International Linguistics Research; Vol. 3, No. 4; 2020
ISSN 2576-2974 E-ISSN 2576-2982
https://doi.org/10.30560/ilr.v3n4p57

The Sociolinguistic Sampling. Does it Need to be Redefined?

Francisco García Marcos[1]

[1] Facultad de Humanidades, University of Almería, Spain

Correspondence: Francisco García Marcos, Dpto. Filología. Ctra. Sacramento, s/n. 04120-Almería, Spain. E-mail: fmarcos@ual.es

Received: October 12, 2020; Accepted: November 4, 2020; Published: November 22, 2020

Abstract

The present article analyses a classic in the methodology on the analysis of the social variation of languages: the application of the ratio of 0'0025 % to obtain a representative sample of the population of a speaking community. This ratio, established empirically by Labov in 1966 for New York City, nevertheless presents important limitations when moving to communities with smaller populations. Replicating the empirical experimentation in four Spanish populations of different demographic size, it is shown that the empirically representative samples correspond to the confidence intervals already provided by the general statistics. Likewise, it is shown that these were the parameters between which 0,0025 % in the city of New York was developed. Consequently, the problem was not in the formulation of the ratio by Labov (1966), but in the subsequent indiscriminate application that has been made of it.

Keywords: sociolinguistics, variation, change, methodology, statistical sampling

1. Introduction

The present research critically reviews the sampling criteria employed by variationist sociolinguistics, practically during the last fifty years. These criteria, without a doubt, have sustained a great deal of empirical work. Multiple speech communities around the world have been studied, in principle, with a solid empirical foundation. In comparison with the synchronous linguistics immediately preceding it, the variationist sociolinguistics came much closer to the demographic profiles of the areas it analysed.

For this purpose, I use a common and universal procedure, applied in a uniform way in any kind of community. This way of proceeding implied an evident contradiction. In general statistics, representativeness is directly linked to the size of the reference sample. This means that, by definition, communities with different demographic sizes must follow different representativeness ratios. Variationist sociolinguistics, therefore, introduced a counterexample, at the very least, to general statistics. In view of this, two options should be considered:

A) The sociolinguistic reality is exceptional and, therefore, requires statistical procedures different from the usual ones.

B) Variant sociolinguistics has excessively extended a uniform criterion that, in the end, was not adapted to the reality of the speech communities it studied.

This research tries to solve this dilemma, replicating the methodology used by Labov in New York, to evaluate the results obtained with general statistical procedures.

2. The Ratio of Representativeness in Variationist Sociolinguistics. Relevant Scholarship

In 1966 Labov established one of the great landmarks of 20th century linguistics. In the immediate field, it was logically for sociolinguistics, and especially for the analysis of linguistic variation. His study of the social stratification of New York English has been the benchmark on which such research has been based in virtually all parts of the world. But it was also for linguistics as a whole, insofar as it meant a radical -and transcendental- change both in the theoretical and methodological order. In theory, it introduced a model that gave priority to the interrelationship between language and society as the epistemological core of linguistics. This implied, among other things, overcoming the classificatory vision of structuralism and the transcendentalism of generativism, almost newly introduced at that time. From the beginning, moreover, Labov's option proved to be a valid solution to the great issues that had preoccupied linguistics since time immemorial. Socio-linguistic variationism, the model directly inspired by Labov, immediately tackled the diversification of languages, change, their historical evolution or the skills needed to use human language. Later, this version of sociolinguistics became increasingly active in

applied linguistics, starting with the teaching of foreign languages (Taronne, 1979, 1982, 1988; Ellis, 1982, 1984, 1985) and ending with translation (Caprara, Ortega Arjonilla & Villena, 2016).

Much of the immediate success of variationist sociolinguistics was due to the fact that methodologically it introduced a substantial transformation in empirical research. Dialectology, its immediate reference within linguistics, had operated with reduced sets of speakers that it considered idiosyncratic. He applied mainly geographical criteria, to which he sometimes added other observations, never systematic and stable. In opposition to this, sociolinguists operated on the basis of representative samples of social reality, filtered through three axes of conditioning on linguistic behaviour (linguistic, social and stylistic factors). This allowed them to access a deeper and more complex dimension of linguistic reality, about which a much greater and more precise knowledge could be obtained.

From the beginning, the variationist sociolinguistics made a special effort to refine a very precise and strict methodology that would allow it to fulfil this great objective. That is why it established equivalence sets to analyse its variables, discriminated the levels of influence of social factors and operated with representative samples of the communities it studied.

On this last aspect it was decisive to specify the minimum number of speakers from which a sample was considered reliable. This was another of Labov's great contributions, establishing a ratio, 0,0025 % to the reference population, which fully guaranteed representativeness. López Morales (1994) emphasized the extraordinary empirical solvency of Labov's proposal, based on the examination of the data (Labov 1966), which were stable up to that point, from which they began to deviate and, therefore, to be unreliable.

The ratio of 0,0025% became a true methodological postulate, completely undisputed and consistently applied in all research that has followed the Labovian perspective. It showed its solvency by obtaining sociolinguistic samples of enormous relevance, thus making an extraordinary contribution to the development of a much more rigorous empirical linguistics. Undoubtedly, it has been a more than decisive reference, thanks to which substantial progress has been made in linguistic research, especially at the empirical level.

This does not mean that you have used the most appropriate criteria to collect your demographic samples. In theory, 0,0025% guaranteed efficient sampling, regardless of the demographic volume to which it was applied. But, in practice, this produced very asymmetrical results. Despite this, the sociolinguistic literature has not discussed the universal validity of the ratio established by Labov.

3. Methodology

To replicate the methodology used in Labov (1966), four Spanish-speaking communities with very different demographic dimensions have been selected: Pedro Martínez (Granada, 1134 inhabitants), Motril (Granada, 58020), Almería (189680) and Barcelona (5.5 M), according to the last statistical census in force in Spain (2018). In each of them, a social factor has been selected through which to measure the stratification of the linguistic variables studied. The aim is to check from what level of representativeness the results deviate. Labov (1966) showed in New York that these did not vary until reaching 0,0025%. Before (with 0,5%, 1%, etc.) the sociolinguistic distribution was stable and offered the same results. Therefore, here we intend to check whether the same empirical methodology reaches the same conclusions.

Specifically, the linguistic variables and social factors listed in Table 1 have been examined.

Table 1. Sampled populations

Town	Population	Variable	Factors
Pedro Martínez (Granada)	1134	/[t͡ʃ/	Redes sociales
Motril (Granada)	58020	/s/- /Ɵ/	Geography
Almería	189680	/[t͡ʃ/ /tr/	Sex
Barcelona	5,5M	Atitudes-1 Atitudes-2	General

4. Results

4.1 Some Previous Problems in the Application of the 0,0025 % Ratio

Before starting to contrast empirical data, already from the theoretical point of view very different results are obtained.

Table 2 summarizes the results of a possible application of the 0,0025% ratio among a very diversified set of populations. As it can be easily observed, some hypothetical samples would be impossible (0,01 persons in Latour de Carol), others would be irrelevant (2.14 in Flensburg) or not acceptable (14 in Málaga) and, finally, one could work with reasonable samples above one million inhabitants (Paris, Moscow…); that is, in contexts similar to New York.

Table 2. Application of the 0,0025% ratio to different speech communities

Type	Place	Population	0'0025%	Census
A-Less than 5.000	Latour de Carol (France)	421	0'01	2015
	Diezma (Spain)	775	0'01	2018
	Marinaleda (Spain)	2.626	0'06	2018
B- 5.000-10.000	Cheste (Spain)	8.319	0'21	2018
	Sentmenat (Spain)	9 078	0'22	2019
	Pinos Puente (Spain)	10.144	0'25	2018
C- 10000-100.000	Eckenförde (Germany)	22.798	0'56	2008
	Flensburg (Germany)	85.942	2'14	2015
	Roquetas (Spain)	94.925	2,37	2018
D- 100.000-500.00	Terrassa (Spain)	218.535	5'46	2018
	Kiel (Germany)	246.306	6'15	2016
	Konstanz (Germany)	285.325	7'13	2019
E- 500.000-1 M	Florencia (Italy)	358.079	8,95	2011
	Toulouse (France)	458 298	11'45	2013
	Málaga (Spain)	569.005	14'22	2016
F- 1 M-5 M	Hamburg (Germany)	1.841.179	46'02	2019
	Paris (France)	2.187.526	54'68	2017
	Roma (Italy)	2.617.175	65'42	2011
G- More than 5 M	London (United kingdom)	8.982.000	224,55	2019
	Sao Paulo (Brasil)	12.180.000	304'50	2018
	Moscow (Russia)	12.500.123	312'50	2018

As noted, table 2 shows the serious difficulties involved in the indiscriminate application of the 0,0025% ratio. In effect, up to group F (more than 1M inhabitants) the samples that would be obtained would not be qualitatively acceptable. Of course, in cities with less than 10,000 inhabitants it would be impossible to have even one speaker. In practice, these limitations have been resolved by managing much higher percentages of representativeness (García Marcos, 1989, 2020) for obvious reasons. Thus, at least in a specific type of speaker communities, this proportion has not been maintained in studies of sociolinguistic variation.

In the specific case analysed here, three of the four speaking communities examined would have serious difficulties if the 0,0025% ratio were applied. Except for Barcelona, all the others would not be able to establish minimum samples with which to work. Precisely, the procedure to be used tries to establish empirically what the relevant ratio is in each case.

4.2 The Case of / [t͡ʃ]/ in Pedro Martínez (Granada)

First, we examine a small rural community, Pedro Martinez, located in the province of Granada. Here, the phoneme is analysed the phoneme /[t͡ʃ]/ ("CH", in orthographic notation). This phoneme has a very low frequency of appearance in Spanish. It is the penultimate in all counts from Zipf and Rogers (1939) to Perez (2003), with an average appearance of no more than 0,34 %. Despite this, it has shown significant variability in the Spanish language, both in Europe and in America.

In Andalusia, the socio-cultural environment, which includes Pedro Martínez and the province of Granada, has also behaved in this way. *The Linguistic and Ethnographic Atlas of Andalusia* (*ALEA* from 1959), a reference work on the Spanish dialectology of the time, documented a considerable diversification in its domain. The appearance of the fricative allophones [ʃ], present in practically all of Andalusia, although in a somewhat irregular form, is noteworthy. The province of Granada was no exception. Fricative achievements were documented in the Costa de Granada region, in the capital and in its metropolitan area. In the interior of the province, the *ALEA* detected another variant of /[t͡ʃ]/, a cacuminal realization that the surveyor of the atlas (G. Salvador) noted in Iznalloz, the main town of the Montes Orientales region, to which the town of Pedro Martínez belongs.

This description of the dialectal situation in the area has remained valid until practically the present day. However, a mere impressionist approach to the linguistic reality of the area creates certain doubts, since other variants could be found that were not detected by the dialectologists. The suspicion was confirmed in García Marcos (2019a), when the spectrographic results attested the presence of an unexpected variant, a clear [tj], closer to the adherent realizations of this phoneme that have appeared in the Canary Islands.

The distribution in apparent time of the variants of /[t͡ʃ/ in Pedro Martínez indicated that at that time it was in a situation of notorious regression, although it must have had an expansion, if not complete, at least considerable through the whole sociolinguistic spectrum of the locality. At least this seemed to be deduced when examining the generational spectrum, which included the non-normative variants in older speakers. Thus, in 2019 the sociolinguistic distribution of [tj] in Pedro Martínez was mainly conditioned by three social factors:

1. Age, which described the process of generational regression just discussed.
2. Academic background, which distributed the adherent variant following a descending curve from speakers with a lower level of education (higher rates of appearance of the variant) to university speakers (testimonial appearance of [tj]).
3. The social network among which the speakers were located. The endogamic type (the dense ones in the sociolinguistic bibliography) favoured the appearance of [tj]. On the contrary, the exogamous type (the diffuse ones) provoked a remarkable increase of the [t͡ʃ] normative.

Thus, for the study that is being considered here, a considerably precise characterization of the sociolinguistic dynamics between which the alternation of [tj] and [t͡ʃ] in Pedro Martínez's Spanish was developing was available. Therefore, the most reasonable thing to do was to maintain the same set of equivalence that had been used in García Marcos (2019a), in which basically the two variants that had been documented would be considered.

Table 3. Set of variants of /t͡ʃ/ in Pedro Martínez (Granada)

Variant	Descripción
CH1	Affricate [[t͡ʃ]
CH2	Adhesive [tj]

Also, to adjust the precision of the experiment, one of the three factors mentioned above has been selected. Specifically, the social network has been included, as it contained the most internal variability. Age and school training were very much in the foreground in its internal stratification, so it seemed preferable to opt for the third possibility in terms of social determinants. Thus, the distribution of these two variants was examined, both in the general distribution of their results and through their sociolinguistic paths among speakers with dense or diffuse social networks.

Following these criteria and conditions, the tabulation of the corpus of García Marcos (2019a) has offered the results shown in table 4.

Table 4. Distribution of CH in Pedro Martínez (Granada)

			General		Endogamic		Exogamous	
N	n	**Ratio**	**CH1**	**CH2**	**CH1**	**CH2**	**CH1**	**CH2**
1134	56	R-3,3	80,3	19,7	60,96	39	96,8	3,1
	45	R-3,2	78,9	21	61,32	38,6	97,1	2,8
	40	R-3,5	80,1	19,8	60,59	39,4	96,9	3
	36	R-3,1	79,7	20,29	61,12	38,87	97,28	2,71
	30	R-2,6	80,3	19,6	61,2	38,7	98,1	1,8
	25	R-2,2	80,1	19,8	60,9	39	96,8	3,1
	20	R-1,7	80,2	19,7	61,3	38,6	96,7	3,2
	16	**R-1,41**	**79,6**	**20,3**	**62,4**	**37,5**	**95,8**	**4,1**
	15	R-1,32	71,3	28,6	48,24	51,75	86,7	13,2
	14	R-1,23	65,2	34,7	51,13	48,86	82,4	17,5
	13	R-1,14	63,9	36	49,7	50,2	79,4	20,5
	12	R-1,06	61,2	38,7	48,6	51,3	73,4	26,5
	11	R-0,97	60,8	30,1	47,2	52,7	71,3	28,6
	10	R-0,88	59,8	40,1	46,6	53,3	69,7	30,2

N = Statistical reference universe, **n** = Number of speakers sampled; **CH1** =[t͡ʃ]; **CH2** =[tj]

As can be easily seen, the point from which the data deviation starts is not at 0,0025 %, but at 1.41 %, which would correspond to a minimum sample of 16 speakers. Below that figure the data are completely distorted. Moreover, the sample behaves homogeneously in all its components. The figures have remained constant up to 1.41% in terms of the general distribution of variants in the speaking community, but also when observing the intervention of its social factors. In fact, the distance between inbred and exogamous speakers has been stable just up to that point, so it seems to be a common feature of the whole sample, regardless of the magnitude being measured.

4.3 [s-Θ] in Motril (Granada)

For the second point of analysis, the Granada town of Motril, one of the most constant processes of variation and change in the history of Spanish has been examined, with a case study that is especially unique in Andalusia. The transition from medieval Spanish to the Golden Age meant, among other things, a complete readjustment of the sibilant order. As a result, three variants concurred to occupy the same position in the new consonantal system: the distinction of [s-Θ] and two neutralisations, one seseant [s] and another lisp [Θ]. The normative Spanish opted for the first option, although the seseant was imposed in America and in part of the peninsular South and the Canary Islands. To these three solutions was added another variant, [h], which was much more sporadic and dialectally localised, although it was present in some parts of Andalusia (Alonso, 1951; Lapesa, 1962; Llorente, 1962).

The Costa Granadina, the region of which Motril is part, documents the four variants (García Marcos, 1989, 2020) that are included here.

Table 5. Variants of [s-Θ] in Motril (Granada)

Variant	Description
S/Z	Distinction [s-Θ]
S	Neutralization [s]
C	Neutralization [Θ]
H	Neutralization [H]

Sociolinguistic research (García Marcos, 1989, 2020) has shown that Motril, like the rest of the Costa Granadina, is no longer the zone of absolute predominance of [Θ], as it had been registered by the *ALEA* (1959). On the contrary, the four previous variants coexist, distributed according to the intervention of very active social factors, such as age, occupation, cultural level and, above all, the type of habitat. Urban environments favour the appearance of [s-Θ] and [s]. Rural and marine environments, on the other hand, increase neutralisation [Θ] and, to a lesser extent, [H].

The city of Motril brings together both scenarios. The nucleus of the population is located in the interior. It has an important seafaring neighbourhood two kilometres away, with very idiosyncratic sociological characteristics. In this way, it has been possible to incorporate the urban/seafood social factor into the analysis, in order to obtain the following table of results.

Table 6. Distribution of /s-Θ/ in Motril (Granada). General data

			General			
N	**n**	**R**	**S/Z**	**S**	**C**	**H**
58020	580	R-1	38,29	4,32	54,09	3,3
	240	R-0,5	37,34	5,03	54,71	2,9
	180	R-0,3	38,57	4,13	54,29	3
	120	R-0,25	38,32	4,47	54,12	3,1
	100	R-0,17	37,98	4,32	54,57	3,1
	50	**R-0,086**	**38,63**	**4,39**	**53,91**	**3,1**
	49	R-0,084	40,87	3,59	52,71	2,8
	48	R-0,082	45,22	5,01	48,12	0,7
	47	R-0,081	40,13	1,92	50,26	7,6
	46	R-0,079	41,09	1,84	47,53	9,5
	45	R-0,077	43,11	3,08	51,15	2,4

Table 7. Distribution of /s-Θ/ in Motril (Granada). Urbane/sailor

			Urbane				Sailor			
N	**n**	**R**	**S/Z**	**S**	**C**	**H**	**S/Z**	**S**	**C**	**H**
58020	580	R-1	38,29	4,32	54,09	3,3	38,29	4,32	54,09	3,3
	240	R-0,5	37,34	5,03	54,71	2,92	37,34	5,03	54,71	2,92
	180	R-0,3	38,57	4,13	54,29	3,01	38,57	4,13	54,29	3,01
	120	R-0,25	38,32	4,47	54,12	3,09	38,32	4,47	54,12	3,09
	100	R-0,17	37,98	4,32	54,57	3,13	37,98	4,32	54,57	3,13
	50	**R-0,086**	**38,63**	**4,39**	**53,91**	**3,05**	**38,63**	**4,39**	**53,91**	**3,05**
	49	R-0,084	40,87	3,59	52,71	2,82	40,87	3,59	52,71	2,82
	48	R-0,082	45,22	5,01	48,12	0,74	45,22	5,01	48,12	0,74
	47	R-0,081	40,13	1,92	50,26	7,64	40,13	1,92	50,26	7,64
	46	R-0,079	41,09	1,84	47,53	9,52	41,09	1,84	47,53	9,52
	45	R-0,077	43,11	3,08	51,15	2,39	43,11	3,08	51,15	2,39

As in the previous case, there is again a clear break point, this time at 0,086 %. Similarly, it remains constant in all the dimensions of the analysis, both when approaching the variable as a whole, and in the social specifications considered in this study.

Therefore, what had started to be described down in Pedro Martinez is confirmed in Motril. To meet the requirements of sociolinguistic representativeness, 50 speakers would be needed, that is, a sample of 0,086% of its population, 34.4 times more than 0,0025 %.

4.4 /tr/ in the City of Almería

The third observation point was located in the city of Almería, where two linguistic variables have been analysed, /tr/ and, again, [t͡ʃ/.

The articulation of the first of these variables, the consonantal group /tr/, in the city of Almeria presented another variation not detected by dialectology. Although the *ALEA* paid attention to the city, its point Al-508, did not perceive any variation in that phonic sequence. However, the exploratory sociolinguistic research of García Marcos (1999) did find a tendency to affricate the group. In the Spanish of America, it had been a perfectly delimited phenomenon, especially in Chile. From Lenz's initial research to the present day there has been an evident continuity in these studies (Figueroa, 2008). In the Peninsula it was not a completely unknown phenomenon either. In fact, when examining the Chilean situation, A. Alonso (1925, 1933) insisted that there were multiple solutions for /tr/ within the Hispanic domain, also in the Peninsula. Only that its radius of peninsular extension was located in the north of Spain, more specifically, in the Basque Country, Navarra and La Rioja. The possibility that it existed beyond that environment had not been openly and systematically considered either. The southern solutions of Almeria, in principle, were not within the foreseeable.

As it happened in Pedro Martínez, the spectrographic analysis confirmed the articulation of a variant close to the Chilean affricate solutions (García Marcos, 2019a). Once again, the set of equivalence made up of two variants was maintained, as it had been used at the time.

Table 8. Variants of /tr/ en Almería

Variant	Description
TR1	[tr]
TR2	[t˔˩˧]

On this occasion the social factors that conditioned the variation were, firstly, sex, and more distantly, cultural level and age. In general terms, there was considerable variation and presence in practically all social groups. But, in any case, affrication predominates among women, especially in those with average or scarce cultural training. Furthermore, it is predominant among speakers over 50 years old. So, finally, here the sex factor was incorporated into the analysis, the social conditioning with the greatest stability of all, to complete the general information on the variation of /tr/.

Table 9. Distribution of TR in Almería

			General		Men		Women	
N	n	**RATIO**	**TR1**	**TR2**	**TR1**	**TR2**	**TR1**	**TR2**
189680	121	0,063	97,54	2,45	99,21	0,78	94,81	5,18
	110	0,057	97,51	2,48	99,37	0,62	94,77	5,22
	100	0,052	97,48	2,51	99,26	0,73	94,65	5,34
	90	0,047	97,63	2,36	98,7	1,2	95,32	5,67
	85	0,044	97,78	2,21	99,01	0,9	94,87	5,12
	84	0,044	97,77	2,22	99,14	0,85	94,67	5,32
	83	**0,043**	**97,53**	**2,46**	**99,01**	**0,98**	**94,95**	**5,04**
	82	0,043	94,47	5,52	95,5	4,4	92,86	7,13
	81	0,042	93,37	6,62	94,7	5,2	90,91	9,08
	80	0,042	92,25	7,74	92,2	7,7	86,04	13,35
	70	0,03	91,07	8,92	89,3	10,6	85,13	14,86
	60	0,03	90,14	9,85	88,7	11,3	85,12	14,87
	30	0,01	90,78	9,21	88,1	11,8	83,24	16,75
	20	0,001	89,71	10,28	87,91	12,08	82,42	17,57
	15	0,0007	84,83	15,16	87,78	12,21	81,64	18,35

As in the two previous cases, a distortion of the results from a sample point other than 0,0025 % is empirically corroborated. On this occasion, it is situated at 83 speakers, which is equivalent to a ratio of 0,043 % of the 189680 inhabitants of the city. This figure also remains unchanged across the whole sociolinguistic spectrum.

4.5 /t͡ʃ/ in Almeria

As indicated, a second variable was included for Almeria. Garcia Marcos' exploratory research (1999) uncovered another variable that had not been found in the dialectology literature on Almeria either. The enormous variation of the phoneme / [t͡ʃ/ that had been recorded by ALEA (1959) in Andalusia, which has been referred to when analysing Motril (Granada), in Almeria also located a fricative variant [/ʃ/], on this occasion. In García Marcos (1999) its presence was confirmed in the sociolinguistic spectrum of the city, fundamentally conditioned by the factors of sex and age. Fricative variants were predominantly found among male speakers. At a smaller distance, it was a phenomenon with an increased presence among older speakers, although it was also located among the rest of the age groups. Therefore, the main social conditioning came from the sex factor, which also allowed maintaining the same contrasting criterion in the previous variable.

Table 10. Distribution of /[**t͡ʃ**/ en Almería

Variant	Description
CH1	affricate [t͡ʃ]
CH2	Fricative [ʃ]

In this way, a second approach to the same sociolinguistic spectrum, Almería, was produced, although controlling a new linguistic variable. The results of this new exploration are the following:

Table 11. Distribution of [t͡ʃ] in Almeria

			General		Man		Women	
N	n	**RATIO**	**CH1**	**CH2**	**CH1**	**CH2**	**CH1**	**CH2**
189680	121	0,063	93,51	6,48	77,54	22,45	96,52	3,47
	110	0,057	93,78	6,21	77,96	22,03	96,47	3,52
	100	0,052	93,41	6,58	77,32	22,67	96,78	3,21
	90	0,047	93,06	6,93	76,9	23,09	96,26	3,73
	85	0,044	93,77	6,22	77,34	22,65	96,63	3,36
	84	0,044	93,68	6,31	77,27	22,72	96,37	3,62
	83	**0,043**	**92,84**	**7,15**	**77,45**	**22,54**	**96,14**	**3,85**
	82	0,043	89,3	10,6	65,1	34,89	91,84	8,15

81	0,042	88,01	11,9	64,87	35,12	91,03	8,96
80	0,042	80,5	19,4	63,65	36,34	92,83	7,16
70	0,03	79,92	20,07	63,08	36,91	92,75	7,24
60	0,03	78,36	21,63	62,67	37,32	92,2	7,79
30	0,01	77,97	22,02	62,34	37,65	91,28	8,21
20	0,001	76,84	23,15	61,68	38,31	91,2	8,7
15	0,0007	75,78	24,21	61,16	38,83	90,8	9,19

As can be seen, the ratio that keeps the results stable is around 0,043 %, exactly the same as in the analysis of the previous Almeria variant. This complete coincidence confirms that it is a value that affects the whole of sociolinguistic variation within the speech community and, therefore, reinforces the hypothesis among which this work is being developed.

4.6 Sociolinguistic Attitudes in the City of Barcelona

The last point of analysis has been located in Barcelona, where two other variables will be examined, although they now refer to sociolinguistic attitudes. Catalonia in general, and Barcelona in particular, have been the object of permanent sociolinguistic attention. Since Badia's pioneering and almost foundational study (1969) of the language of the people of Barcelona, sociolinguistics has not neglected a community with an obvious attraction for analysing the relationships that link societies and languages. Barcelona is a macrocosm of 5.5 million inhabitants, 1.5 times smaller than New York, which has been the recipient of migratory contingents practically since the beginning of the 20th century. In Catalonia, moreover, there has been a secular linguistic contact with the coexistence of Spanish and Catalan. There have been different stages in relation to this linguistic contact, among which two recent ones stand out, which are very marked from the sociolinguistic point of view. The dictatorship of General Franco (1939-1975) meant a strong repression for Catalan, reduced to the most informal levels of the socio-functional spectrum. After the reinstatement of Spanish democracy, from the achievement of political autonomy in 1979 Catalonia began a progressive linguistic planning aimed at the normalisation of its vernacular.

This means that, at least in theory, the aim was to modify the diglossia inherited from the previous stage, in which Catalan had been reduced to the status of a B-language. Naturally, the direction and intensity of this transformation of the diglossia that was in force until 1979 has involved different and even opposing positions, so that a gradation of options has been established, sometimes at great social risk:

1. The complete suppression of the Spanish language from the functional repertoire of Catalonia.
2. The inversion of the diglossia, placing Catalan as an A-language and Spanish as a B-language.
3. The formal and socio-functional equalization of both languages, which would mean the establishment of the co-officiality they enjoy.
4. The maintenance of the previous diglossia.

This has meant an obvious linguistic debate, very present in social life, within which reality has not always been in step with political provisions. In García Marcos (2019b) it was found that the official regulations on compulsory signage in Catalan were not applied regularly, especially in the urban centres of significant populations in the capital's metropolitan area. This was a significant indication of the sociolinguistic conflict referred to above.

4.6.1 Throughout this process it has been essential to know the attitudes of citizens towards the two main languages of the long and dense multilingualism that is registered in Catalonia, and in Barcelona in particular. For the experiment to be developed here, the studies by Huguet (2007) and Gracia (2012) on this issue have been taken as a reference. After collecting materials in September 2019, we first examined attitudes towards the languages involved in this contact and the hypothetical predominance of one of them. Empirical research showed that positions in real society are less polarized than reflected in the political debate. Even so, the differences exist and are therefore relevant for a sociolinguistic analysis.

The first group took over the assessments of the Barcelona people towards the Catalan and Spanish. To do this, three possible attitudes were considered:

A. in favour of each of the languages (+)

B. refusals (-) or

C. neutral (=).

The latter possibility occurs when respondents have either explicitly stated that they are or have refused to make a statement in any direction. It should be made clear that these solutions have not been offered as mutually exclusive possibilities. Respondents could assign two attitudes: negative and positive or give up in both cases. Thus, after tabulating the data, the following table of results was obtained.

Table 12. Attitudes towards Spanish and Catalan in Barcelona

		Spanish			Catalan		
	R	**(+)**	**(-)**	**(=)**	**(+)**	**(-)**	**(=)**
500	R-0,009	71,59	2,75	25,66	69,27	3,42	27,31
475	R-0,0086	72,17	3,47	24,36	69,31	3,49	27,2
450	R-0,0081	72,75	3,27	23,98	69,32	3,31	27,37
425	R-0,0077	72,43	3,6	23,97	69,25	3,27	27,48
400	R-0,0072	72,48	3,4	24,12	69,29	3,36	27,35
375	R-0,0068	72,17	3,7	24,13	69,31	3,17	27,52
350	R-0,0063	72,09	3,39	24,52	69,34	3,25	27,41
325	R-0,0059	72,25	3,16	24,59	69,28	3,42	27,3
300	R-0,0054	72,22	3,21	24,58	69,41	3,48	27,11
275	R-0,005	72,3	3,11	24,59	69,35	3,32	27,33
274	**R-0,00498**	**72,31**	**3,08**	**24,61**	**69,28**	**3,38**	**27,34**
273	R-0,00496	72,31	6,14	27,83	71,07	3,62	25,31
272	R-0,00494	62,38	7,19	30,43	73,83	3,96	22,21
271	R-0,00492	59,25	8,21	32,54	74,91	3,98	21,11
250	R-0,0045	55,23	8,32	36,45	76,34	4,08	19,58
200	R-0,0036	49,97	9,12	40,91	78,93	4,48	16,59
100	R-0,0018	42,42	9,21	48,36	81,37	5,53	13,09
55	R-0,001	35,39	10,26	54,34	86,42	6,14	7,44

This time, the results remain stable at 0,00498 %, which translates into 274 speakers. With only one speaker less, at 0,00496 %, the first significant deviations are already beginning to appear. It could be discussed whether they are sufficient to not accept the relevance of that sampling point. But, in any case, from that point on, the deviations are too ostensible.

Therefore, once again, a larger sample is required than that theoretically contemplated, although only 1.9 above. It is therefore confirmed that the ratio is more effective in large populations, demographically close to the empirical reference used by Labov, New York City.

4.6.2 The second group of attitudes focused on a much-debated question, practically from the beginning of the new language policy developed from 1979, about which language should be the predominant one in communication. An analogous criterion of attitude assignment was maintained, so that speakers could choose one, or several, options from among the four proposed:

A. Catalan (only)

B. Spanish (only)

C. Spanish and Catalan (jointly)

D. None of the two in particular.

The results of this second approach to the sociolinguistic attitudes of the Barcelona population also showed very heterogeneous profiles, without a clearly predominant trend, as shown in table 13.

Table 13. Attitudes towards the predominance of one language in Catalonia. Barcelona data

N	**n**	**Ratio**	**Catalan**	**Spanish**	**Both**	**Neutral**
5,5 M	500	R-0,009	8,72	58,42	23,42	9,44
	475	R-0,0086	8,73	58,31	23,39	9,57
	450	R-0,0081	8,68	58,41	23,4	9,51
	425	R-0,0077	8,71	58,43	23,41	9,45

400	R-0,0072	8,49	58,32	23,37	9,82
375	R-0,0068	8,52	58,37	23,4	9,71
350	R-0,0063	9,53	58,45	23,4	9,53
325	R-0,0059	8,73	58,42	23,38	9,47
300	R-0,0054	8,81	58,39	23,39	9,41
275	R-0,005	8,64	58,31	23,42	9,63
274	R-0,00498	**8,72**	**58,29**	**23,41**	**9,58**
273	R-0,00496	10,14	55,62	20,14	14,1
272	R-0,00494	12,36	53,29	17,32	17,03
271	R-0,00492	15,28	50,02	15,51	19,19
250	R-0,0045	17,31	47,26	11,25	24,82
200	R-0,0036	20,23	44,78	9,31	25,68
100	R-0,0018	23,49	42,93	7,63	25,94
55	R-0,001	27,03	40,26	6,13	26,58

As happened in Almeria, the behaviour of this second sample has been completely equivalent to that of the first incursion into the attitudes of the people of Barcelona. Once again, the relevant ratio is 0,00498, which confirms it as the appropriate parameter for the preparation of a sociolinguistic sample on the city.

5. Conclusions. Towards the Revision of the Reliability Ratio of Sociolinguistic Sampling

The results of the analyses carried out in this work warn of the inadequacy of using the ratio of 0,0025 %, as a methodological universal for the measurement of sociolinguistic variation. In all the cases examined, the deviation of the results has been above this parameter. It has done so, moreover, in a quite reliable way: it has worked both in action data and in attitude data, it has remained stable in more than one variable of the same community (Almería, Barcelona) and, finally, it has demonstrated its applicability in the analysis of a variable as a whole, but also in its social subspecifications. Therefore, it is obvious that 0,0025% introduced a very important methodological novelty. But, in the same way, it is also evident that it is not a criterion that should be maintained with a universal character.

Naturally, more axes of observation could have been included, to observe also linguistic conditioning and stylistic variation. However, they would not have modified the final results. The aim here was not to measure the correlation of sociolinguistic variation, but a preliminary question, such as the minimum number of speakers to guarantee the reliability of the sample.

At this point, it is necessary to ask whether sociolinguistics should use special statistical methods or, on the contrary, may resort to standardized methods. Obtaining representative samples has a long tradition in statistics. Moreover, sociolinguistics is easily integrated into the field of stratification-type probability statistics. As a general criterion, these samples are configured on the basis of four main factors: the size of the reference universe (the inhabitants of a community of speakers), the level of confidence, the admitted margin of error, and the internal diversity of the reference universe. This is usually done by means of a classical calculation, summarized in the following formula (Cochran, 1982: 104; Vivanco, 2005: 53).

$$\text{Statistical simple size} = \frac{\frac{z^2 \times p\,(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p\,(1-p)}{e^2 N}\right)}$$

where N = size of the population; e = margin of error (percentage expressed with decimals); p = probability and z = z-score. This last parameter is in charge of measuring the amount of standard deviations that are proportionally far from the mean.

In principle, the ratios obtained empirically in the assumptions handled here correspond to wide confidence intervals and small margins of error, and therefore perform positively within the general statistics. Table 13 provides relevant information in that direction.

Table 14. Comparison of empirical and theoretical ratios

	Empirical ratio		Theoretical sampling			
City	N	R	n	MS	NC	ME
Barcelona	5,5M	0,00498	274	273	90	5
Almería	189680	0,043	83	83	93	10
Motril	58020	0,08	50	50	84	10
Pedro Martínez	1134	1,32	15	15	75	15

N = statistical reference universe, n = sample, R= ratio, MS = Sample, ME= statistical error margin, NC = statistical confidence level.

It has its limitations, however, because there is still a tendency for the sample volume to increase as the demographics of the reference communities decline. Even the empirical ratios have shown the lowest levels of confidence in the smallest populations: 75% (Pedro Martínez) and 84% (Motril) compared to 93% (Almería) and 90% (Barcelona). Obviously, the margin of error follows a downward trend in the opposite direction: 15% in Pedro Martínez, 10% in Motril and Almería, 5% in Barcelona.

It is certainly a paradoxical situation. In fact, Labov's ratio (1996) follows standardized statistical criteria, although only for New York. That year, 1996, the official US census registered 8,399,000 inhabitants in the city. The 0,0025% of that population resulted in a sample composed of 210 speakers. This figure coincided with the results that would have been obtained had standardized criteria been applied, with a 95% confidence level and a 7-8% margin of error. Labov's empirical method - probably without being aware of it - had brought it to the same place as the general statistics.

The problem was that what was valid for New York could not be transferred as a criterion of universal representativeness. Moreover, by definition, statistical representativeness depends on the demographic dimensions of each community. Therefore, what was valid for New York (0,0025%) had to be invalid in different populations. Contrary to what has been common practice, the only certainty was that the 0,0025% could not be universally applicable.

In that sense, the sociolinguistic variationist has been making a systematic error in the preparation of its samples. It is true that, in spite of this, it has obtained better empirical results compared to previous synchronic linguistics. But, at the same time, it is evident that it cannot continue to perpetuate such a large methodological error. Especially when the solution is as simple as following the general statistical procedures for the preparation of representative samples, sufficiently contrasted and verified.

References

Alonso, A. (1951). Historia del seseo y ceceo españoles. *Thesaurus, VIII*(3), 111-198.

Alonso, A. (1925). El grupo tr en España y América. *Homenaje a Menéndez Pidal*. Madrid: Hernández y Galo, Vol. (II), 167-191.

Alonso, A. (1953). *Estudios lingüísticos. Temas hispanoamericanos.* Madrid: Gredos.

Alvar, Manuel. (Antonio Llorente y Gregorio Salvador, cols.) (1961). *Atlas Lingüístico y Etnográfico de Andalucía*. Madrid: CSIC.

Badía, A. M. (1969). *La llengua dels Barcelonins: resultats d'una enquesta sociològico-lingüística*. Barcelona: Edicions, 62.

Caprara, G., & Emilio, O. Arjonilla y Juan Andrés V. (2016). *Variación lingüística, traducción y cultura: de la conceptualización a la práctica profesional*. Frankfurt am Main: Lang. https://doi.org/10.3726/978-3-653-06875-7

Cochran, W. G. (1982). *Contributions to statistics*. New York: John Willey.

Dalarna: Univ. Dalarma.

Ellis, R. (198). *Discourse Processes in Classroom Second Language Development.* London: University of London.

Ellis, R. (1984). *Classroom second language development.* Oxford: Pergamon.

Ellis, R. (1985). *UnderstandingSecond Language Acquisition.* Oxford: O.U.P.

Figueroa, M. (2008). *Prestigio de las variantes de /tr/ en la comuna de Concepción. Estudio sociolingüístico.* Concepción: Universidad de Concepción (Chile).

García, M. F. (1987). El segmento fónico *VOCAL*+S en ocho poblaciones de la Costa Granadina. Aportación informática, estadística y sociolingüística al reexamen de la cuestión. *EPOS*, III: 155-180. https://doi.org/10.5944/epos.3.1987.9490

García, M. F. (1989). *Estratificación social del español de la Costa Granadina*. Almería: Depto. Lingüística General, Universidad de Granada.

García, M. F. (1999). *Patrones sociolingüísticos del español de Almería.* Granada: Mágina.

García, M. F. (2019a). Una lectura caotológica de la vida lingüística. A propósito de algunas soluciones andaluzas de "tr" y "ch". In M. Baran, ed. *El andaluz polifacético. Acercamientos desde la comunicación y la didáctica*. Gdansk: Wydawnciwto Uniwersytetu Gdańskiego, 53-75.

García, M. F. (2019b). El lenguaje de las rotulaciones de establecimientos comerciales en las ciudades contemporáneas. Los casos de Almería, Łódź y Tarrasa. *Signa: Revista de la Asociación Española de Semiótica, 28*, 699-732. https://doi.org/10.5944/signa.vol28.2019.25075

García, M. F. (2020). *Variación y cambio sociolingüísticos en tiempo real. El español de la Costa Granadina (1987-2017).* Jaén: Universidad de Jaén.

Gracia Sánchez, J. C. (2012). *Un estudio sociolingüístico sobre el catalán: Los efectos de la inmersión lingüística*.

Huguet, Á. (2007). Multilingüismo y actitudes lingüísticas. Un estudio en el contexto bilingüe de la Cataluña actual. In *Hizkunea*. Retrieved 04.05. 2020 from http://www. euskara. euskadi. net/r59738/es/contenidos/informacion/artikl6_l_linguistika_07_02/es_linguist/artikl6_l_linguistika 07_02.Html

Labov, W. (1966). *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Lapesa, R. (1962). Sobre el seseo y ceceo andaluces. Homenaje. In *Homenaje a A. Martine*. La Laguna: Universidad de La Laguna, Vol. III, 99-165.

Llorente, A. (1962). Fonética y fonología andaluzas. *Revista de Filología Española, XLV*(1), 227-240. https://doi.org/10.3989/rfe.1962.v45.i1/4.925

López, M. H. (1989). *Sociolingüística.* Madrid: Gredos.

López, M. H. (1990). *Métodos de investigación lingüística.* Salamanca: El Colegio de España.

Pérez, H. E. (2003). Frecuencia de fonemas. In *Revista Electrónica de la Red Temática en Tecnologías del Habla*, *1*. Retrieved 04. 05. 2020 from http://lorien.die.upm.es/~lapiz/e-rthabla/numeros/N1/N1_A4.pdf

Taronne, E. (1979). Interlanguage is a chamaleon. *Language Learning, 29*, 181-191. https://doi.org/10.1111/j.1467-1770.1979.tb01058.x

Taronne, E. (1982). Systematicity and attention in interlanguage. *Language Learning, 32*, 69-82. https://doi.org/10.1111/j.1467-1770.1982.tb00519.x

Taronne, E. (1988). *Variation in Interlanguage*. London: Arnold.

Vivanco, M. (2005). *Muestreo estadístico. Diseño y aplicaciones.* Santiago de Chile: Editorial Universitaria.

Zipf, G., & Francis, R. (1939). Phonemes and Variphones in four present day Romance Languages and Classical Latin from the viewpoint of dynamic Philology. In *Archives Néerlandaises de Phonétique Expérimentale, 15*, 111-147.