hizkuntza baten eta bestearen ezaugarri tipologikoak ere aipatu dituzte (Almgren y Barreña, 2001; Barreña, 1995; Ezeizaberrena, 1996; Idiazabal, 1988). Ondorioz, analisi gehiago egin beharko liriteke, luzetarako datuetatik abiatuz eta lagin handiagoa izanik, haur elebidunetan hizkuntzaren jabeakuntza zelan gertatzen den jakiteko. Hizkuntzaren jabeakuntzaren inguruko beste ikerketa batzuetan lortutako datuak kontuan

izanik, argumentatu daiteke haur hauen hiztegia askoz urriagoa dela haur elebarrarekin konparatuz (Pearson, Fernández y Oller, 1993).

Emitza hauek kontuan izanik, inportantea da azpimarratzea L1 euskara duten haurrek eta L1 gaztelania dutenek emitza desberdinak dituztela, L1 ingelesa-gaztelania duten haurrek bezala. Bai galdetegien eta baita testuinguruan kokatutako behaketaren emaitzetan ere aztertu daiteke L1 euskara duten haurrek hizkuntzaren jabeakuntza prozesu arinagoa dutela L1 gaztelania dutenek baino. Hori dela eta, ondoriozta daiteke adin goiztiarrean bi hizkuntza aldi berean ikasten dituzten adin bereko haurrek, hizkuntzaren esposizio denboraren arabera, patroia desberdinak garatu ditzaketela. Egoera hau gerta daiteke hizkuntza nagusia euskara den gune soziolinguistikoetan ere, ikerketa honetan bezala. Edozein kasutan, ikerketa honetan erregistratutako konputoak ikerketa antzekoetan azaldu direnekin konparatuz, batezbestekoa baino baxuagoak dira.

Azkenik, etorkizunean analizatzea interesgarria izan daitekeen eta aipatu ez den aspektu bat azpimarratu nahi genuke, honekin erreferentzia egiten diogu errepertorio komunikatiboaren barietatea kontuan izateari. Tasa komunikatibo desberdinak kontuan izateaz gain, haur bakoitzak, adin bakoitzean, zenbat elementu eta konbinazio desberdin erabiltzen dituen eta nola egoera hau erlazionatzen den etorkizuneko hitz kopuruarekin analizatzea interesgarria litzateke.

Eskerrak

Eskerrak eman nahi dizkiegu ikerketa honetan parte hartu duten familia eta haurrei, eta Bermeoko Eleizalde Ikastolari haren laguntza eta kolaborazioagatik.

5. Bibliografia

- Almgren, M. eta Barreña, A. (2001). Bilingual acquisition and separation of linguistic codes: Ergativity in Basque versus accusativity in Spanish. In K.E. Nelson, A. Aksu-koc eta C.E. Jonson (Eds.), *Children's Language 11: Interactional Contribution to Language Development* (pp. 27-48). New Jersey: Lawrence Erlbaum Associates.
- Baltes, P.B., Reese, H.W. eta Nesselrode, J.R. (1977). *Life-span developmental psychology*. Monterrey, CA: Brooks/Cole.
- Barreña, A. (1995). *Gramatikaren jabeakuntza-garapena eta haur euskaldunak*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Bosh, L. (2004). *Evaluación fonológica del habla infantil*. Barcelona: Masson.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Ezeizabarrena, M.J. (1996). *Adquisición de la morfología verbal en euskera y castellano por niñas/os bilingües*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Fenson, L., Dale, P.S., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethick, S. eta Reilly, J. (1993). *The MacArthur Communicative Development Inventories: user's guide and technical manual*. San Diego: Singular Publishing Group.
- Fernández López, I. (2009). *¿Cómo hablan los niños? El desarrollo del componente fonológico en el lenguaje infantil*. Madrid: Arco/Libros.
- García, I., Arratibel, N., Barreña, A. eta Ezeizabarrena, M^aJ., (2008). Adaptación de los inventarios MacArthur-Bates al euskara: desarrollo comunicativo entre los 8 y 30 meses. *Infancia y Aprendizaje*, 31(4), 411-424.

- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Halliday, M.A.K. eta Hasan, R. (1985). *Language, context and text: a social semiotic perspective*. Victoria: Deakin University Press.
- Hymes, D. eta Gumpertz, J.J. (1964). "The Ethnography of Communication". *American Anthropologist*, 66(6).
- Idiazabal, I. (1988). *Adquisición del lenguaje en niños/as bilingües y monolingües*. VII Curso de verano. San Sebastián: Servicio Editorial de la Universidad del País Vasco.
- Kotliarenco, M. A.; Mendez, B. (1988). *Mirando hacia nosotros: El método de observación naturalista*. Santiago de Chile: Unesco, Oficina Regional de Educación para América Latina y el Caribe, Proyecto Principal de Educación en América Latina y el Caribe.
- 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. eta Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41(3), 841-849.
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, S., Mariscal, S. eta Martínez, M. (2005). *Inventario de desarrollo comunicativo MacArthur*. Madrid: TEA ediciones.
- Pérez Pereira, M. eta García-Soto, X.R. (2003). El diagnóstico del desarrollo comunicativo en la primera infancia: adaptación de las escalas MacArthur al gallego. *Psicothema*, 15(3), 352-361.
- Pearson, B.Z., Fernández, S.V. eta Oller, D.K. (1993). *Lexical development in bilingual infants and toddlers: Comparison to monolingual norms*. *Language Learning*, 43(1), 93-120.
- Romero, A., De Pablo, I., Etxebarria, A., & Romero, A. (2017a). Teorización sobre la construcción de corpus orales en la adquisición del lenguaje: una propuesta metodológica para el procesamiento de lenguas con soporte tecnológico, *Analecta Malacitana*, 42, 123-155. http://www.anmal.uma.es/AnMal42/Corpus_orales.pdf
- Romero, A., Etxebarria, A., De Pablo, I. & Romero, A. (2017b). Interrelación entre gestos y vocalizaciones en funciones comunicativas tempranas: evidencias desde la lengua vasca. *Signos*, 50(93), 96-123. doi:10.4067/S0718-09342017000100005

EUSKALKIEN SAILKAPEN ZIENTIFIKOA: LEXIKOA²¹**Gotzon Aurrekoetxea, Jose Luis Ormaetxea eta Xarles Videgain****Euskal Herriko Unibertsitatea eta Université de Pau et des Pays de l'Adour**
gotzonaurre@gmail.com, txipirrin@gmail.com, charles.videgain@unipau.fr**Laburpena**

Ekarpen honek ohiko euskalkien sailkapen zientifikoa proposatzea du helburu. Horretarako, EHHA egitasmoan bildutako datuak erabiliko dira, zehazki lehen bi liburukietan biltzen den lexikoa. Horrek esan nahi du gaur egun 100 urte inguru dituen edo izango lituzkeen jendearen lexikoa duela langai. Bi liburuki horietako datuekin datu-basea sortzean irizpide zehatz batzuk erabili dira galderak eta erantzunak aukeratzeko orduan; lehenengoa, lexiko orokorra aukeratzeko izan da; bigarrena, aukeratu den galderak gehienez ere erantzunik gabe bost herri edukitzea. Guztira 183 galdera aukeratu dira eta datuak datu-basean ezarri dira transkribaketa ortografikoa erabiliz. Datu-basea sortu ondoren, *Diatech* programara igo da eta programa hori erabiliz egin da analisi estatistikoa eta mapagintza. Analisisian similaritate-mapa esanguratsu batzuk eskaintzen dira, herri batetik abiatuz sareko gainerako herrietara dagoen desberdintasun lexikala nolakoa den zehazteko; horretan mapez gain, hizkuntza-distantzien taula ere agerraraziko da, hizkuntza-distantziak zenbakiz ere erakutsarazteko. Bigarren analisia clusterra izango da, zeinaren bidez hizkeren sailkapen hierarkikoa erakutsiko den. Sailkapen hau bi modutan erakutsiko da: batetik, dendrograma edo zuhaitz hierarkikoaren bidez, zeinetan aztertzen diren 145 hizkerak hizkuntza-hurbiltasunaren hierarkia baten barnean agertzen diren; eta bestetik, datuak mapa batean erakutsiko dira.

Gako-hitzak: bariazio geolinguistikoa, dialektometria, EHHA, lexikoa**1. Sarrera**

Euskalkien banaketa eta sailkapena aspaldiko kontua da gurean, euskaldunok betidanik izan dugu-eta gure hizkuntzaren bariazioa gogoan, sano deigarria delako guretzat ere. Alabaina, orain arteko sailkapen eta banaketa gehienak dialektologia tradizionalaren prozedurak erabiliz egin izan dira. Honako ekarpen hau, aldiz, metodologia eta prozedura desberdinak erabiliz gauzatu da.

Izenburutik beretik abiatuz, “sailkapen zientifikoa” terminoa erabili da. Zergatia eskatuko du batek baino gehiagok, eta arrazoiz. Eta guri dagokigu azaltzea: batetik, zientifikoa da ikerketa zientifikoetan, objetibotasun eta sinesgarritasunari begira, datu-kopuru handiak erabili ohi direlako; azterketa honetan 26.535 erregistro erabili dira. Ez da beharbada ikaragarritzko datu-kopurua eta handitu daiteke erregistro hori, zalantzarik gabe, eta etorkizunean egiteko asmoa ere badela jakinarazi gura dugu, baina orain artean sailkapen geolinguistikoak egiteko erabili den zabalena dela uste dugu. Bigarren, zientifikoa da zeinahi zientzian ezer sailkatu behar denean erabiltzen diren teknikak erabili direlako azterketa honetan ere; hots, sailkapen hierarkikoa edo, bestela esanez, cluster analisia.

Datu-kopuru handiak eta sailkapen-teknikak erabiltzeak ematen dio ekarpen honi duen garrantzia, gure ustez ezohiko garrantzia. Horrek ez du esan nahi dena eginda dagoenik, ezta hurrik eman ere, baina bai bide egokian jarri garela eta bide ziurrean ipini ditugula gure oinatzak. Egia esan, zientzian ez da ezer finkoa, dena da aldagai; gaur ziurtzat ematen dena bihar agian zalantzatan jartzen da eta etzi beharbada baztertu.

²¹ Lan hau GIU 22/16 eta IT 1028/16 erreferentziadun proiektueen barnean egin da.

Eta horretaz ere jabetzen gara. Horregatik, unean uneko teknika eta teknologia erabiliz egiten dugu ekarpen hau. Honetan gure egiten dugu gaur egun nazioartean jarraituenean eta aipatuenean artean kokatzen den J. Nerbonnek dioena (2008: 7): “Good dialectological analysis need to adopt techniques of aggregation more extensively.”

2. Metodologia

EHHako datu-bilketaren metodologiak ez da hemen ezer gehituko jada ezagutzen baita, gaian aditua denak aukera izan baitu orain arteko argitalpenetan (Aurrekoetxea 1986, 2002; Aurrekoetxea & Videgain 1987). Hemen EHHaren datuetatik datu-base berriaren sorrerako bidea nola eman den azalduko da.

2.1. EHHaren metodologia eta emaitzak

Irakurlea ondo koka dadin, EHHak emaitzak argitaratzean datuak bi moldetan ematen ditu: batetik, erantzunen zerrenda alfabeto fonetikoan, lurraldeka eta herrika antolaturik. Eta bestetik, mapetan, emaitzen lematizazio baten bidez, hitzen aldaerak lema berean bilduz. Adibidez, “sábado / samedi” kontzeptuari dagokion mapa (EHHa-II, 275. mapa) lau multzotan banatzen da: *zapatu* eta haren aldaerak mendebaldean, *larunbat* haren aldaerekin erdigunean, *ibiakoitz* nagusiki behe-nafarreraren eremuan eta *neskenegün* ekialdean. Oso mapa argia, banaketa dialektal sinplea duena eta hitz bakoitzari dagokion eremua oso modu bisualean erakusten duena. Hala ere, ez dira mapa guztiak modu horretakoak; bigarren adibide bat emateko, “peña / rocher” kontzeptuari dagokion mapa (EHHa-II, 377. mapa) aurrekoa baino ñabarragoa da, erantzun mota gehiago ditu eta banaketa dialektala ez da hain argia. Batetik *atx* eta *aitz* ditugu mendebaldean; bestetik *arkaitz*, *burkaitz*, *malkor*, *peña*, *barga* eta *arkide* erdigunean; hirugarren eremu batean *harroka* da nagusi eta, azkenik, ekialdean *botxu* da forma zabalena. Aipatu gabe gelditzen dira, baina ez garrantzi txikiagoa dutelako, oso eremu txikietan erabiltzen diren berbak.

Lan hau egiteko ez dira batere kontuan izan mapak ezta mapak sortzeko egin zen lematizazio eta legenda. Abiapuntua erantzunen zerrenda izan da. EHHak erantzunen zerrendako emaitzak alfabeto fonetikoan ematen dituen arren, proiektu honetan emaitzak zuzenean eraldatu egin ditugu eta erantzun ortografikoan idatzi dira gure datu-basean. Galderaz galdera eta herriz herri datuak banan-banan idatzi dira sortu dugun datu-base berrian.

2.1. EHHatik datu-basera

Ikerketa honen abiapuntua EHHa proiektuko lehen eta bigarren liburukien paperezko bertsioak dira. Paperezko bertsioa esku artean, ondoko baldintzak betetzen zituzten galderak aukeratu dira:

a) erantzun hutsik ez edo huts gutxi dituzten galderak bakarrik hartu dira kontuan. Galderaren erantzunen zerrendan bost herri edo gehiagotan erantzunik ez denean izan, galdera hori ez da kontuan hartu.

b) lexiko orokorra bakarrik aukeratu da, edozein gai dela kontzeptu orokorrak biltzen dituzten galderak bakarrik onartu dira. Hori dela eta, adibidez, baztertu egin dira intsektuei, arrainei e.a. dagozkien galderak.

Bi baldintza hauek kontuan izanik eratu da datu-base berria. Datuen kopuruari dagokionez, EHHako lehen bi liburukietan 509 galdera argitaratu diren arren, datu-base berrian lan honetarako 183 galdera baino ez dira lortu ahal izan; gainerako galdera guztietan edo erantzun hutsak direla-eta edo galderetan lexiko berezitua (edo berezitzat jotakoa) biltzen delako, bazter utzi dira.

Liburuetako emaitzak datu-base batean ezarri dira. Horretarako aurretik datu-base eredia zehaztu da. Hauek dira sortu den datu-basearen ezaugarriak:

- formatua: .csv formatuan eraiki da (*Openoffice*-eko *Calc* formatua erabili da).
- Unicode UTF-8an kodetu da formatu hori datuak gordetzean.
- 183 galdera ditu datu-base berriak.
- 145 herri biltzen dira datu-basean, EHHAn erabili diren herri guztiak.
- Guztira 26.535 erregistro biltzen ditu datu-baseak.

Lan hau guztia eskuz egin da, ekarria izenpetzen dugun hiruron artean, leku berean eta aldi berean hirurok langintza bera izan dugularik, galderak banandu ondoren eta bakoitzaren ordenagailuan datu-base eredia ezarri ondoren. Horretarako aste osoa behar izan da, bost egunetako jarduera latz eta gogorra, zeinetan aldian-aldian zalantzaren bat izan denetan elkarri galderak eginez erantzun egokiena bilatu izan den. Paperezko liburukietatik datu-basera pasatzean datuak alfabetoz ere aldatu egin dira, gorago esan bezala: liburukietan, ezaguna den legez, erantzunak alfabeto fonetiko transkribatuta dauden bitartean datu-basean idaztean alfabeto ortografikora pasatu dira, horretarako euskaraz idazteko baliatzen den alfabetoa erabiliz.

Aipatu datuen eraldatze-lanean EHHAn agertzen diren erantzunak (lekukoek euren kasa emandako erantzunak) eta proposamen onartuak (inkestazaileek proposatu eta lekukoek herriko forma edo berbatzat onartuak) maila berean ezarri dira. Uste dugu lekukoek bere kasa emandako erantzuna eta inkestagileak hitz bat iradokiz gero herrikotzat jotzen duenean, eta herriko edonork erabil lezakeela dioenean, herriko hitza dela, nahiz ez zaion hiztunari bere kasa atera inkestaren unean.

6236	44	1	kopo
6237	44	2	alpar,matasa
6238	44	3	matasa,txingor
6239	44	4	oja,txotor
6240	44	5	metxa
6241	44	6	edurtxotor
6242	44	7	edurmata,txingor
6243	44	8	edururtantakada,edurtxingor

6244	44	9	tantakada,matasa
6245	44	10	edurtxistil,edurmako
6246	44	11	matasa,edurlapatz
6247	44	12	edurgarrafa,edurmatasa
6248	44	13	edursati,edurlume
6249	44	14	edurmatasa,edurtxingor

1. taula: datu-base berrian 44. galderari (EHHako 09090) dagokion datu-base zatia

Datu-base honen egitura honako hau da (1. taula): lehen zutabea datu-baseak automatikoki sortzen duen erregistro zenbakia da, bigarrena galdera-zenbakia (datu-base berrian 1-etik hasi da kontatzen), kasu honetan 44. galderaren zenbakia da. Hurrengo zutabea herriaren kodea da (1 lehen herria, Getxo..., mendebaldetik ekialderanzko norabidean). Beraz, taula honetan agertzen diren erantzunak Bizkaiko mendebaldeko herrietan jaso direnak dira.

2.3. Hizkeren arteko desberdintasunak neurtzeko sistema

Behin datu-basea sortu denean, *Diatech* tresnara igo dira datuak. Ezaguna da honezkero analisi geolinguistiko kuantifikatuak egiten online tresna (<http://eudia.ehu.eus/diatech/index/>). *Diatech* EUDIA ikerketa-taldeak bariazio geolinguistikoa ikertzeko analisi estatistikoak egin eta mapa kuantitatiboak lortzeko tresna da. Tresna hau doanekoa da eta edonork erabil lezakeena; horretarako behar duen gauza bakarra izena ematea da, besterik ez.

Datuak *Diatech*era igo ondoren, hizkeren arteko desberdintasunak zehazteko erabiliko den neurria erabaki behar da lehenik. Dialektometriari neurri bat baino gehiago erabiltzen da: RIV neurria (*Relative Index Value*) Salzburgo-ko eskolak proposatu zuen eta haren jarraitzaileek erabiltzen dute, nagusiki (Goebel 1978). *Levenshtein algoritmoa* Groningen-go eskolakoek eta haren jarraitzaileek erabiltzen dute (ikus besteak beste, Nerbonne & Heeringa 2010). Aurrekoetxeak bi neurriak erabiltzea proposatzen du helburuen arabera, eta biak erabilgarri ditu *Diatech* programarekin analisi estatistikoak egin nahi dituenak. Neurriok desberdinak dira oso: lehena, RIV delakoa, neurri kategorikoa da eta konparatzen diren hitzak bere osotasunean hartzen ditu; nahikoa da zeinu edo hizki bat desberdin izatea hitzak desberdintzat hartzeko (neurri horren arabera eta adibide bat jartzeko, *ixar* eta *izar* %100ean desberdinak izango lirateke). Neurri hori egokia da distantzia fonologikoa neurtzeko, adibidez. Bigarren neurria, *Levenshtein algoritmoa*, ez da distantzia kategorikoa, eta hizkien galtzean, aldatzean edo sortzean oinarritzen da (neurri honen arabera, *ixar* eta *izar* hitzen arteko desberdintasuna %25 da, lau hizkitatik bat dute desberdin); neurri hau oso da egokia distantzia fonetikoak neurtzeko. Bata ala bestea erabiltzea ikertzaileari dagokio eta ikerketaren helburuen arabera izaten da (ikus zehatzago Aurrekoetxea 2016: 28-30).

Ekarpen hau erantzunetan dago oinarrituta, lekukoengandik jasotako erantzunetan, hauek ahalik eta modu zehatzenean transkribatuta, baina alfabeto arauemailea erabiliz, ez transkripzio fonetikoak. Emaizta hauetan ez da lematizaziorik egin eta ez dira lemak

erabiliko ikerketa honetan. Beraz, distantziak neurtzeko datuak erantzunak direnez, distantzia fonetikoak lortuko direla diogu, nahiz erantzun ortografikoetan oinarritu; ikerketa honetan distantzia fonetikoak fonologikoei kontrajartzeko termino moduan erabiltzen da. Distantzia fonetikoak lortu nahi direnez, ikerketa honetan *Levenshtein algoritmoa* edo neurria erabiliko da.

2.4. Erabiliko diren estatistika-teknikak

Bariazio geolinguistikoa aztertzeko garatu diren teknika kuantitatiboak askotarikoak dira. Ekarpen honetan horietatik aukera bat egin da, hain zuzen ere hizkeren sailkapen hierarkikoa aztertzeko erabiltzen direnak dira aztertuko direnak. Baina hori baino lehen eta jomuga hori izan arren, hizkeren arteko hizkuntza-distantzien taula eta similaritate-mapa bat aurkeztuko da. Hizkuntza desberdintasunak zenbakietara iraulgiz gero, hizkeren arteko desberdintasunak zenbakiz neur daitezke eta ondoren hizkeren arteko hizkuntza-desberdintasuna zenbakiz zehatz daiteke. Berrikuntza eta aurrerakuntza ikaragarria da hau, eta aurrerantzean hizkera batetik besterako desberdintasunaz hitz egitean ez da zertan “desberdintasun handia ala txikiaz” hitz egin behar, zehatz kuantifikatuta eman daitekeelako.

Ondoren, hizkeren sailkapenari begira, sailkapen hierarkikoa (cluster analisia) burutuko da. Cluster analisia teknika erabiltzen da ikerketa bateko datuak sailkatu nahi direnean; teknika unibertsala eta zientzia guztietan erabiltzen dena da. Azterketa honetan sailkapen honen bi emaitza azalduko dira: dendrogramak edo zuhaitz hierarkikoak lehenik; eta mapak bigarrenik. Mapetan balizko euskalkien eta azpieuskalkien mapak aurkezten dira.

3. Analisia

Gogora dezagun lehenik, analisi honetan hizkuntza-distantziatzat *Levenshtein* algoritmoa erabiltzea erabaki dela eta datuak lematizatu gabe erabili direla. Analisisan lehen-lehenik hizkuntza-distantzien taula aurkeztuko da, ondoren similaritate-mapa bi (bata Getxori dagokiona, ezker muturrean kokatuta dagoen herria; bigarrena Eskiulari dagokiona, hau eskuin muturrean, ekialdean) eta hirugarren sailkapen hierarkikoari dagokion dendrograma eta mapak.

3.1. Hizkuntza-distantziak

Esan legez, jada hizkeren arteko desberdintasunak zenbatzeko aukera dugu egun dialektometriarako sortu diren programa informatikoen bidez. Kasu honetan hizkeren arteko distantzia-etaula *Diatech* programaren bidez lortu da. Programa honek honelako taulak jaisteko aukera ematen du eta hori da oraingoan egin den eragiketa (ikus 2. taula).

	Abaurregaina	Ahetze	Aia	Aldude	Alkotz	Altzai	Altzurükü	Amezketeta	Andoain	Aniz
Abaurregaina	0,00	47,52	51,73	44,87	46,73	61,12	62,93	50,70	53,56	41,09
Ahetze		0,00	51,09	36,41	50,98	57,02	57,65	51,71	51,39	42,66
Aia			0,00	53,93	50,13	66,02	66,18	37,06	39,46	47,84
Aldude				0,00	51,57	56,14	58,13	52,63	54,33	42,38
Alkotz					0,00	66,23	67,34	49,69	49,15	36,90
Altzai						0,00	31,26	65,92	66,11	61,65
Altzurükü							0,00	67,62	67,70	63,49
Amezketeta								0,00	33,26	46,47
Andoain									0,00	47,61
Aniz										0,00

2. taula: hizkeren arteko distantzien taula (zatia)

Taulak erakusten duen bezalaxe, Abaurregainetik Ahetzeko hizkerara datuen %47,52 desberdintasuna dago, Aiakora %51,73, Aldudekora %44,87, e.a. Hots, Abaurregaineko hizkeratik Ahetzekora baino Aldudekora desberdintasun txikiagoa dago, eta Ahetzekora baino Aiakora handiagoa. Beste konparaketa bat eginez, Amezketatik Aiako hizkerara dagoen desberdintasuna Altzürüküra (67,62) baino ia erdia txikiagoa da (37,06). Zenbaki hauek adierazten dutenez, aztertzen diren herriak binaka elkarrekin konparatuz, eta gure datu-basea oinarri izanik, lexikoan hizkeren artean aurkitzen den desberdintasun handiena %73,26 da (Altzürükü eta Getxoko hizkeren artean jaso). Desberdintasun txikiena, aldiz, %22,87 da (Aniz eta Lekarozeko hizkeren artean jaso da).

Zenbaki hauek begi aurrean, berretsi egiten da datu zehatzik gabe beti esan izan den baieztapena: euskararen bariazioa handia dela oso; edo bestela esanda, euskararen dialektalizazioa gainerako inguruko hizkuntzetakoa baino handiagoa dela. Hizkera batetik besterako distantzia txikiena %22,87 izatea oso desberdintasun handia da; lexikoari dagokionez, bi hizkeren hizkuntza aldetik hurbilen direnen artean ia laurdena desberdina dela esan nahi du horrek. Kontuan hartu behar du irakurleak, berriro diogu, Levenshtein distantzia unitatea erabili dela estatistika hauetan.

Datuok zehaztasun handia ematen diote azterketari. Ikaragarritzko aurrerapena da hizkuntza-desberdintasunak xeheki aztertu nahi dituenarentzat. Hizkuntza-datuak estatistikaren galbahetik iragaziz aitzinatze bortitza lortzen da, datuok erakusten duten bezala.

Behin zenbakizko datuen taula lortuz gero, distantzia handienak zein txikienak zein herrik dituen ere azter daiteke:

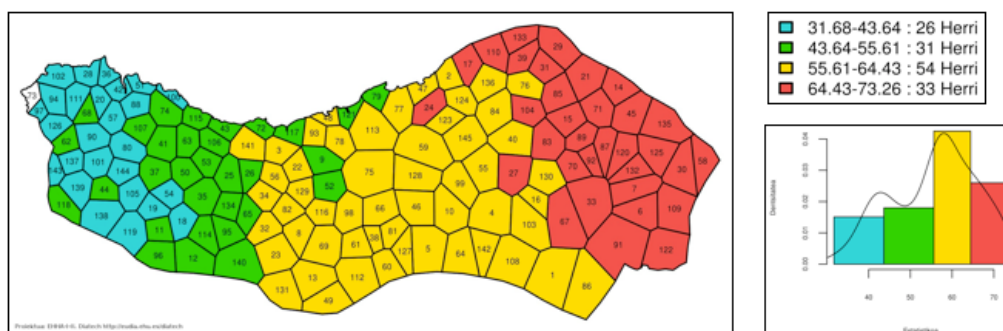
a) batetik, mendebaldeko muturretik abiatuz distantzia handienak biltzen dituzten mendebaldeko herriak zein diren jakin daiteke: Getxotik Altzürüküra dago hizkuntza-distantzia handiena (%73,26), ondoren Laukizetik Altzürüküra (%73,17), gero Getxotik Domintxainera (%72,21) eta Getxotik Montorira (%72,23). Hots, Getxo eta Altzürüküren artean lortzen da hizkuntza-distantzia handiena aztertzen diren datuekin, kontaketa mendebaldeko muturreko herrietatik abiatuz.

b) Ekialdeko muturretik abiatuz distantzia handiena Altzürükütik Getxora kontatzen da (%73,26), bigarren Altzürükütik Laukizera (%73,17), hirugarren Altzürükütik Lekeitiora (%72,61), laugarren Altzürükütik Etxebarrira (%72,29), ondoren Santa Grazitik Getxora (%71,85) eta Larrainetik Getxora (%71,66). Altzürükü da hizkuntza-kontuetan gainerako hizkeretatik desberdintasun handienak erakusten dituen herria: Getxorekin, Laukizekin, Etxebarrarekin eta Lekeitiorekin. Ondoren Getxo aurkitzen da: aipatu Altzürüküz gain, Domintxaine, Montori, Larraine eta Santa Graziarekin lortzen ditu guarismo handienak.

Bestalde, basean aurkitzen diren desberdintasun txikienak edo, bestela esanda, hizkuntza-hurbiltasun handienak hauek dira: Aniz eta Lekaroz artean (%22,87), Ahetze eta Senpereren artean (%25,36), Erratzu eta Anizen artean (%26,44) eta Lemoiz eta Bakioren artean (%27,40). Ikusten denez, desberdintasun txikienak edo hizkuntza-hurbiltasun handienak geografikoki oso hurbil diren herrien artean ematen dira.

3.2. Similaritate-mapak

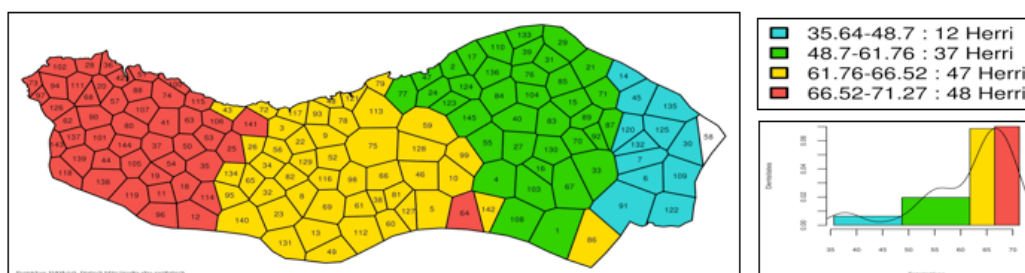
Mapetara iraganez, ikus dezagun nolako similaritate-distribuzioa duten bi muturretako herriek: Getxo mendebaldeko muturrean eta Eskiula ekialderen kokatuta dagoen herria. Lehen herriari dagokionez (1. mapa), Getxok ikertzen diren gainerako herrietako hizkeraren aldean dituen desberdintasunak lau multzotan bananduz, lehen multzoan (urdinez) hurbilak diren herriak agertzen dira; guztia 26 herri, denak Bizkaian kokatutakoak. Bigarren multzoan (orlegiz) 31 herri aurkitzen dira, Bizkaia eta Gipuzkoan kokatuta. Hirugarren multzoan (horiz) 54 herri, Gipuzkoan, Nafarroan, Lapurdin eta Nafarroa Beherean kokatua. Azkenik, laugarren multzoan 33 herri, denak Iparraldean kokatuta.



1. mapa: Getxo herriari dagokion similaritate-mapa

Ez da hemen sakonduko, baina etorkizunean mereziko luke sakontzea lehen multzoaren barnean kokatzen diren bigarren multzoko herrien (Etxebarri-62, Fika-68, Dima-44 eta Orozko-118 herriak) ezaugarrietan. Edo hirugarren multzoko herriz inguratuta aurkitzen diren Gipuzkoakoak (Andoain-9, Elduain-52, Hondarribia-79 eta Pasaia-121 herriak).

Ekialderen dagoen herria aukeratuz, Eskiulak gainerako herriekin dituen similaritateak agertuko dira (2. mapa). Eskiulak (mapan zuriz dagoen poligonoa dagokio) gainerakoekin dituen similaritateak lau multzotan bananduz lau multzo geografikoki aski koherenteak lortzen dira: lehen multzoan Zuberoako eta Amikuzeko 12 herri agertzen dira, bigarren multzoan (orlegiz) 37 herri kokatzen dira: gainerako Nafarroa Behereko, Lapurdiko guztiak eta Nafarroako ekialdeko herri batzuk. Hirugarren multzoan (horiz) 47 herri agertzen dira: Nafarroako eta Gipuzkoako herriak dira. Eta, azkenik, laugarren multzoan Gipuzkoako ekialdeko eta Bizkaikoak. Etorkizunean sakontzeko utziko da Eugi-64 herriko hizkeraren ezaugarriak beste multzo baten kokatuta dauden herrien artean egotearen arrazoiak aztertzea.



2. Mapa: Eskiulari dagokion similaritate-mapa.

Bi mapa hauek, bigarren honek garbiago, erakusten dute dialektologoen artean betidanik izan den axioma: zenbat eta geografian urrunago joan orduan eta hizkuntza-desberdintasuna handiagoa izango dela. Interesgarri oso eta aipagarri une honetan behinola Séguyk (1971) defendatu zuena, alegia, Europan zehar herri batetik bestera joanez elkar ulertzen zela eta zenbat eta urrunago izan orduan eta jatorritik hizkuntza bestelakoagoa zela.

3.3. Sailkapen hierarkikoa

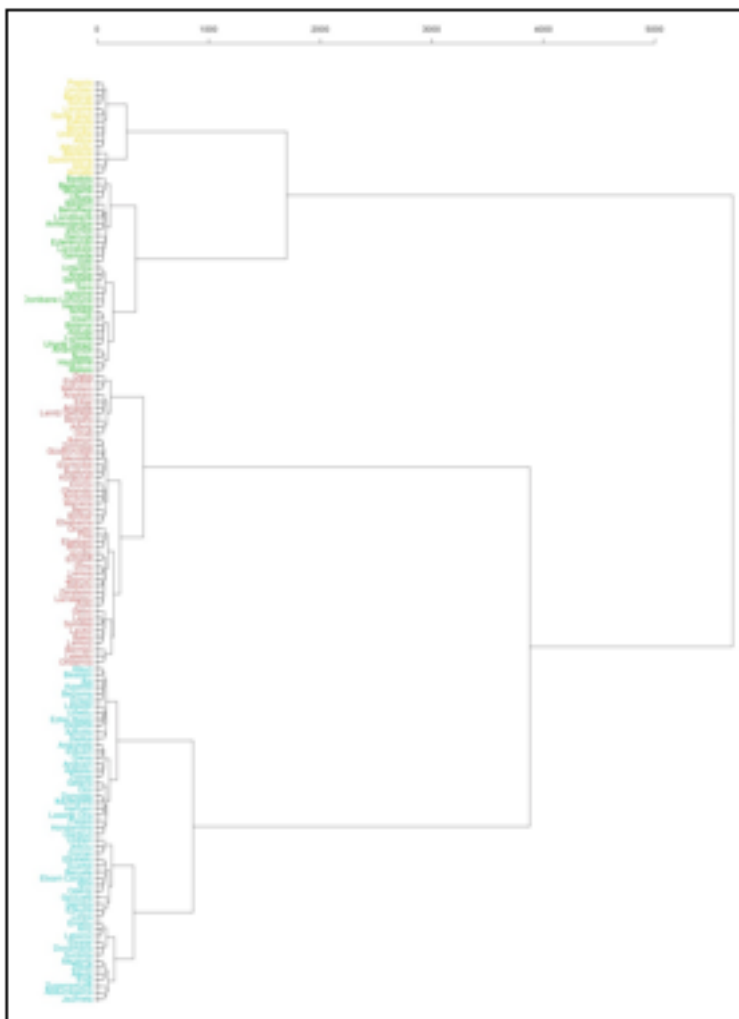
Hizkeren sailkapen hierarkikoa egiteko cluster analisia erabiliko da. Dialektologia modernoan hizkeren sailkapenaz eta banaketaz nazioartean ari diren ikertzaile gehien-gehienek erabiltzen duten teknika da hau. Teknika estatistiko honek bi emaitza ematen ditu ohiz: dendrograma edo zuhaitz hierarkikoa deitua eta honetatik ondorioztatzen diren mapak. Ekarpenean honetan biak aurkeztuko dira.

3.3.1. Dendrograma

Cluster analisiaren lehen emaitza dendrograma deitzen den zuhaitz hierarkikoa da. Zuhaitz honek hizkerak hierarkia baten barnean ezartzen ditu elkarren artean dituzten hizkuntza-ezaugarrien arabera hurbiltasuna erabiliz. Dendrograma eraikitzeke Ward metodoa erabili da ikerketa honetan, metodo hau delako aseptikoena eta distortsio gutxien sortzen duena. Gainerako zehaztasunak aurreko analisian erabilitako berak dira.

Dendrogramaren mozketarako bi irizpide erabiliko dira: Guiter-en irizpidea eta dendrogramaren besoen luzera. Guiterrek (1973) hizkeren arteko desberdintasunak mailakatzea proposatu zuen; haren arabera, %0-25 bitarteko desberdintasunek hizkerak bereizteko balio lezakete, %25-%50 bitarteko desberdintasunek azpidialektoak bereizteko, %50-%75 bitartekoek dialektoak bereizteko, eta hortik aurrera hizkuntzak bereizteko. Aurrekoetxeak (2012) hizkeren arteko hizkuntza-mugan beharrezko mailakatzeaz jardutean, berrartu zuen ideia eta hiru maila proposatu zituen. Ikerketa honetan sortutako dendrogramaren eskalak orotara 5.800 maila ditu: Guiterren arabera jokatzuz, eskalako %25erainoko tartean barietateak edo hizkerak bereizten dira, %50erainokoa azpidialektoak eta %75erainokoa dialektoak. Gorago aipatu dendrogramara itzuliz eta %25ean moztuz, eskalako 1.450 mailan, lau dialekto genituzke; %50ean eginik, 5.800 eskalako 2.900 mailan, hiru dialekto izango genituzke; eta azkenik, %75ean eginik, eskalako 4.350 mailan, bi dialekto edo euskalki lortuko liriteke. Gure kasuan eta aurreko arlo honetako literatura kontuan izanik, ez dirudi sistema egokiena denik.

Bigarren irizpidea estatistikan cluster analisia egiten dutenen artean oso ezaguna da: dendrograma bat izanik (ikus 1. irudia), hizkeretatik (irudiaren ezkerrean koloreetan) edo hizkera multzoetatik sortzen diren adarren luzera hartzen da kontuan.



1. irudia: cluster analisiaren bidez lortutako dendrograma

Adar horien luzetasunak hizkuntza-desberdintasuna edo hizkuntza-disimilaritatea adierazten du. Hizkuntza-desberdintasun edo disimilaritatearekin hizkuntza-distantzia esan nahi da, beste molde batez. Zenbat eta adarra luzeagoa izan, orduan eta hizkuntza-distantzia handiagoa da adar-muturrean biltzen diren bi multzoen artean.

Dendrograman agertzen den sailkapena mapa batean islatu behar denean, dendrograma hori nonbait moztu behar izaten da; mozketa horretan sortzen diren multzoen kopurua mapan koloreetan bereizten diren eremuetan islatuko da. Beraz, garrantzi handia du dendrograma non moztu behar den zehazteak, edo zein den mozteko lekurik egokiena zehaztea, banaketa dialektal egokia lortzeko.

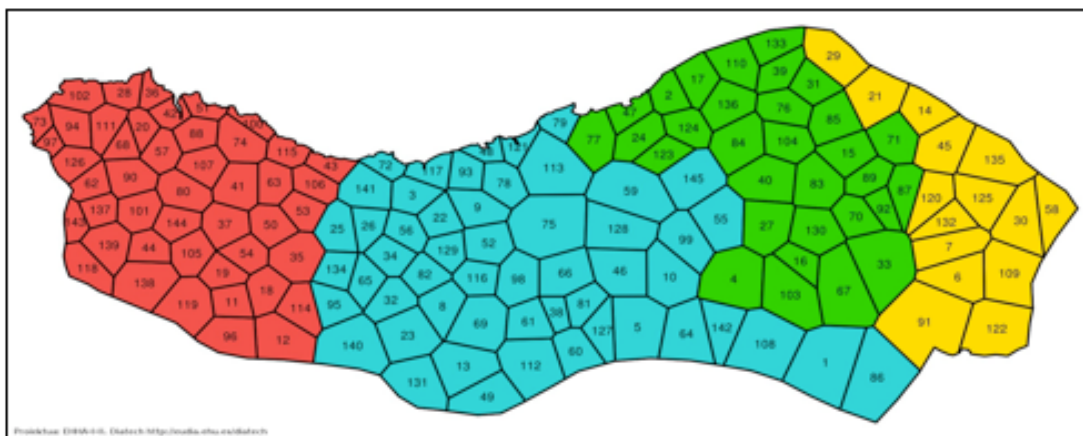
Funtsean honek zera esan nahi du: muga dialektal guztiek ez dutela garrantzi bera, muga batzuk garrantzitsuagoak direla, hizkuntza-desberdintasun handiagoa erakusten dutelako, eta alderantziz. Gauzak horrela, dendrograma moztu nahi denean garrantzi handia dauka mozketa-leku edo maila ondo aukeratzea. Badira mota honetako estatistika ikertzen dutenen artean tresna automatizatuak dendrogramaren mozketa-maila egokiena egiten laguntzen dutenak. Denek bat egiten dute mozketa adarraren luzetasunarekin lotzean. Gure kasura etorritz, eskalako 900 maila inguruan 5 adar aski esanguratsu agertzen dira: bat, oso luzea, gorriz idatzita agertzen diren herriek osatzen

duena, beste bi ere aski luzeak, hori eta orlegiz idatzita agertzen diren herriek osatuak, eta azkenik, beste bi adar, laburragoak urdinez idatzitako herriek sortuak.

Dendrogramak osotara 5.700 maila ditu bere eskalan; horren %25a 1.425 maila da, %50a 2.820 eta %75a 4.275 maila. Desberdintasuna %25 delarik, lau multzo agertzen dira: horia, gorria, orlegia eta urdina; desberdintasuna %50 bada (dialektoak bereiziko lirateke Guiterren esanetan), hiru multzo (Iparraldeko herriak, mendebaldeko gorria eta gainerako guztiak: Gipuzkoa gehiena eta Nafarroako guztiak), eta %75ean multzo bi: Iparraldeko herriak batetik, eta Hegoaldekoak bestetik.

3.3.2. Mapa

Demagun dendrograma 1.425 mailan mozten dela (hizkuntza-desberdintasunen %25a). Arestian esan bezala, lau multzo agertu dira dendrograman mozketa maila honetan eginez gero. Lau multzo horiek mapan kokatuz ondokoa lortzen da (ikus 3. mapa).

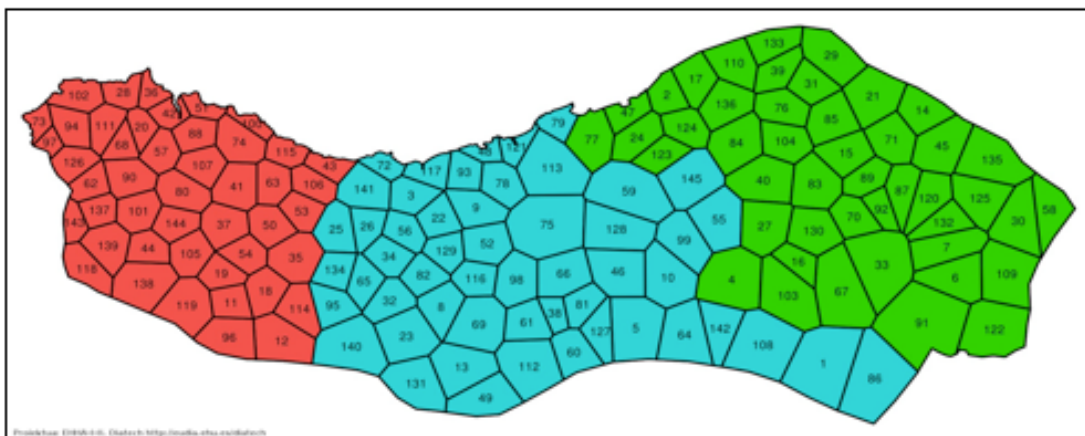


3. mapa: cluster analisiaren bidez lortutako mapa (ward, 4 multzo, Levenshtein).

Mapan lau multzo azaltzen dira: mendebaldekoa gorritz margotutakoa, erdialdekoa urdinez, ipar-ekialdekoa orlegiz eta ekialdekoa horiz. Lehen aldia da datu lexikalekin euskararen eremu osoko datuetan oinarrituz eta prozedura kuantitatiboak erabiliz mapa bat egiten dena. Garrantzi handiko emaitza da, zalantzarik gabe, orain arte egin izan diren euskalkien mapak prozedura tradizionalekin eginak izan direlako, eta hau oso bestelako metodologia eta prozedurak erabiliz. Laburbilduz, mapa hau lortzeko Euskal Herri osoan barreiatutako 145 herritan lortutako lexiko orokorreko 183 galderaren erantzunak erabili dira, gorago esan bezala. Guztira 26.535 datu bildu dira sortu den datu-basean. Datu horiek erabili dira ondoren, *Diatech* programarekin mapa hau egiteko. Mapa sortzean cluster analisia egin da; cluster analisia hizkeren arteko sailkapen hierarkikoa egiteko erabiltzen da. Kasu honetan Ward algoritmoa erabili da, Levenshtein hizkuntza-distantziarekin batera.

Mapa hau sortzeko, dendrograma eskalaren %25 mailan moztu da. Honek esan nahi duena honako hau da: multzo bakoitzaren barneko hizkuntza-desberdintasuna %25 baino txikiagoa dela eta multzoen arteko desberdintasuna %25 baino handiagoa dela.

Aldiz, lau multzotan banandu ordez hirutan egingo balitz, hiru eremu dialektaletako mapa lortuko litzateke (ikus 4. mapa). Mozketa hau dendrogramaren eskalako 1.750 maila inguruan legoke (desberdintasunen %30,70).



4. mapa: cluster analisisian dendrograma 3 multzotan banandu osteko mapa

Eta azkenik, 5 multzoko mapa egin gura izanez gero, dendrogramaren eskalan 850 mailaraino (desberdintasunen %14,9) jaitsi beharko litzateke. Maila honetan orain arteko mapetan urdinez agertzen den erdiguneko eremua bitan banatzen da: batetik multzo batean Gipuzkoako herriak biltzen dira, bestetik Nafarroakoak. Ondorioz, bost eremuak hauek lirateke (3. mapan oinarrituz): mendebaldean gorritz margotutako eremua; eremu urdina bitan bananduz, Gipuzkoako herriek eta Nafarroakoek osatzen dituzten eremuak; laugarren eremua orlegiz margotutakoa da, eta bosgarrena horiz margotua.

3.3.3. Euskalkien mapa

Metodologia hau hizkeren sailkapena egiteko baliabide eskuragarria dugu. *Diatech* tresnak, gainera, erabiltzailearen esku uzten du eremu-kopurua zehaztea. Ikertzaileak baliabide handiak ditu eskura gaur egun hizkerak sailkatzeko orduan. Kontua, ostera, bestelakoa da: sortzen diren eremu horien mailakatzea; hots, zer den azpidialektoa, dialektoa, eta abar. Nola zehaztu eremu dialektal bat dialektoa den ala beste zerbait? Marraztu diren mapa biak kontuan izanik, eta Guiterrrek planteatu zuena aintzat hartuz, euskarak hiru dialekto baino ez litzateke. Honek, ostera, ez du aintzat hartzen euskal dialektologiaren historian egindako euskalkien mapetan egin den banaketa.

Guiterreren proposamenaren aurka esan behar da berea, zehaztu ez zuen arren itxura guztien arabera, hizkuntza erromanikoetarako dela baliagarri. Edonola ere, ez da aplikagarri edozein kasutan, hizkuntza batzuetan bariazioa oso handia izaten delako, eta besteetan, aldiz, askoz ere urriagoa. Egia da berea izan dela saio bakarra muga eta eremu dialektalen hierarkiak planteatzen. Bere proposamena aho betekoa ez bada ere, lehen saiakeratzat egokia dela esan behar da. Zelan aurreratu teorikoki arlo honetan?

Aurrez aurre dialektologia tradizionalerako banaketa dialektalak ditugu. Arlo honetan ere hizkuntza bakoitzean eta ikertzaile bakoitzak bere kasa asmatu ditu mugak eta eremuak. Horren ondorio zuzenena honako hau da: hizkuntza batean egin den banaketa dialektalean sortu diren eremu dialektalen (dialektoen) arteko

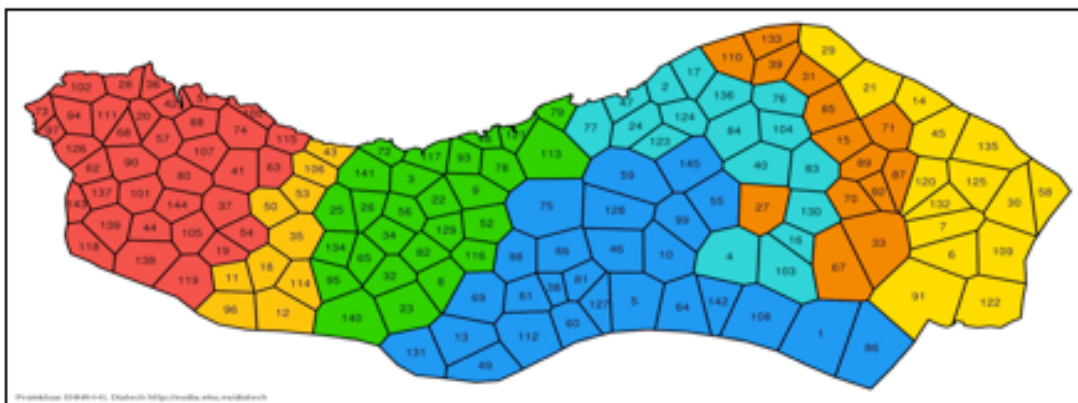
desberdintasunak oso handiak izan daitezkeela; hizkuntza batean %25 inguruko desberdintasunean koka daitekeen bitartean, beste batean %50ean izan litekeela. Eta, gainera, ezinezkoa da zehaztea zein ehunekotan eman izan den. Gure dialektologiari etorri, nola zehatz daiteke L. L. Bonaparteren edo berrikiago egindakoen euskalkien banaketan zein den bizkaiera eta gipuzkeraren arteko mugako desberdintasunen kopurua? Dialektologia tradizionala ez da horretaraino heltzen, mugak zehaztu eta eremu dialektalak bereiztea zuen helburu eta ezinezko zitzaien zehaztasun gehiagotara heltzea. Hala ere, ez zuen arazorik izan sortzen diren eremuei “dialekto” izendapena emateko, dialektoen arteko mugak oso desberdinak izan arren.

Aurreko hori guztia kontuan izanik, eta aurrerapen edo lehen hurbilketa gisa, desberdintasunen %25eko mailari dagokion mapa hurbilago dago ohiko euskal dialektologian egin izan diren euskalkien banaketarekin, %50eko mailari dagokiona baino. Hau ontzat emango balitz, benetako euskalkien mapa 3. mapa izango genuke: lau euskalki genituzke, mendebaldean bat (aurreko sailkapenetan bizkaiera edo mendebaldekoa deitua izan dena), erdigunean bigarrena (aurreko banaketetan gipuzkera eta goi-nafarrera, edo erdialdeko eta nafarrera bezala izendatuak), hirugarrena orlegiz margotua (ohiko banaketetan lapurtera eta behe-nafarrera, edo nafar-lapurtera izendatuak) eta horiz margotua ekialdean (aurreko banaketetan zuberera deitua).

Euskalkien sailkapen hau ez dator bat ez L. L. Bonaparteren banaketarekin, ezta K. Zuazorenarekin (1998). Desberdintasunak alde guztietan aurkitzen dira: batetik, datuak desberdinak dira oso: lexikoaz bestelakoak (nagusiki) Bonaparte eta Zuazorenak, lexiko hutsak hemengoak. Bestetik, prozedurak ere oso bestelakoak dira: Bonaparte-Zuazoren kasuan eremuak definitzetik abiatu ziren, hemen hizkuntza-datuak desberdintasunetatik. Lehenengoen eremuen hizkuntza-ezaugarriak bereizten zituzten, hemen eremuak ondorio hutsa dira. Hirugarren, kuantifikazioa, haiek eskuz eta oraingo honetan metodo sailkatzaile-konputerizatuak erabiliz. Laugarren, emaitzak: Bonapartek 8 euskalki banandu zituen bitartean, Zuazok 5etan utzi zuen, eta hemen 4tara ekarri dira.

3.3.4. Azpieuskalkien mapa

Dendrograma eskalaren 275 mailan (desberdintasunen %4,82 biltzen da maila honetan) zatituz gero, gorago definitu diren lau euskalkiak bitan zatituko lirarteke (5. mapa), Zuberoakoa salbu. Mendebaldeko euskalkiak bi azpieuskalki lituzke maila honetan: mendebaldekoa (gorritzatua) eta ekialdekoa (laranja kolorekoa); erdigunekoak beste bi: Gipuzkoakoa (orlegiz) eta Nafarroakoa (urdin indartsuz); nafar-lapurterak ere bi: mendebaldekoa (urdin argiz) eta ekialdekoa (laranja ilunez); eta Ekialdeko euskalkiak bakarra (horiz).



5. mapa: cluster analisisan 7 multzoak egindako mapa

Mapak erakusten digu nafar-lapurtera euskalkiaren multzoa lasterrago zatitzen dela mendebaldeko euskalkia baino; eta euskalkietan azkena zatitzen erdiguneko dela (Gipuzkoako eta Nafarroakoak sortzen dutena). Mapan eremuak oso zehatz irudikatzen dira, salbuespen bakarrarekin, mendebaldeko nafar-lapurteran kokatuta dagoen baina ekialdeko nafar-lapurterari dagokion Baigorriko hizkera (mapan “27” zenbakiarekin identifikatua).

Ohiko banaketa dialektalarekin erkatuz, esan behar da azpieuskalkietako mugek hor dirautela nagusiki, leku berean edo oso gutxi aldatuta: ohiko bizkaieraren Gipuzkoako azpieuskalkiaren muga eta prozedura sailkatzaileekin lortutakoa bera da, ez da batere aldatu. Ohiko gipuzkeraren barneko banaketa ez da prozedura berriekin maila honetako banaketa, oraindik maila apalago batean kokatuta agertzen da; gune dialektal hau oso trinkoa da eta horren erakusgarri dendrograma dugu: oso maila apalean, behaldez, biltzen da eremu honetako hizkera guztietan cluster edo multzoa.

Badira muga batzuk mugitzen ez direnak eta horien artean kokatzen da Nafarroako hizkerena. Hizkera guztiak biltzen dira erdiguneko euskalkiaren Nafarroako azpieuskalkian. Eremuaren ekialdean, ohiko zubererak Amikuzeko hizkerak ere bereganatu ditu, baina ez denak: Garrüze (71 mapan) ez da eremu dialektal horretan kokatzen, ekialdeko nafar-lapurteran baizik. Aldiz, Bardozeraino (29) heltzen den eremua da.

4. Eztabaida

Euskalkien sailkapen hierarkikoa egitea izan da helburu ekarpen honetan, gaur egun dialektometriari erabiltzen diren prozedura estatistikoak erabiliz, horretarako datu lexikalak erabiliz. Ohiko euskal dialektologian, aldiz, banaketa dialektala egiteko nagusiki datu gramatikalak erabili dira batetik, eta bestetik, eskuzko prozedurak, ez automatizatuak. Erabilitako datuak eta prozedurak oso bestelakoak izan dira, guztiz.

Emaitzak ere guztiz desberdinak direla egiaztatzen da; ohiko euskalkien mapa eta ikerketa honetako ondorio bezala lortu den euskalkien mapa desberdinak dira. Ez dute bat egiten ez euskalki kopuruan, ezta batzuetan euskalkien arteko mugetan ere. Euskalkietatik azpieuskalkietara pasatzen bagara, desberdintasunak handitu egiten dira.

Ikerketa honetan erabili den metodoaren abantaila handiena lortzen diren mugak mailakatzeko edo hierarkia batean kokatzeko gaitasuna da. Eta horretan konpontzen du ohiko dialektologiak zuen gabezia handiena. Muga guztiak ez dira maila eta pisu edo garrantzi berekoak, eta hori nabarmen argi gelditzen da azpieuskalkien izaera kontuan hartzean: ez da maila bereko muga ohiko bizkaieraren ekialde eta mendebaldeko banatzen duena eta, adibidez, ohiko gipuzkeraren ipar eta hegoaldeko azpidialektoak banatzen dituena. Horrek erakusten du sailkapen hierarkikoa egiteak ikaragarritzko garrantzia daukala, argi erakusten baitu mugen maila edo garrantzia.

Ohiko dialektologia gai izan da mugak non diren zehazteko, baina inola ere ez muga horien desberdintasuna zehazteko. Horretarako sailkapen hierarkikoa ezinbesteko teknika da, eta dialektologoak behar-beharrezko du erabiltzea bere lana zehatz egin behar badu.

5. Ondorioak

Lexikoa erabiliz, euskalkien sailkapen hierarkiko automatizatuen lehen emaitzak dira hemen aurkezten direnak. Eta agian euskal dialektologian datu lexikalak erabiliz burutzen den lehen euskalkien sailkapena dugu. Hala ere, azaldu behar da datu lexikalak erabili direla, hots, EHHako emaitzetan oinarritutako datuak erabiliz eta datuoi inolako ukiturik eman gabe. Datu lexikalak alfabeto ortografikoan emanak (beraz, alfabeto fonetikotik ortografikora iragaziz), baina bestelako doitzerik jasan gabe.

Abiapuntutzat hartzen den ekarri honek etorkizunean izango du segida, eta gramatikari dagozkion datuak erabiliz eta prozedura sailkatzaile berak erabiliz sortuko dira mapa berriak. Mapa hauek berriro erkatuko dira ohiko dialektologian lortutakoekin eta erkatzean lortzen diren berdintasun-ezberdintasunak aztertuko dira.

6. Erreferentzia bibliografikoak

- Aurrekoetxea, G. (1986). Euskal Herriko Hizkuntz Atlas (EHHA): inkesta metodologia eta ezezko datuak. *Euskera*, 31, 413-424.
- Aurrekoetxea, G. (2002). Algunas consideraciones sobre la contrapregunta en las encuestas lingüísticas. In L. Rabassa (ed.). *Mélanges offerts à Jean-Louis Fossat (57-65 orr.)*, Université de Toulouse II-Le Mirail, CerCLid 11/2. http://artxiker.ccsd.cnrs.fr/docs/00/07/14/00/PDF/Contrapregunta_encuestas_ling.pdf
- Aurrekoetxea, G. & Videgain, X. (1987). Euskal Herriko Hizkuntz Atlasaren ezaugarri nagusiak. In askoren artean, II. *Euskal Mundu Biltzarra*, 2, II, 57-462.
- Aurrekoetxea, G. (2012). Towards a scientific measurement of linguistics boundaries. In Á. Pérez, X. Alfonso, E. Carrilho & C. Magro (Eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr). Lisbon, 2011*, (23-34). Lisboa: Centro de Lingüística da Universidade de Lisboa. http://limiar.clul.ul.pt/proceedings_en.html
- Aurrekoetxea, G. (2016). Analysis of the morphological variation of Basque. *DiG*, 24, 21-41
- EHHA: Euskaltzaindia, 2010-2017, *Euskararen Herri Hizkeren Atlas*, I-VIII, Bilbo: Euskaltzaindia.
- DIATECH: <http://eudia.ehu.es/diatech/index/>

- Guitier, H. (1973). Atlas et frontières linguistiques. In G. Straka & P. Gardette (eds.) *Les dialectes romans de France à la lumière des atlas régionaux* (61-109). (col. de Strasbourg, 1971). Paris.
- Goebel, H. (1978). Analyse dialectométrique de quelques points de l' AIS (italien standard valdotain provençal alpin turinois milanais). In G.E. Clivio & G. Gasca Queiraza (arg.), *Lingue e dialetti nell'arco alpino occidentale. Atti del Convegno Internazionale di Torino (1976)*, (282-294). Turin.
- Nerbonne, J. (2008). Variation in the Aggregate: An Alternative Perspective for Variationist Linguistics. In K. Dekker, A. MacDonald, & H. Niebaum (Eds.), *Northern Voices: Essays on Old Germanic and Related Topics offered to Professor Tette Hofstra*, (365-383 orr.). Leuven: Peeters.
- Nerbonne, J., & Heeringa, W. (2010). Measuring dialect differences. In J. Erich Schmidt & P. Auer (arg.) *Language and Space: An International Handbook of Linguistic Variation, Volumen 1*, (550-565). Berlin: Mouton De Gruyter.
- Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane*, 35, 335-357.
- Zuazo, K. (1998). Euskalkiak, gaur. *FLV*, 78, 191-233.

1. Jardunaldietako kartela:

EUDIA-6 WORKSHOP
LINGUISTIC VARIATION IN THE BASQUE
LANGUAGE AND EDUCATION
Euskararen bariazioa eta bariazioaren
irakaskuntza



2017ko ekainaren 9a
9 de junio de 2017

Bilboko Irakasleen Unibertsitate Eskola
Escuela Universitaria de Magisterio de Bilbao
(UPV-EHU)

Antolatzaileak / Organizadores: UPV-EHUko EUDIA Ikerketa-taldea

Laguntzaileak / Colaboradores:

 Basque Language Laboratory	 Basque Language Laboratory
 Basque Language Laboratory	 Basque Language Laboratory

2. Jardunaldietako programa

Antolatzaileak / Organizadores



UPV-EHUko EUDIA Ikerketa-taldea

Laguntzaileak / Colaboradores



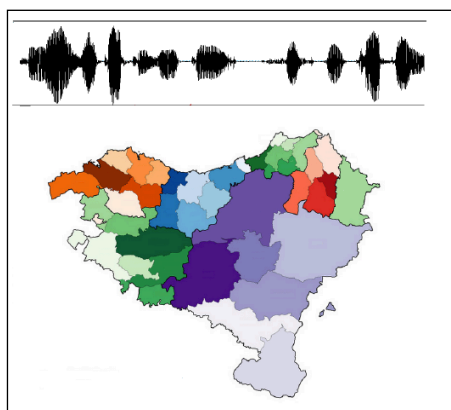
UPV EHU Bilboko Irakasleen Unibertsitate Eskola



UPV EHU Hizkuntza eta Literaturaren Didaktika Saila



UPV EHU Bizkaiko Campuseko Errektoreordetza



EUDIA-6 WORKSHOP

Linguistic variation in the Basque language and Education

Euskararen bariazioa eta bariazioaren irakaskuntza

2017ko ekainaren 9a
9 de junio de 2017

Bilboko Irakasleen Unibertsitate Eskola
Escuela Universitaria de Magisterio de Bilbao
(UPV-EHU)

EUDIA-6 WORKSHOP

Linguistic variation in the Basque language and Education

Euskararen bariazioa eta bariazioaren irakaskuntza

Egitaraua / Programa

Tokia / Lugar

UPV-EHUko Bilboko Irakasle Eskolan (Bilera gelan /Sala de reuniones)

- 9:00-9:15 Harrera eta kredentzialak hartzea / Recepción y entrega de credenciales
 9:15-9:30 Irekiera / Acto de inauguración
 9:30-10:30 Ponentzia (1) / Ponencia (1) (mahaiburua A. Etxebarria)
 Borràs-Comes, J. (Universitat Pompeu Fabra & U. Autònoma de Barcelona):
Metodología experimental y tratamiento estadístico: aplicación al estudio de la prosodia del catalán
- 10:30-11:30 Txostenak / Comunicaciones (mahaiburua I. De Pablo)
 Orbegozo, I., Berro, A. & Fernandez, B. (UPV-EHU): Euskal aldakortasun morfosintaktikoa erakusten duen sareko datu basea.
 Gaminde, I., Eguskiza, N., Romero, A. & Etxebarria, A. (UPV-EHU): Informatzaileen sailkapenerako arazo metodologiko batzuen gainean.
 Etxebarria, A., Eguskiza, N., Gaminde, I., Romero, A. (UPV-EHU): Ahots seduzitzaien ezaugarriak generoaren ikuspegitik.
- Eztabaida / Discusión
- 11:30-12:00 Atsedena / Descanso
- 12:00-13:00 Txostenak / Comunicaciones (mahaiburua X. Etchebest)
 Beristain, A. (University of Illinois): The Acoustic Realization of /s/ and /tʃ/ by L1 and L2 Basque Speakers.
 Iglesias, A., Eguskiza, N., Gaminde, I. & Unamuno, L. (UPV-EHU): Igorreko azentuaren korrelatu akustikoez.
 Atutxa, U., Ansa, O., Iruskietza, M. & Molina, A. (UPV-EHU): Compress-EUS: i(ra)kasleen laburpenak lortzeko tresna.
 Eztabaida / Discusión
- 13:00-15:00 Bazkaria / Comida
 15:00-16:00 Ponentzia (2) / Ponencia (2) (mahaiburua A. Romero)
 Murillo, E. (Universidad Autónoma de Madrid):
Acciones, gestos y palabras: el desarrollo multimodal del lenguaje en la etapa infantil.
- 16:00-17:00 Txostenak / Comunicaciones (mahaiburua A. Iglesias)
 Etchebest, X. (UPV-EHU & UPPA): Zubererazko Dardarkarien azterketa sozio-fonetikoa.
 Eguskiza, N., Gaminde, I., Olalde, A., Etxebarria, A. & Gaminde, U. (UPV-EHU): Larrabetzuko euskaldun berrien seme-alaba euskaldun zaharren ezaugarri prosodikoez.
 de Pablo, I., Romero, A., Etxebarria, A. & Gaminde, I. (UPV-EHU): Metodologia ezberdinak haur euskaldunen komunikazio-garapena neurtzeko.
- Eztabaida / Discusión
- 17:00-17:30 Atsedena / Descanso
 17:30-19:00 Txostenak / Comunicaciones (mahaiburua N. Eguskiza)
 Videgain, X., Aurrekoetxea, G. & Ormaetxea, J.L. (UPV-EHU & UPPA): Euskalkien sailkapen zientifikoa: lexikoa.
 Khetitib, S. (University of Oran 2): Metacognition in enhancing students' productive skills.
 Eztabaida / Discusión
- 19:15 Mintegiaren amaiera / Acto de clausura



EUDIA Ikerketa-taldea

3. Argazkiak:

Irekiera / Acto de inauguración

Jon Zarate (UPV/EHU Euskararen arloko errektoreordea), Gurutze Ezkurdia (Bilboko Irakasleen Unibertsitate Eskolako zuzendaria), Kepa Larrea (Hizkuntzaren eta Literaturaren Didaktika Sailako koordinatzailea) eta Iñaki Gaminde (EUDIA ikerketa taldeko ikertzaile nagusia)



Argazki orokorrak





PONENTEAK:

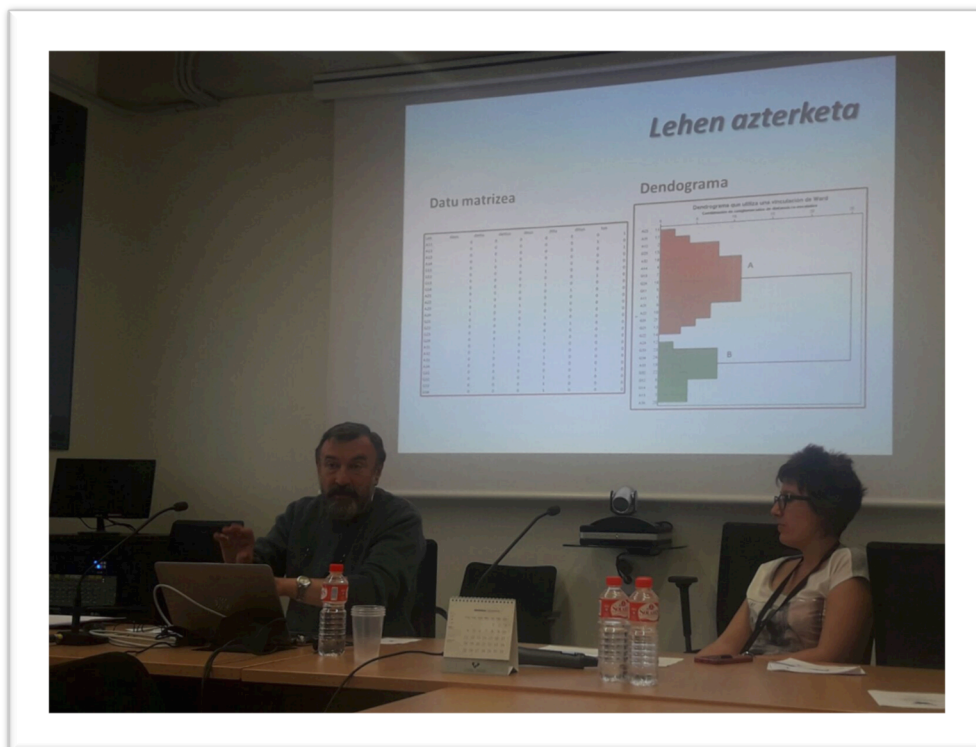
Borràs-Comes, J. (Universitat Pompeu Fabra & U. Autònoma de Barcelona): Metodología experimental y tratamiento estadístico: aplicación al estudio de la prosodia del catalán.



Orbegozo, I., Berro, A. & Fernandez, B. (UPV-EHU): Euskal aldakortasun morfosintaktikoa erakusten duen sareko datu basea.



Gaminde, I., Eguskiza, N., Romero, A. & Etxebarria, A. (UPV-EHU): Informatzaileen sailkapenerako arazo metodologiko batzuen gainean. azentuaren korrelatu akustikoez.



Etxebarria, A., Eguskiza, N., Gaminde, I., Romero, A. (UPV-EHU): Ahots seduzitzaileen ezaugarriak generoaren ikuspegitik.



Beristain, A. (University of Illinois): The Acoustic Realization of /s/ and /ts/ by L1 and L2 Basque Speakers.



Iglesias, A., Eguskiza, N., Gaminde, I. & Unamuno, L. (UPV-EHU): Igorreko azentuaren korrelatu akustikoez.



Ansa, O., Atutxa, U., Iruskieta, M. & Molina, A. (UPV-EHU): Compress-EUS: i(ra)kasleen laburpenak lortzeko tresna.



Murillo, E. (Universidad Autónoma de Madrid): Acciones, gestos y palabras: el desarrollo multimodal del lenguaje en la etapa infantil.



Etchebest, X. (UPV-EHU & UPPA): Zubererazko Dardarkarien azterketa soziofonetikoa.



Eguskiza, N., Gaminde, I., Olalde, A., Etxebarria, A. & Gaminde, U. (UPV-EHU): Larrabetzuko euskaldun berrien seme-alaba euskaldun zaharren ezaugarri prosodikoez.



de Pablo, I., Romero, A., Etxebarria, A. & Gaminde, I. (UPV-EHU): Metodologia ezberdinak haur euskaldunen komunikazio-garapena neurtzeko.



Videgain, X., Aurrekoetxea, G. & Ormaetxea, J.L. (UPV-EHU & UPPA): Euskalkien sailkapen zientifikoa: lexikoa.

