

Sociolinguistic and Geolinguistic Variation in the Basque language¹

GOTZON AURREKOETXEA

*University of the Basque Country (UPV-EHU),
/Faculty of Letters, Unibertsitatearen Ibilbidea, E01006 VITORIA-GASTEIZ/
gotzon.aurrekoetxea@ehu.es*

SCN III/1 [2010], 88–100

V prispevku so predstavljeni prvi rezultati *Socio- in geolingvističnega atlasa baskovskega jezika – EAS*. Raziskovalni projekt je bil predstavljen na 5. konferenci SIDG, ki je bila v Bragi. Predstavljamo nekatere podatke, zbrane na isti lokaciji pri informatorjih različnih generacij (starejša, srednja, mlajša). Raziskujemo jezikovne razlike, ki se pojavljajo med govorci teh generacij, in analiziramo, ali so medgeneracijske razlike enake na vseh lokacijah ali pa se med njimi pojavljajo kakšne posebnosti. Prikazujemo tudi različne tipe geolingvističnih variacij med generacijami.

In this contribution, we show the first results of the “Socio-geolinguistic atlas of the Basque language-EAS”. This research project was presented at the Vth Congress of Dialectology and Geolinguistics held in Braga. On the one hand, we present some data collected in the same locality from informants of different generations (the elderly, adults and young people). We examine the linguistic differences that we found among people of these generations and we analyse whether the difference among generations is similar in all localities or not, or whether some of them have peculiar characteristics. On the other hand, we show different types of geolinguistic variation: variation among elderly people, adults or among young people.

Ključne besede: sociolingvistična variacija, geolingvistična variacija, jezikovni korpus, baskovska dialektologija

Key words: sociolinguistic variation, geolinguistic variation, corpus linguistics, Basque dialectology

¹ This project has received financial support from the Ministerio de Ciencia e Innovación of the Spanish Government (HUM2007-65094/FILO) for the following three years (2007–2010).

0 Introduction

In spite of the fact that the Basque language has had active production of literature since the 16th century, all of the Basque literature has been written in dialect. During all that time, four literary dialects have been developed. It was in 1968 when the Academy of the Basque Language decided to create a standard Basque language (the ‘unified’ *batua*, as it is called). The standardisation of the Basque language is very recent being only 40 years old. The usage of standard Basque started with education in the Basque language in the 1970s, first in the *ikastola* (‘monolingual Basque education created by private initiative’), later in bilingual public schools and finally in private schools. Media outlets in Basque (public and private channels) were created in the eighties.

With the creation of the standard variety and its use in education, in media, etc. the dialects have, on the one hand, fallen more and more into disuse in formal relationships and businesses. In the last 30–35 years, more than half of the young people have been educated in the Basque standard language (only in Basque, or in Basque and Spanish or French). People older than 40–45 were educated only in Spanish or French, and they did not learn Basque; even if they are Basque speakers they did not learn to write and to read in Basque.

On the other hand, with the growing use of the standard variety there has been an increase of levelling phenomena between dialects and the standard variety. Dialects experience a double situation: a **diglossic** situation and a **diaglossic** one. We say ‘diglossic situation’ because the use of dialect is looked down on and its use in the media and culture is more or less prohibited or not advised. That is to say, the standard has become the prestigious variety and the dialects have been moved away from the cultural and public life. Moreover, dialects have taken shelter in family life, and they are only used in private life. Dialects know also a diaglossic situation, in the sense used by Bellmann (1998: 24); that is to say, “... dialects, as part of the substandard, are getting more or less deprived of their features, especially of their primary ones”. There is a large amount of people speaking dialect and standard varieties in different everyday contexts: people who only speak the standard variety, even in the informal contexts such as in familiar conversations, etc., and people who do not speak the standard variety (they are not able to speak it because they have not learned it, but they listen to the standard variety every day on the TV, radio or even in the street). Even in familiar conversations, people use some features of the standard variety, when they are speaking in dialect.

Since the standard variety was created, all dialects have changed and in all of them we can notice cases of levelling, above all among young people, but even among adults, who decided to educate their children in the Basque language, taking active part in the revival of the language.

In this context, with diglossic and diaglossic situations of the language, with the increasing energy of the standard variety and a levelling phenomena spread through the entire territory of the Basque language, it is very important to analyse the sociolinguistic situation from the variationist point of view.

The objective of this paper is to show the first results of our researches about the sociolinguistic variation that is happening nowadays in the Basque language, taking into account only one factor: speaker age. For that, we will study linguistic variation between three generations: the elderly (more than 75 years old), adults (average of 45 years old) and young people (around 25 years old).

1 The Corpus of Oral Dialectal Basque (EDAK)

The data for this paper were gathered in the EDAK corpus. This corpus has been created using data collected in two projects: the first one is called Sociolinguistic atlas of the Basque language-EAS, which provides phonological, morphological, syntactic and lexical data, and the second one is EDAK, that provides only prosodic data (stress and intonation). Both sets of data form part of the EDAK corpus². This corpus collected data from 100 localities of the Basque Country. In each locality we gathered information of different generations. These two projects have two parts: the geolinguistic part (different generations) and the sociolinguistic one (data collected in 100 localities from the Basque Country). For that, we use only one informant for each generation. We know that the use of only one informant for each locality is usual in geolinguistic, but it is not in sociolinguistic studies. For that, this study has to be considered as the first step in the sociolinguistic way to study the variation in the Basque language.

The sources of this paper are data that come from the EAS project. This project collects data in each locality employing a questionnaire of 201 questions: 120 questions of lexicon (the words to say ‘open’, ‘rust’ etc.), 23 questions about noun morphology (all the questions are about flexion morphology), 39 about verb morphology (simple and compound verbs) and 20 about syntax (agreement between the verb and the direct object, etc.).

This contribution is a compilation of all of the papers that have been published or that have been accepted to be published. We have selected investigations carried out in 8 localities (see Fig. 1): Dima (Aurrekoetxea 2008), Bolibar and Etxebarria (Ezenarro 2008), Otxandio (Ormaetxea 2008), Aramaio (Ormaetxea 2009), Gizaburuaga (Unamuno 2010), Zegama, and Luzaide (Santazilia 2010). There is no other work published about sociolinguistic variation in the Basque language, using EDAK corpus.

These papers constitute the first step of our research. We would like to measure the linguistic variation between generations; this constitutes research in apparent time. We compare three generations: elderly people who do not have any contact with the standard variety, adults who have weak contact and young people who have been educated in the standard variety. Taking into account

² For more information about EAS see Aurrekoetxea & Ormaetxea, 2006; for EDAK see Aurrekoetxea / Sánchez / Odriozola, 2009.

that the EDAK corpus has data from 100 localities, the studies discussed here account for 8 % of the total data.

It is also important to take into account the location of the villages where the data reference in this paper was gathered: there are 6 localities belonging to the dialect of Biscay, which is one of the most remote varieties from the standard, one locality belonging to the Guipuscoan dialect, which is one of the closest varieties to the standard, and one situated in the frontier between France and Spain, which shows different sociolinguistic situations (elderly people have more social ties to people of the north part of the territory, belonging to France and the Basse Navarre dialect, whereas young people have more social ties to the south).

We do not show a large amount of data and they will not furnish us with a definitive pattern of the current intergenerational linguistic variation of Basque. But they allow us to see some of the most important tendencies of the present-day sociolinguistic situation that we can find in different parts of the Basque Country.

This contribution is only the first step in analysing the most important sociolinguistic parameters that are influencing variation in the Basque language today.



Fig. 1: map of the localities of the EDAK and the localities chosen for this paper

2 Linguistic variation across generations

It is evident that the missing of the specific linguistic features of the base dialect has to do with the levelling phenomena. The Basque language is experiencing this phenomenon nowadays. We understand that the disappearance of features from the base dialect in young people is synonymous of the loss of dialectalism, and it shows the loss of dialectal dynamism. In this context, we will analyse the dialectalism of each locality comparing speakers from three generations. It will be shown that in the sociolinguistic situation of the Basque language we find many cases of dialect levelling. And we will find out which are the linguistic features that have a tendency to change or change easiest among generations.

2.1 Differences across generations

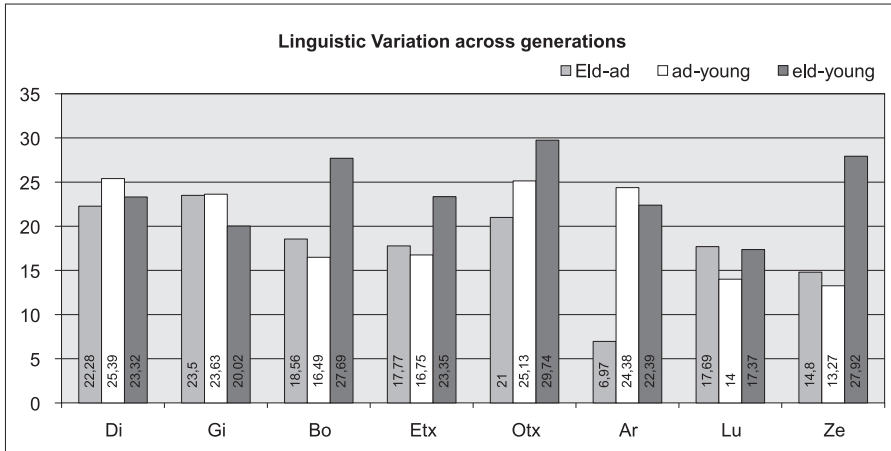


Fig. 2: Variation across generations

This diagram shows many linguistic differences, in percentages: the blue column shows differences between elderly people and adults, the orange column presents differences between adults and young people and the yellow one shows differences between elderly and young people. It takes in all the differences we have found among localities (phonological, morphological, syntactic and lexical differences). Therefore, the diagram shows the total difference we have found in the EDAK corpus.

The diagram points out that the linguistic variation is not the same in all localities³; each locality has a different pattern.

On the one hand, our first hypothesis was that, on the whole, linguistic variation would increase across generations: that is to say, there should be less variation between elderly people and adults than between elderly and young people. On the other hand, we hypothesized that linguistic difference between elderly people and adults would be greater than between adults and young people.

According to the data, and with regard to the variation between the elderly and adult-young people, our hypothesis has been confirmed. In fact, in only one locality (Luzaide) is the variation between the elderly and young people slightly lower than between the elderly and adults (17,39 % and 17,69 % respectively), whereas in the other localities (Dima, Bolibar, Etxebarria, Otxandio, Aramaio and Zegama) it is greater or far greater: the difference between the two percentages is runs between 13,12 % in Zegama and 1,04 % in Dima.

If we regard the variation between adults and young people, that variation between adults and young people is greater than in the other two generational

³ The localities are abbreviated as following: Dima (Di), Gizaburuaga (Gi), Bolibar (Bo), Etxebarria (Etx), Otxandio (Otx), Aramaio (Ar), Luzaide (Lu) and Zegama (Ze).

comparisons in Dima and Aramaio; it is lower than the other two comparisons in Bolibar, Etxebarria and Zegama. These data show different patterns: in some of them there is more variation than in others. In some of them there is a critical break between generations; one of the best examples is Zegama: there is a difference of 13,89 % between two older generations (average data from these two generations) and young people. We can speak in similar way about Aramaio (the difference between the elderly, and adult-young people (average data from these two generations) is 16,45 %). These are the most critical breaks between generations. But these differences have distinct characteristics: in the first case it is the young generation which moves away from the older generations, and in the second one it is the elderly generation that moves away from the other two.

In the following subsections we will analyse the differences between two generations at a time:

2.1.1 Elderly people and adults

Analysing thoroughly the figure 3 we can see the differences between two generations better.

We will begin analysing the differences between old people and adult generations.

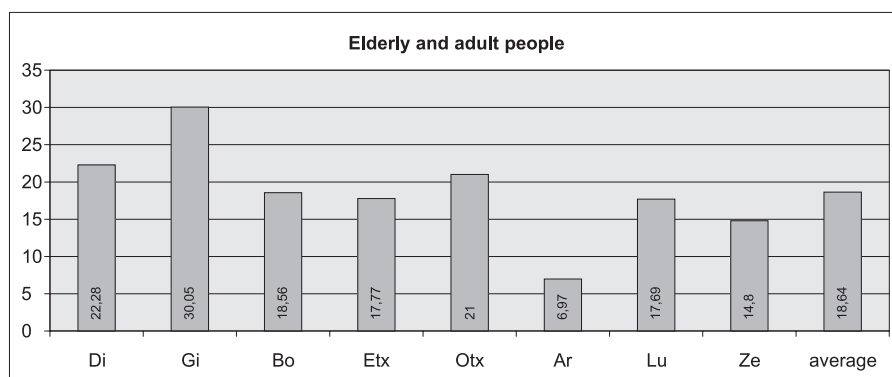


Fig. 3: Percentage of linguistic differences between the elderly and adult people

Although the majority of the scores is situated between 30,05 (Gizaburuaga) and 14,80 (Zegama), there is only one locality that scores outside this range (Aramaio which has the lowest).

We can also see that the first five localities located in the Biscayan dialect (in the West part of the territory of Basque language) have clearer differences than the localities located in the middle (Zegama, in the Guipuscoan dialect and Luzaide, in the Basse navarre dialect). The average of the differences across all localities is 18,64.

2.1.2 Adults and young people

If we take the adult and young generations, the differences are also greater in the localities located in the West than in the localities in the East (Zegama).

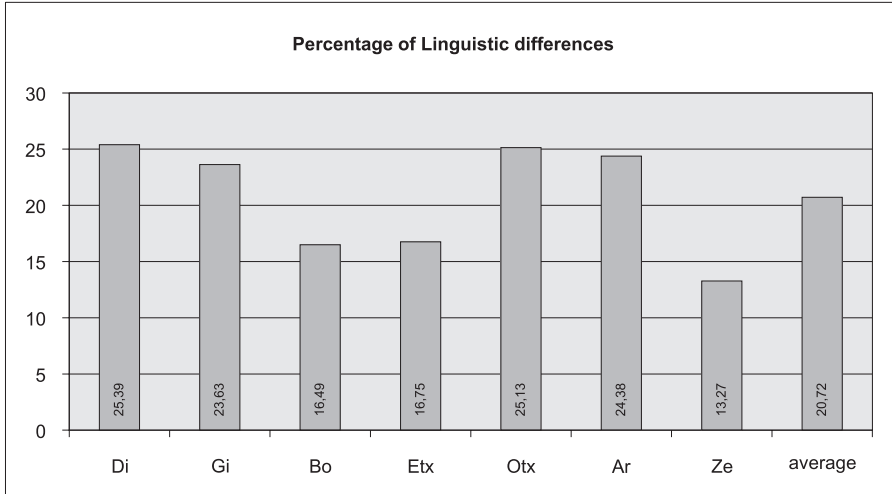


Fig. 4: Percentage of the linguistic differences between adults and young people

On the whole, the differences between adults and young people are a bit bigger (20,72 %) than between the elderly and adult generations (18,64 %).

2.1.3 Elderly people and young people

If we take the elderly and the young generation, obviously the differences get bigger, but they are not the sum of the differences of elderly-adult and adult-young generations.



Fig. 5: Linguistic differences between elderly and young people

All differences are over 16,11 % and the average is situated at 23,82 %.

At this moment we must emphasize the importance of the coexistence of different linguistic features in the linguistic system of any locality. We can hear in lexic, for instance, in Zegama that old people say *erruki* for ‘pity or compassion’, adults say *erruki* and *pena*, and young people say only *pena*. Or in Bolibar old people say *junio* and *bagill* for ‘Jun’, adults say *junio*, and young people *junixo* and *ekañe*. In this case, the autochthonous form *bagill* has disappeared, adults use the loan form *junio* and young people use two forms: *junio* and the standard form *ekañe*, learned in school.

Although it is very interesting we will not mention here the variation across other linguistic parameters that has been taken into account in the figures. In other words, we will not show which parameter has more variation (noun morphology, verb morphology, syntax or lexicon).

2.2 Diatopic variation among generations

2.2.1 Diatopic variation among the elderly people

If we analyse the diatopic variation among the elderly people and the young generation (that is to say, the diatopic variation of two generations), we will have different pictures: we will have four diagrams for each generation, (one for each linguistic field: lexicon, noun morphology, verb morphology and syntax).

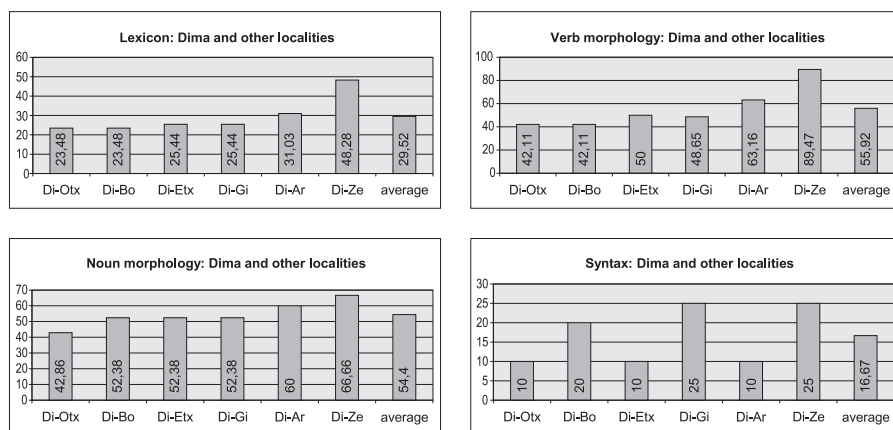


Fig. 6: Relationship between Dima and other localities in the elderly generation⁴

⁴ Di-Otx (between Dima and Otxandio), Di-Bo (Dima and Bolibar), Di-Etx (Dima and Etxebarri), Di-Gi (Dima and Gizaburuaga), Di-Ar (Dima and Aramaio), Di-Ze (Dima and Zegama).

In the elderly generation the linguistic variation between Dima and the other localities is located between 10 % (Dima and Otxandio, Dima and Etxebarri, and Dima-Aramaio, all of them in syntax), and 89,47 % (between Dima and Zegama in verb morphology).

The graphics show, it is obvious, that Dima has a closer relationship with localities located in the same dialect than with localities situated in different dialects (Zegama).

The linguistic parameter that shows the most significant difference is verb morphology (55,92 %) followed by noun morphology (54,44 %), and the least is syntax (16,67 %). The lexical difference is situated at 29,52 %.

2.2.2 Diatopic variation in the young generation

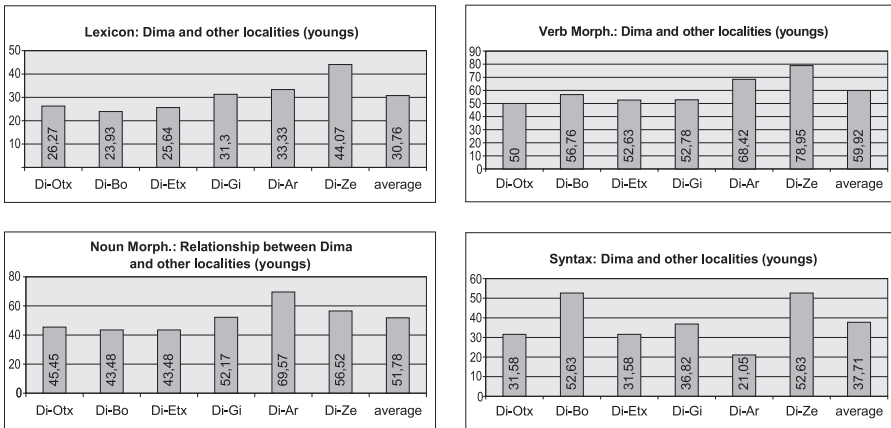


Fig. 7: Relationship between Dima and other localities in the young generation

In the young generation, the linguistic variation between Dima and the other localities is located between 21,05 % (Dima and Aramaio in syntax parameter), and 78,95 % (between Dima and Zegama in verb morphology). In both cases — elderly and young people— the least difference is situated in syntax parameter (between Dima and Aramaio, even if it is not the nearest locality, but located in the same dialect), and we find the highest difference in verb morphology (between Dima and Zegama).

On the whole, the graphics of the young generation show also that Dima has a closer relationship with localities located in the same dialect than with localities situated in different dialects (Zegama), except in noun morphology, in which Aramaio has more differences than Zegama; and in syntax, in which Bolibar has the same difference as Zegama.

The linguistic parameter that shows the greatest difference is verb morphology (it shows 59,92 %) followed by noun morphology (51,78 %), and the least, in this case, is the lexical difference (30,76 %). The syntax difference is 37,72 %.

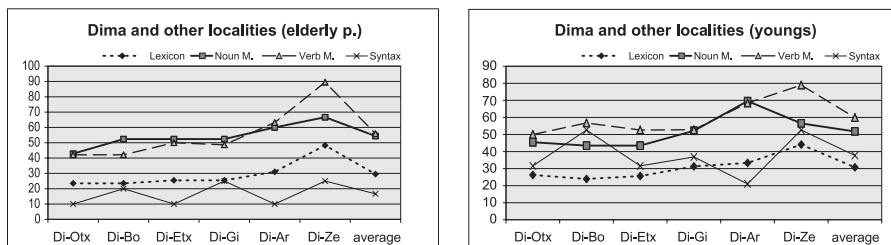


Fig. 8: Relationship between Dima and other localities (the elderly and the young)

Looking at both graphics (Fig. 8), we can see the linguistic differences among different localities better, taking into account data from both generations. The greatest difference is produced between Dima and Zegama (except Dima and Aramaio in noun morphology, and Dima and Bolibar in syntax, in the young generation). And fewest linguistic differences are located between Dima and Otxandio among elder people, and between Dima and Etxebarria in the young generation.

It is the syntax parameter that shows the most irregular variation. The other parameters show a lineal differentiation, growing from Otxandio to Zegama in two generations (except for verb morphology in young generation).

The following picture shows the linguistic differences between elderly and young generations from the diatopic point of view (Fig. 9). We have obtained it summarizing the averages of fig. 6 and 7.

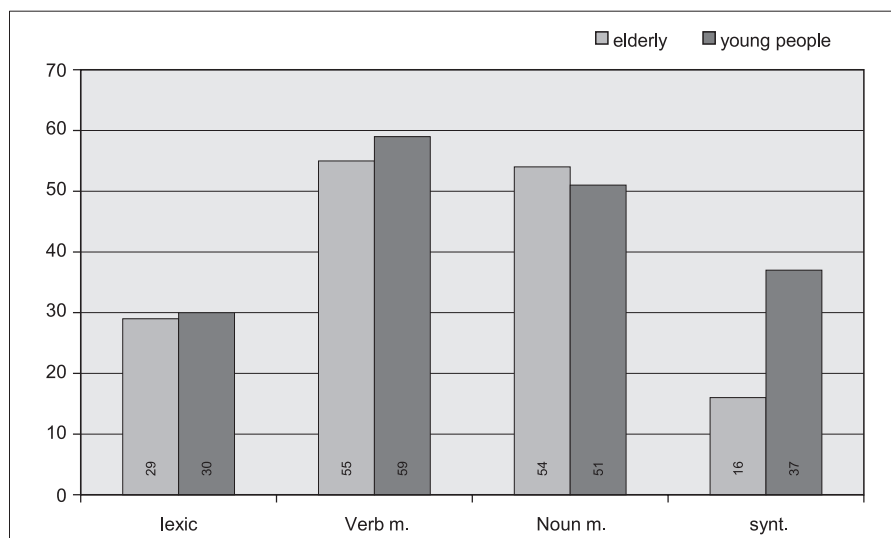


Fig. 9: Diatopic differences among generations

As we can see, it is the verb morphology that shows greatest linguistic difference between Dima and other localities analysed, both among the elderly

(55 %) and among young people (59 %). The noun morphology shows also great differences (54 % in the elderly and 51 % in young people). The least linguistic difference is produced by syntax parameter (16 % in the elderly and 37 % in young people).

2.3 Diatopic differences between generations

Considering all the differences among all the localities, and comparing the elderly generation with young people, we will find the following differences (Fig. 10):

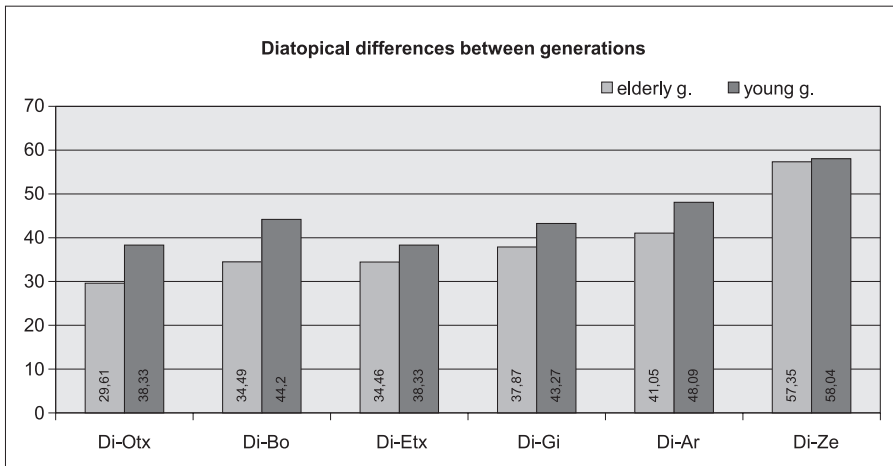


Fig. 10: Diatopic differences between generations taking into account all features

We find greater linguistic differences between young people than between elderly people in all the relationships between Dima and other localities, using all the linguistic parameters.

We supposed that the difference between young people would be lower than the difference between old people.

According to these data, we can determine the percentage of dialect levelling that occurs in these localities: young generation uses more standard features than adults and the elderly people. But these standard features are not the same in all localities; each one takes different features. For instance, in Etxebarri young people say *udazken* (standard) and elderly people *udagoen* (becoming to base dialect); however in Bolibar young people use *junio*, but not *ekain* (standard).

3 Conclusions

We have shown the linguistic differences found between Dima and other localities situated both in the same dialect (Gizaburuaga, Bolibar, Etxebarri, Otxandio and Aramaio) and in different dialects (Zegama and Luzaide), using data from EDAK corpus.

On the one hand, and with regard to the linguistic differences among the elderly and adults, the scores are situated between 23,50 and 14,80. These scores are higher in the West part than in the middle part. And if we regard to the differences between adults and young people, they are also greater in the localities located in the West than in the localities in the middle (Zegama and Luzaide). The localities in the West have dialects that differ more from the standard varieties. Consequently, we have to say that if the differences between the standard and the dialect are big, the differences among generations are bigger.

On the other hand, and taking into account the pattern of the variation, in our data we have not found any pattern of sociolinguistic variation between generations: in some localities there is a break between the elderly and the adults, in others between the adults and the young people. However, we can see that verb and noun morphology produce more variation than lexicon and syntax, according to data of EDAK corpus.

And finally, if we compare Dima locality with the others we can see that the linguistic differences are bigger in the young generation than in the elderly, using all the linguistic parameters.

REFERENCES

- Borja ARIZTIMUÑO, 2010: Tolosako eta Ataungo hizkerak: hizkuntz bariazioa eta konbergentzia joerak [Linguistic Variation and levelling in Tolosa and Ataun dialects], *Uztaro* 72, 79–96.
- Gotzon AURREKOETXEA, 2008: Bariazio soziolinguistikoa Dimako euskaran [Sociolinguistic Variation in Dima Dialect]. *Euskalingua* 12, 17–26 [on-line version: www.mendebalde.com].
- Gotzon AURREKOETXEA & Jose Luis ORMAETXEA, 2006: Research project – “Socio-geolinguistic atlas of the Basque language”. *Euskalingua* 9, 157–163 [on-line version: www.mendebalde.com].
- Gotzon AURREKOETXEA / Jon SÁNCHEZ / Igor ODRIOZOLA, 2009: EDAK: A Corpus to Analyse Linguistic Variation. *A Survey on Corpus-based Research Panorama de investigaciones basadas en corpus*. Ed. Pascual Cantos Gómez / Antonio Sánchez Pérez. Murcia: Asociación Española de Lingüística del Corpus. 489–503.
- Günter BELLMANN, 1998: Between Base Dialect and Standard Language. *Folia Linguistica* XXXII/1–2, 23–34.

Amaia EZENARRO, 2008: Etxebarria eta Bolibarko bariazio linguistikoa [Linguistic Variation in Etxebarria and Bolibar]. *Uztaro* 67, 59–84 [on-line version: www.uztaro.com].

Jose Luis ORMAETXEA, 2008: Otxandioko hizkera: adinaren araberako bariazioa [Dialect of Otxandio: Age Variation]. *Fontes Linguae Vasconum* 108, 249–262.

– –, 2009: Linguistic differences between generations in Basque. Paper presented at ICLaVE 5 (Copenhagen, 2009).

Ekaitz SANTAIZILIA, 2010: Luzaideko hizkuntz bariazioa [Linguistic Variation in Luzaide]. *Fontes Linguae Vasconum* 111, 219–248.

Lorea UNAMUNO, 2010: Adinaren araberako bariazioa Gizaburuagako hizkeran [Age Variation in Dialect of Gizaburuaga]. *Euskalingua* 16, 41–48 [on-line version: www.mendebalde.com].

SOCIO- IN GEOLINGVISTIČNA VARIACIJA BASKOVSKEGA JEZIKA

Raziskovalni projekt *Socio- in geolingvistični atlas baskovskega jezika – EAS* zbira podatke različnih generacij na 100 lokacijah baskovskega jezikovnega področja. Delo se je pričelo leta 2007 in bo končano 2010 – prikazuje podatke z različnih lokacij, primerja podatke različnih generacij.

Čprav je bil baskovski jezik standardiziran šele pred kratkim, njegova raba zelo hitro narašča, tako da se danes že večina Baskov izobražuje v standardnem (baskovskem knjižnem) jeziku. Projekt raziskuje diglosična in diaglosična razmerja med baskovskim in španskim jezikom ter med narečji in normo, hkrati pa pripravlja korpus EDAK (Korpus baskovskih ustnih in dialektoloških inačic), ki oblikoslovne, skladijske in besedne informacije prejema od projekta EAS, prozodične informacije pa nastajajo v projektu EDAK.

V članku predstavljamo podatke z osmih lokacij in analiziramo jezikovne variacije generacij iste lokacije, diatopične variacije med generacijami (med starejšo in mlajšo generacijo) ter narečno izenačevanje.

Podatki kažejo, da je jezikovna variacija med starejšo in srednjo generacijo večja kot med srednjo in mlajšo generacijo. Glede na diatopično medgeneracijsko različico je več jezikovnih razlik med mladimi kot med starejšimi, in sicer v vseh odnosih med področjem Dima in drugimi analiziranimi lokacijami.