CORPUS BASED PROSODIC VARIATION IN BASQUE:
Y/N QUESTIONS MARKED WITH THE PARTICLE AL

VARIACIÓN PROSÓDICA EN VASCO BASADA EN CORPUS:
PREGUNTAS SÍ/NO MARCADAS CON LA PARTÍCULA AL

GOTZON AURREKOETXEA  
Universidad del País Vasco  
gotzon.aurrekoetxea@ehu.es

INAKI GAMINDE  
Universidad del País Vasco  
inaki.gaminde@gmail.com

AITOR IGLESIAS  
Universidad del País Vasco  
txadonak@gmail.com

Artículo recibido el día: 8/11/2010
Artículo aceptado definitivamente el día: 13/07/2011
Estudios de Fonética Experimental, ISSN 1575-5533, XX, 2011, pp. 11-31
ABSTRACT

This paper describes the intonational variation between two generations in three different localities of the Basque Country, using data recorded and organized in the EDAK corpus (Dialectal Oral Corpus of the Basque Language) and analysing the usage of just one type of sentence, namely y/n questions. In the selected localities, there are two morphological ways of constructing this type of sentence: using the morphological marker al before the verb (or between the verb and the auxiliary) or not using it. First, we identify the phonological patterns that exist in these localities. Then, we analyse intergenerational variation, according to the five different phonological patterns found. Finally, we study the geo-prosodic variation which exists between older and younger people from these localities.

Keywords: linguistic corpus, Basque language, prosody, socio-prosodic variation, geo-prosodic variation.

RESUMEN

El artículo trata la variación entonacional entre dos generaciones en tres localidades distintas situadas en el centro del espacio lingüístico vasco en frases interrogativas absolutas con datos del EDAK (Corpus dialectal oral del euskera). En dichas localidades hay dos posibilidades morfológicas para construir esta clase de oraciones: usando la marca interrogativa al delante del verbo (o entre el verbo y el auxiliar) o sin ella. En el análisis primeramente se han determinado los modelos fonológicos entonativos de las localidades en esta clase de oraciones. En segundo lugar se ha analizado la variación intergeneracional. Y finalmente, se ha analizado la variación geoprosódica entre dos generaciones de dichas localidades.

Palabras clave: corpus lingüístico, vasco, prosodia, variación socio-prosódica, variación geo-prosódica.
1. INTRODUCTION

This paper aims to show the prosodic variation that exists in several localities of the Basque Country, using one type of sentence: y/n questions. In order to do this, we use the Dialectal oral corpus of the Basque language or EDAK (Aurrekoetxea, Sánchez and Odriozola 2009). This research was carried out by the EUDIA research team, located at the University of the Basque Country (UPV-EHU). The current project has two aims: first, to obtain basic prosodic information from all varieties of the Basque language, and second, to complete the first study of sociolinguistic variation in Basque, using recorded data from two generations in each locality. To date, there are no studies on sociolinguistic variation in Basque other than the partial ones by Aurrekoetxea (2008 and 2010) and Ormaetxea (2008, 2011), and none on prosodic sociolinguistic variation. The EDAK corpus constitutes the first step to explore the basic principles of this kind of variation, and hence should be used as the basis for future work in this field.

This paper is organized into five main sections following the introduction: in section 2, we look at the strategies employed in Basque to form y/n questions; section 3 deals with the methodology used to gather information; in section 4 we provide a phonological analysis of the data; in sections 5 and 6 respectively, we analyse the socio-prosodic and the geo-prosodic variation in each of the dialects; and finally, we draw conclusions and make several proposals for future research.

2. Y/N QUESTIONS IN BASQUE

It is known that the languages of the world employ different strategies for constructing y/n interrogative sentences (Dryer 2005). These kinds of sentence have been one of the most often-analysed items of intonation research in neighbouring languages, including Catalan (Carrera et al. 2004; Fernández Planas et al. 2007; Martínez Celdrán et al. 2005; Pradilla and Prieto 2002; etc.), Galician (Fernández Rei et al. 2005) and Spanish (Dorta and Hernández 2005). The intonation of y/n questions has also been studied in Basque (Aurrekoetxea and Iglesias 2010; Gaminde 2004, 2006, 2010; etc.). A number of research teams and projects (GrEP¹, AMPER², etc.) have also taken them into account in their studies.

¹ http://prosodia.upf.edu/home/en/index.php
² Contini et al. 2002.

EFE, ISSN 1575-5533, XX, 2011, pp. 11-31
Apart from intonation contours, there are other strategies in Basque for building y/n questions, namely the use of interrogative markers. In some varieties, the particle *al* is used before the verb or between the verb and auxiliary in periphrastic constructions. A different strategy is to add the suffix -a to the verb or to the auxiliary in periphrastic forms. From a morphological point of view, there are thus three possible types of y/n questions: morphologically unmarked questions, questions marked with the particle *al* and questions marked with the suffix -a.

In this paper we test the hypothesis which states that the intonation curve is different in questions which use only intonation tools and in questions which use a morphological marker.

### 3. METHODOLOGY

#### 3.1. Questionnaire, localities and informants

The questionnaire used for the EDAK project has 201 questions, of which 23 deal with questions of intonation, and out of these 23, 6 concern y/n questions. The questionnaire was created in line with the existing literature on Basque prosody. The aim of the corpus is to present material gathered from two generations in 100 localities of the Basque Country, selected after a thorough analysis of dialectal variation in Basque, and including the whole of the Basque-speaking territory on both sides of the Spain/France political border. This is the first step for analysing sociolinguistic variation in the Basque language. The research treats all of the Basque Country as one unit and gathers information from only one person from each generation in each locality. We are aware that this does not strictly satisfy the number of informants required by sociolinguistic research, but it allows us to isolate the first preliminary features, which must subsequently be developed in future research. We took into account the weakness of the corpus on this point, and took great care both in the selection of the informants and during the interviews themselves in order to gather appropriate data from each generation.

---

3 In the sentences analysed in this paper, the particle *al* is always placed between the main verb and the auxiliary.

4 There are phonological rules for the combination of the suffix -a with the last vowel of the auxiliary verb: da + -a > dea, dü + -a >dia, etc.
To elicit the data, a member of the research team read each sentence in Spanish or French and asked the informant to translate it orally into his or her usual variety of Basque. This part of the questionnaire, related to intonation, was repeated three times, but at different moments during the interview, following the methodology of the AMPER project (Contini et al. 2002).

The research team tried to obtain an unbiased attitude from the informants when doing the questions. Although the interview is a formal task, informants were asked to speak as they would in a normal conversation with their friends. Informants were explicitly asked to employ their usual variety of Basque, without thinking about other ways of saying the same thing in other dialectal varieties or in the standard variety.

3.2. Data gathering, sound annotation and labelling, and storage

For data gathering and sound recording, laptops equipped with Audacity software\(^5\) and USB microphones (PC Headset 960 USB)\(^6\) were used.

The annotation task was carried out using the SFSWin software program\(^7\). Although this tool offers the option of automatic annotation, we opted for manual annotation because of the structure of the data gathered.

The output of the SFSWin program is a text file which connects the annotations to each question with their sound location in the recording. Once the sound had been annotated, we carried out the labelling task using the “txertatu_etiketak” ActivePerl script\(^8\). This script was adapted by Aholab, the Signal Processing Laboratory (http://aholab.ehu.es) of the University of the Basque Country (UPV-EHU), which we work with.

The data storage was carried out using two formats: TEI and MySQL. Each of these formats has tools for performing different analyses. The change from one format to the other is made automatically.

---


\(^6\) [www.logitech.com](http://www.logitech.com)

\(^7\) [http://www.phon.ucl.ac.uk/resource/sfs/download.htm](http://www.phon.ucl.ac.uk/resource/sfs/download.htm)

\(^8\) [http://www.activestate.com/](http://www.activestate.com/)
The following step was the acoustic analysis of the data. For this task we used Praat\(^9\). The Segment-data script\(^{10}\) is used to extract quantitative information automatically, so it can be used in quantitative analyses.

In this paper we examine the following 6 y/n questions:

1. Has he/she come [Etorri da]?
2. Has he/she bought it [Erosi dau]?
3. Has the friend arrived [Laguna sartu da]?
4. The friend, has he/she arrived [Laguna, sartu da]?
5. The friend has arrived [Laguna etorri da]?
6. Has he/she bought the bread [Erosi dau ogia]?

We examine data from three localities (see map of figure 1): Amezketa (which belongs to the Gipuzkoan area according to traditional dialect classification), Leitza and Sunbilla (these latter are included in the High Navarrese dialect area). The reasons for this choice of localities are connected with the aim of this paper, namely to study the differences in the intonation curve of y/n questions with and without a question cue. In all three selected localities, we found two ways of forming y/n questions: the use of the particle \(\textit{al}\) (figures 2 and 3), and a change in the intonation contour in the absence of a particle (figures 4 and 5).

For this research, we use the Autosegmental/metrical (AM) model of intonational analysis to examine the recorded data (Hualde 2003; Prieto 2003 and 2006; Toledo 2007). This model is one of the most widely used for analysing intonation in different languages. It has also been applied to the Basque language by Elordieta (2003, 2007a and 2007b, etc.), among others.

Leaving aside the presence or absence of interrogative particles, and taking into account previous literature on this subject, we consider that the most relevant feature of y/n interrogative sentences in Basque is the pitch of the last part of the sentence, and more precisely, the part which follows the last accented syllable. For this reason we focus on the final syllables of the sentences to see whether the intonation curve is upward or downward.

\(^9\) http://www.fon.hum.uva.nl/praat/

\(^{10}\) Lyko, K. (2008): "Segment_data.praat" script. GNU General Public License.
With regard to the type of boundary tones, and following previous literature about Basque intonation (Elordieta 2000, 2003; Gaminde 2003, 2007, 2010), we consider four contours: H%, L%, HL% and LH%. These are the main border contours we have found in the Basque language in these kinds of sentence.

Figure 1. Localities of the EDAK corpus and localities chosen for this study\(^\text{11}\).

4. PHONOLOGICAL ANALYSIS OF THE DATA

The phonological models found in the data from Amezketa, Leitza and Sunbilla are the following.

\(^{11}\) Map made with Google Maps.
4.1. Y/n questions marked with the particle al

We observe two main types of intonation contour: curves with a final H% boundary tone, as in figure 2, and curves with a final L%, as in figure 3. All of the data gathered from these three localities follows one of these two patterns.

Figure 2. Has he/she come? (Leitza).

Figure 3. Has he/she bought it? (Amezketa).
4.2. Y/n questions without the particle *al*

Looking at the pitch on the final syllables, we find three patterns: H% (figure 4), HL% (figure 5) and LH% (figure 6).

![Figure 4. Has he/she bought it? (Amezketa).](image)

![Figure 5. Has he/she come? (Leitza).](image)
As we can see in figures 2 and 4, final rises (H%) are found with and without the particle *al*. The other three intonation contours illustrated in the figures are each found in only one of the two types of interrogative sentence: L% is used exclusively in questions with *al*, and LH% and HL% are only used in questions without an interrogative particle.

At this point, a question which naturally arises is whether the use of the particle *al* is obligatory or optional in specific contexts for any given speaker, or for speakers from the same town and belonging to the same generation. Amongst speakers from Amezketa who belong to the older generation, we have found that its use is optional. Interestingly though, the presence or absence of the interrogative particle *al* conditions a difference in the final tonal contour. As shown in figures 4 and 5, we have a final L% boundary tone when the interrogative particle is used, but an H% boundary when this particle is absent. In Leitza, in turn, the same sentence recorded with *al* (older informant) and without it (younger informant), shows different pitch contours (see figures 2 and 5).

### 5. SOCIO-PROSODIC VARIATION

In this section we compare the data from the older and younger generations in order to look for any cues which would indicate the existence of age-related
sociolinguistic variation. To this purpose, we examine two features: pitch contour and the use of the particle *al* (see table 1).

<table>
<thead>
<tr>
<th>LOCALITY</th>
<th>GENERATION</th>
<th>With al (H%)</th>
<th>Without al (H%)</th>
<th>al (L%)</th>
<th>No al (H%)</th>
<th>No al (HL%)</th>
<th>No al (LH%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amezketa</td>
<td>Older</td>
<td>61%</td>
<td>39%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Younger</td>
<td>89%</td>
<td>11%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Leitza</td>
<td>Older</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Younger</td>
<td>39%</td>
<td>61%</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Sunbilla</td>
<td>Older</td>
<td>67%</td>
<td>33%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>Younger</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. Percentages of the use of the different y/n question types.

5.1. Amezketa

61% of the questions produced by the older generation are marked with the particle *al*. Older informant produced a sentence with the *al* mark (figure 3) and without it (figure 4). By contrast, the younger generation *al*-marked 89% of their questions.

If we look at the differences in pitch movement of the final syllables, we see that both older and younger people use low intonation contours in all questions marked with *al*. But in unmarked questions, older informant employs a high contour while younger speaker uses a high-low pattern.

5.2. Leitza

According to our data, speaker from the older generation consistently uses the particle *al* in all cases, whereas younger speaker produces y/n questions both with and without *al* (figures 7 and 8).
Looking at table 1, we see that the member of the older generation produces an H% pattern for all marked questions, whereas younger people produce all of them with a low tone. In unmarked questions, all of which are produced by younger speaker, we find an HL% pattern in all tokens.
5.3. Sunbilla

67% of the y/n questions produced by the older generation are al-marked and 33% are unmarked. Younger speaker never uses al, that is, they use only one way to form y/n questions, while older people make use of the two morphological possibilities.

In terms of intonation, older informant uses two patterns: HL% (67%) and LH (33%), while younger people only use the HL% pattern.

6. GEO-PROSODIC VARIATION

6.1. Geo-prosodic variation in the older generation

Here, we focus on three aspects: the presence or absence of the particle al, the intonation pattern of al-marked questions, and the intonation pattern of unmarked questions.

As far as the presence or absence of the particle al is concerned, the utterances of older speaker from Amezketa and Sunbilla present more similarities than those of the informants from these two localities and Leitza.

Older speaker from Leitza uses the particle al in all instances; however, in Amezketa and Sunbilla, marked questions are used in respectively 71% and 65% of the utterances of that age group. This is somewhat surprising because Leitza is geographically located between Amezketa and Sunbilla, as can be seen from the map. Moreover, according to traditional Basque dialectology, Amezketa is located within the Gizpuzkoan dialect area, whereas Sunbilla and Leitza both belong to the High Navarrese zone.

With regard to the intonation patterns of al-marked questions, the data from Amezketa and Sunbilla presents the same intonation pattern (older speakers only use the L% pattern), but the data from Leitza shows a different contour (an H% pattern in all cases).

Figures 9 and 10 show the typical final L% (Sunbilla) and H% (Leitza) intonation patterns respectively.
As for unmarked questions, in Amezketa we find only the H% pattern (figure 11), while in Sunbilla there are two occurring patterns: HL% in 65% of the cases (figure 12) and LH% in the remaining 35% (figure 13). As mentioned above, older informant from Leitza use al-marked questions.
Figure 11. Has he/she bought it? (Amezketa).

Figure 12. Has he/she bought it? (Sunbilla).
6.2. Geo-prosodic variation in the younger generation

Younger speakers use both marked and unmarked questions in Amezketa and Leitza but only unmarked questions in Sunbilla. So there is a change occurring in the use of the particle *al* in y/n questions by older and younger people in Sunbilla where the older generation produces both types of questions, but younger people never use the particle *al*. Indeed, while in Amezketa *al*-marked questions are widely used by younger people (89%), in Sunbilla they never occur.

All *al*-marked questions follow a low intonation contour in Amezketa and Leitza, and all unmarked questions follow a high-low pattern in all three localities.

7. CONCLUSIONS

In this paper we have described the intonation contours used in y/n questions in three towns of the Basque Country, in an area where the interrogative particle *al* is known to be used. Taking into account the final pitch contour and the use of *al*, our data shows that there are five intonation patterns: two patterns for questions marked with *al* (H% and L%), and three patterns in morphologically unmarked
questions (H%, HL% and LH%). H% contours can thus be used in interrogative sentences both with and without *al*.

The distribution of the two possible patterns for *al*-marked questions is as follows: a) The H% contour is only found in older informant from Leitza (100%); b) The L% pattern is found in both generations from Amezketa (100%) and from Sunbilla (100%), and in younger people from Leitza (100%).

For unmarked questions, the distribution of the three possible patterns is the following: a) The H% pattern is only found in members of the older generation from Amezketa (100%); b) The HL% pattern is found in the younger generation from Amezketa (100%), Leitza (100%) and Sunbilla (100%), and in older people from Sunbilla (67%); c) The LH% pattern is only found in the utterances of older people from Sunbilla (33%).

Regarding socio-prosodic variation, the main intergenerational difference is located in Sunbilla, and to a lesser extent, in Leitza.

As far as geo-prosodic variation is concerned, we find that in the case of the older generation, Leitza differs from Amezketa and Sunbilla. In Leitza, all y/n questions produced by the older speaker were marked with the particle *al*, whereas in Amezketa and Sunbilla *al* was used in 71% and in 65% of the cases respectively. The intonation contours of *al*-marked y/n questions uttered by older speaker from Leitza (H%), and from Amezketa and Sunbilla (L%) are different.

Nevertheless, when we turn to the younger generation, the Amezketa and Leitza varieties are more alike in that younger speakers from these two localities employ *al* optionally. However, this particle never appears in the data collected from younger speaker from Sunbilla. Finally, and with regard to the intonation contours, all *al*-marked questions produced by younger speakers from Amezketa and Leitza have an L% contour (as mentioned above, younger people from Sunbilla do not use the particle *al*). In unmarked questions, the HL% contour is used by younger speakers from all three localities.

**ACKNOWLEDGEMENTS:** We are grateful to J. I. Hualde, G. Elordieta and P. Prieto for their contributions to this study. This project has been possible with the economical support from the Ministerio de Ciencia e Innovacion of the Spanish Government (HUM2007-65094/FILO) for the period 2007-2010), and UPV/EHU (GIU07/05).
8. BIBLIOGRAPHIC REFERENCES


EFE, ISSN 1575-5533, XX, 2011, pp. 11-31
Corpus based prosodic variation in Basque: Y/N questions...


