

The Sociophonetics of /r/ in Bozen: Modelling Linguistic and Social Variation

Lorenzo Spreafico (Corresponding author)

Alpine Laboratory of Phonetic Sciences, Free University of Bozen-Bolzano, Piazza
Universit  1, 39100, Bolzano, Italy

Tel: 39-047-101-2455 E-mail: lorenzo.spreafico@unibz.it

Alessandro Vietti

Alpine Laboratory of Phonetic Sciences, Free University of Bozen-Bolzano, Piazza
Universit  1, 39100, Bolzano, Italy

Tel: 39-047-012-452 E-mail: alessandro.vietti@unibz.it

Received: August 5, 2016 Accepted: August 13, 2016 Published: October 5, 2016

doi:10.5296/ijl.v8i5.9849 URL: <http://dx.doi.org/10.5296/ijl.v8i5.9849>

Abstract

How do speakers reconstruct the boundaries of an allophonic system? In our paper, we address this question and examine how speakers organize into consistent groups of allophones the array of /r/-variants that are used in South-Tyrol Italian (STI). In addition, we discuss that this process of grouping is based on two intertwining sources of variation: the linguistic source and the socio-indexical source. We argue that the indexical dimension is not disconnected from the linguistic one, but it contributes in an essential way to its structuring.

Our investigation is based on a sample of two thousand tokens of /r/. These occurrences are extracted from a corpus that includes the (semi)spontaneous productions of 14 Italian-German bilingual speakers. The analysis concerns the identification of possible relationships among the allophones with respect to (a) distributional, (b) stylistic and (c) biographical factors. Data are analyzed using a multivariate exploratory technique, namely the multiple correspondence analysis approach. The results clearly show how the aggregation of indexical and linguistic factors determines the emergence of two different allophonic subsystems, that is the Italian of Italian-dominant speakers (STI-i) and the Italian of German-dominant speakers (STI-d).

Keywords: Sociophonetics, Rhotics, Bilingualism, Multivariate analysis

1. Introduction

The ability to recognize and actively use the socio-contextual data present in one language is not part of a modular vision of grammatical competence, neither in a technical nor in a broader sense. Even in the theorization of the Labovian sociolinguistics back in the Seventies, the speaker's knowledge of the social and contextual variations was considered – and it still is if we consider Labov, 2006 and Labov et al., 2011 – almost a component that is not involved in the process of grammatical “production” within a given linguistic message. The reason is twofold: first, for the intrinsic nature of the carried meanings, which are not linguistic, but social and communicative; second, for the “co-grammatical” and “post-grammatical” role of the different phases of the production chain when producing a linguistic message. In this second statement, the socio-contextual variability can be inserted into the output of an optional rule as a sort of social “interface” to the grammatical competence. Basically, despite their probabilistic nature, variable rules work as a tool for formal representation in the social distribution of the variation in the same way that optional rules of the generative grammar stand as a way of representing the contextual variability which is to be found in the variation within the language.

In that respect, the cognitive statute of the sociolinguistic variation has not been extensively explored, whether the variation was to be understood as individual, or to be considered in its extension on the entire community of speakers.

The mechanisms left unexplored are therefore those which allow the speaker to integrate variation and grammatical competence, that is those that allow speakers to identify and produce different variants on the basis of social data. In other terms, the relation between contextual factors and grammatical choices must still be understood, as well as to what extent these factors might contribute to the structuring of linguistic categories.

According to Wolfram (2006:339) we still need to understand, through a proper texting of possibly psycholinguistic nature, what knowledge the speaker has in terms of variation and, in particular, his/her ability to:

Recognize the strength of those factors influencing the variation;

Recognize or at least take into account the probabilistic mechanisms producing competition among factors;

Keep track of the frequency values of the different variants.

The studies conducted in the field of sociophonetics (especially in the United Kingdom; see Foulkes, Docherty, 2006; Docherty, Foulkes, 2014) and, in general, the usage-based approaches to phonology (see Pierrehumbert, 2003; Bybee, 2001) have rehabilitated the role of the social variation, revitalizing some of the “classical” questions stemming from sociolinguistics and trying to approach them with new and partly different techniques and objectives. As regards sociophonetics, if methods and techniques are largely to be found in experimental phonetics, objectives take place specifically in a cognitive frame and aim to understand how speakers perceive, store and reproduce the linguistic variation, creating at the

same time links with contextual elements and using these links as tools for better understanding the meaning. The focus of researchers' attention is the ability of individuals to interact within a social environment by means of their knowledge of the language's social variation.

What are particularly significant and emblematic in this direction are the studies conducted by Foulkes et al., 2004 on the interaction between parents and children during L1 acquisition, (in particular in the *child directed speech*, Vihman, Keren-Portnoy 2013). From these studies, it emerges how the contextual and social variation acquisition does not follow the acquisition pattern of a fixed and "stable" grammar, but the contextualization process is integrated to the linguistic encoding and decoding one.¹

Although the perspective adopted for this study concurs with the line of thinking proposed by sociophonetic studies, it does not fully share its assumptions, (for a critical presentation of sociophonetics see Celata, Vietti, 2011; whereas for an introduction to exemplar theory see Johnson 2005 or Vietti, 2014).

In this study it is our intention to show how speakers can organize a set of linguistic variants by just integrating both linguistic (for instance the distribution of allophones in phonetic contexts) and social data. The analysis we propose can be perceived as a sort of simulation in order to demonstrate how, in order to reconstruct the "grammars" of the language variants, we should proceed only if we take into account both socio-contextual and linguistic data. To a certain extent, the idea at the basis of this study runs as follows: in order to reconstruct the phonology of the language variants present in our social environment we must refer to the socio-contextual data that come together with the linguistic input.

Hence, social factors allow us to correctly combine the linguistic elements in distinct varieties within which we can make phonological choices.

The case study we hereby present focuses on the different productions of /r/ in the Italian spoken in Bozen and the possible modes of aggregation of such allophones in subsets either with or without social data.

Bozen is the capital city of the province of South-Tyrol, in northern Italy. According to the last census, about 74% of the city's inhabitants spoke Italian, 25% Tyrolean (an Austro-Bavarian dialect of the German language) and 1% Ladin as their first language.

2. Method

The analysis we present in this paper is based on an Italian corpus sampled in Bozen, (for a description of the corpus, see Spreafico, Vietti, 2010; Vietti, 2012 and Vietti et al. 2015) and made of a collection of audio recordings and transcriptions of spoken production by 28 speakers, aged between 25 and 42.² The corpus has not been conceived with the aim to elicit

¹ In principle, positions and results are not new and recall prior records in ethnographic sociolinguistics as in Hymes, 1972 or in the studies on the sociolinguistic variation acquisition in children starting from Ervin – Tripp, 1971. What differ are not just the tools of analysis but the aim to empirically show how social data play a role in the cognitive representation of linguistic units.

² The sample was recorded with a Marantz PMD660 recorder and a Behringer B1 microphone. Sampling took place at 22.05 kHz, digitalization at 16 bit.

/r/ (rhotics) but it has been planned with the aim to achieve various research objectives in the field of spoken language.

The subjects' sample has been subdivided into three macro-groups based upon the language or languages learned during primary socialization in the family: a first group of Italian speakers (n^o7), a second group of German speakers (n^o11), and a third group of bilingual Italian-German speakers (n^o7). The investigation presented in this paper operates on a subsample of 14 subjects.

Table 1 lists the most relevant socio-demographic data of the selected subjects.

Table 1. Sample of examined subjects

	Name	Age	Gender	Primary Education ³	L1
1	VER	25	F	German	German dialect
2	HEL	38	F	German	German dialect
3	ULL	26	F	German	German dialect
4	VEW	31	F	German	German dialect
5	MAT	33	M	German	German dialect
6	JAS	27	F	German	German dialect
7	CHR	30	M	German	German dialect
8	AND	38	M	German	German dialect
9	CRI	37	F	Italian	Italian
10	CHI	33	F	German	Italian
11	BIB	27	M	Italian	Italian
12	STE	31	F	Italian	Italian
13	MAN	32	F	German	German dialect/Italian
14	ALE	24	M	German	German dialect/Italian

The corpus used for this analysis takes into account the productions of /r/ of each speaker in each of the three assigned experimental tasks: reading of single words in isolation, reading of sentences and reading of a simulated dialogue in the lab, through the map task technique. Materials and data acquisition methods were refashioned after those used in the spoken Italian corpus CLIPS (Cutugno, 2007; De Masi, 2007). Thus, the statistical analysis is based upon a data sample of 2340 /r/ occurrences.

As the data set of occurrences was being created, we first classified the different productions of /r/ by integrating the auditory perception together with the analysis of the spectrographic features, (for classification criteria see Spreafico, Vietti, 2010). Each occurrence was then codified upon the established values according to the linguistic and extra-linguistic variables as shown in Table 2.

³ It has to be noticed here that within the province of South Tyrol a segregated school system operates whereby the German and the Italian schools barely interact. Usually bilingual or monolingual Tyrolean-speaking children attend German schools, while Italian monolingual children attend Italian schools.

Table 2. Variables taken into account during the analysis together with their values and codes

Variable	Definition	Values
Variant	/r/ to be found in the corpus	Alveolar approximant (Appr_A) and uvular (Appr_U), uvular fricative (Fric_U) ⁴ , retroflex ⁵ (Retro), alveolar tap (Tap_A) and uvular tap (Tap_U), alveolar trill (Trill_A) and uvular (Trill_U), vocalization of /r/ (Voc), deletion ⁶
Phonetic context	Phonetic macro contexts in which /r/ may occur in the corpus	Word beginning (#RV), word ending (#VR), between consonant and vocal (CRV), between vocal and consonant (VRC), between vocals (VRV), geminated sound between vocals (VRRV)
Lexical accent	/r/ occurrence in a syllable with or without lexical accent	Stressed syllable (Stressed), Unstressed syllable (Unstressed)
Style	Type of task for data elicitation	Reading of single words (words), map task (Map Task)
Speaker	Initials of the 14 examined speakers from the corpus	
Family language	Language/s spoken during primary socialization	South-Tyrol dialect (dia), South-Tyrol dialect and Italian (dia+ita), Italian (ita)
Primary school	Primary school system	Italian (ita), German (de)
Secondary school	Secondary school system (first and second level)	Italian (Ita), German (De)

At the end of the encoding process, a matrix of data made of 2340 occurrences x 8 variables (18720 cells) will result. In order to determine what kind of relationships will emerge among the variables, that is which associations could emerge among the values of the different variables, a multivariate statistical analysis of the matrix was conducted. The choice has been addressed towards a technique specifically developed for the treatment of nominal variables, namely the multiple correspondence analysis (MCA)⁷. This technique was successfully tested in previous works, (see for instance Vietti, 2008; Spreafico, Vietti, 2010). The reasons for this choice are mainly motivated by the urge to observe and briefly represent the relationships among variables, which are many and characterized by several values. One of the advantages of this technique is that it allows a graphic portrayal of the results in a geometrical space: relationships among variables' values are depicted in terms of distance among points (objects) on a Cartesian plane, as in a sort of topographic matrix map.

⁴ Voiceless and voiced uvular fricatives are here gathered under one fricative "strategy" in order not to excessively fragment the analytical framework.

⁵ Under this category, which refers to variants of regional Italian (South Veneto, cf. Mioni 1990), both retroflexes [ɽ] and flaps were inserted (see Ladefoged & Maddieson, 1996: 230-232).

⁶ Sound deletion does not distinctively represent an allophone of /r/, nonetheless it was inserted among the values of possible variants with the purpose to describe contexts and factors affecting sound deletion (Khattab, 2002).

⁷ Useful references to study in depth the functioning of this technique can be found in Di Franco (2006) and Greenacre (2007).

In our study, we will try to understand what the linguistic and socio-linguistic factors are that determine the distribution of nine /r/ variants in a sample of speakers residing in Bozen. We are faced with a variable with several discrete values that needs to be correlated with more than one linguistic or social variable in order to identify possible associations among the different values.

3. Multivariate Exploratory Analysis

The matrix of data has been investigated following an analytical path aiming at replying to an ordered sequence of questions.⁸

The first question deals with the possibility, on one hand, to organize the set of allophones of /r/ present in the Italian spoken in Bozen within one grammar (or variety) or, on the other, to split the set in several partially different subsets/grammars. The second question stems from the results of the first one: if different varieties of language do exist, what are the contextual factors that determine them? The third question is: once the possible varieties of language are depicted, in what way(s) do sounds distribute themselves in an allophonic distribution? Under which phonological rules?

The analysis path presented in this work is of an exploratory nature, so the passage from one phase to another is not given by a comparison with a base of reference data, but simply by a comparison among different representations of the same data matrix. This has been done in order to offer a vision of the cases that is as close as possible to the sociolinguistic environment.

The starting point of this analysis is represented by the graphic result of the first multivariate exploratory analysis, a real map where the relationships among the variables of linguistic nature are investigated, in this case *Variant* (allophone of /r/), *Phonetic Context* (the specific context where the allophone appears), and *Accent* (the presence or absence of the lexical accent in the syllable where /r/ occurs). The main criterion for a clear understanding of the maps – of this in particular and of the ones to follow – is the analysis of the points' position and their reciprocal distances: the proximity among points is the index of association among the values' variables: the greater the proximity, the stronger the relationship between two values will be. The association is direct if the points are near and in the same quadrant; in other words when the frequency of a value increases, so will the other. On the contrary, the association will be opposite if the points are near on an axis but with an opposite sign on the other, that is when the frequency of a value increases, the other will decrease proportionally.

As the points-values are distributed on the Cartesian plane according to their reciprocal relationships, the meaning of the axes can be assigned only *a posteriori* through the observation of the whole configuration, (similarly to what happens in the techniques of factor analysis). Generally speaking, those points with the most extreme values on the different axes are those which primarily contribute to pinning down the “meaning” of a particular dimension; on the contrary, those more central and nearer to the origin are those «less marked» and less subject to the influence of the variables.

⁸ Here this technique was carried out with IBM SPSS Statistics v.20 software (*Dimension Reduction*>*Optimal Scaling*).

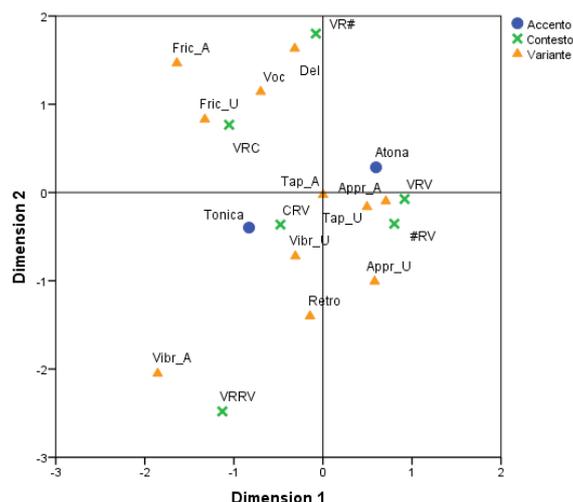


Figure 1. Map of the linguistic variables: lexical accent, phonetic context and allophones of /r/

In order to answer our first question, we examined the relationships among the only encoded “linguistic” variables. According to the given direction we can proceed to examine the map as shown in Figure 1. The adopted perspective in order to interpret the data representation is that of an imaginary speaker-listener who has to deal with the linguistic input of the speakers of Bozen with no given context, a kind of “alien” speaker who is not provided with the normal socio-indexical coordinates commonly present in the speaking act. The task of this hypothetical speaker-listener would be that of assigning several possible variants of /r/ to some phonetic contexts, hence generalizing some rules.

Although the points seem to be randomly scattered with apparently no clear configuration, they are deliberately distributed in that manner and our “alien” can build some (though weak) hypotheses on when to use those different allophones.

The most peculiar aspects of this configuration are the following:

- In the top left quadrant, fricatives, vocalization and the deletion of /r/ are connected to a less prominent position such as the syllabic end (VRC as in [foe'ketta], or VR as in [baʁ]);
- In the central group, around the origin of the plane, the most frequent variants are gathered (as in the case of taps, shown in the analysis with tap) and some associations emerge, like that between approximants and the intervocalic contexts (VCV as in ['kwə:ʁe]);
- Down to the left, there is quite a clear and idiosyncratic relationship between the alveolar trill and the geminated context (for instance ['karro]).⁹

Despite the presence of slight irregularities, there is still a problem of hypervariability in

⁹ Interpretations, both on a phonological and phonetic level, are possible for these distributions, but are not included into the immediate objectives of this study.

which the imaginary speaker would find it hard to select criteria to determine how to use the several variants assigned to different contexts.

If we add the extra-linguistic data pertaining the speakers to the first picture, we will first obtain the map as shown in Figure 2 and then, after a process of reduction of redundant variables, the table in Figure 3.

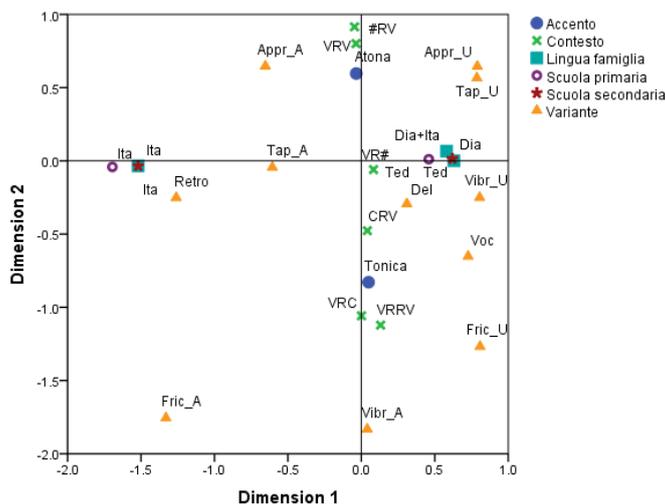


Figure 2. Map of linguistic and social variables I

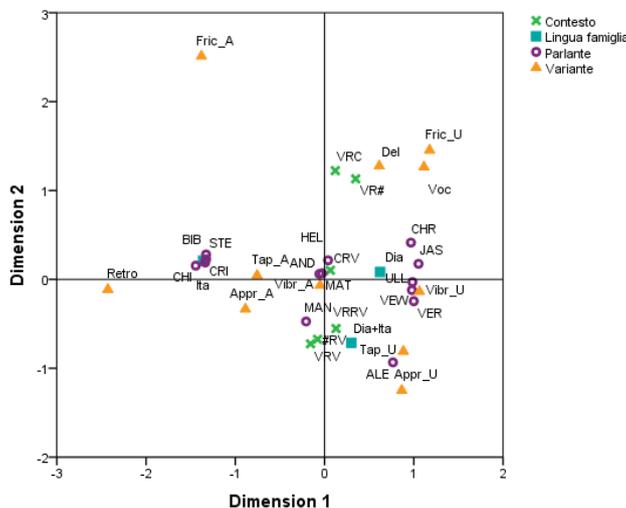


Figure 3. Map of linguistic and social variables II

The passage from the map in Figure 2 to that in Figure 3 is justified by the observation through which the variables related to the Italian or German school path (primary and secondary school) are closely connected¹⁰ to the family language variable, (that is the language or languages the speaker claims to have first learnt at a communicative level in his/her family). In addition, the family language variable has such an explicative power on

¹⁰ Primary school correlates with family language with a ratio of $R=0.815$ and secondary school with a ratio of $R=1.0$

the map's first dimension¹¹ that the other two variables are implicated and therefore redundant. Likewise the accent (such as in the previous representation) does not have a significant role in the variants' distribution¹². In the model shown in Figure 3 another variable has been adopted, the speaker's variable (referring to every single subject of the sample), as a supplementary variable¹³ in order to link possible varieties to groups of speakers.

With respect to the imaginary speaker's perspective described in Figure 1, the realm portrayed in Figure 2 and Figure 3 is more similar to the realm of communicative experience where linguistic and socio-indexical data are always simultaneously present. Even in the real world the speaker must learn to select only the information (in our case the variables) that is more prominent leaving behind those redundant or accidentally copresent, and to generalize common contextual features starting from the instances of each speaker.

A first examination of Figure 3 reveals how the whole configuration has changed from Figure 1 and that the allophones of /r/ organize themselves in the space by exploiting the two dimensions according to a different logic.

Along dimension 1, alveolars (left half-plane) oppose to uvulars (right half-plane), by grouping themselves respectively with Italian values, on one hand and with German dialect and Italian + German dialect, on the other, with reference to the family language variable. In this way two subsets referring to two macro-varieties of Italian emerge, thanks to an association between linguistic elements and socio-contextual elements. Hence a variety of Italian speakers can be distinguished, characterized by alveolars such as the trill, the tap, the approximant, the retroflex – and in a more peripheral position – the fricative; and a variety of German and early bilingual speakers, characterized by post dorsals and uvulars, subdivided into taps, trills, approximants, fricatives to the vocalization and deletion of /r/ and a third variety of German and early bilingual speakers.

Along dimension 2, on the vertical axis, sounds partially distribute according to the phonetic contexts, with a greater sensitivity to the contexts for allophones of the German speakers' variety (on the right), and on the contrary, more independently from the context for the allophones of the Italian speakers' variety. As far as the variety of German speakers is concerned, two clear relationships emerge: the first one between deletion, vocalization, uvular fricatives and the contexts of the syllabic end (VRC and VR#), and the second between approximants and uvular taps and intervocalic - or beginning of the syllable contexts.

Even in this case, an investigation into the social and linguistic implications of this configuration would bring us too far, but the rearrangement among the variants of /r/ on the basis of the “social” and “linguistic” axes seems very fruitful for the phenomenon to be better understood.

¹¹ Among the discrimination measures, the variable has a contribution of 0.941 (the maximum value is 1) on dimension 1.

¹² A further confirmation of the validity of the simplification of the model is given by the proportions of variance explained by the dimensions of the multiple correspondence analysis. While in Fig. 2 the model was able to realize the 79% of the variance of the data matrix, the model in Fig. 3 succeed in explaining 99% of the variance.

¹³ Supplementary variables do not actively intervene in the calculation that determines the factorial axes, but the values of the variable are calculated *a posteriori* and therefore projected on the plane, making it more soundable.

For instance it is possible to observe in Table 3 how the retroflex, although less frequent, is clearly characterized as belonging to Italian speakers, whereas the sheaf of uvular /r/ is to be associated to the Germanic trait¹⁴ (with the uvular fricative on the very right side). At the same time on the linguistic dimension, according to a less evident tendency, variants distribute in macro-contexts along a visible opposition between contexts at the beginning and at the end of the syllable. In particular, the majority of sounds, and most of all those belonging to the variety of Italian speakers, appears more assembled around the centre of the dimension, thus tending to be less dependent to the context.

Table 3. Allophones of /r/ placed according to the two dimensions.

Dimension 1 – Social					
+ Italian ←			→ + German		
Retroflex	Fricative A	Approx. A, Tap A	Trill A	Deletion	Approximant U, Tap U, Trill U, Vocalization, Fricative U
Dimension 2 – Linguistic					
+ end (VRC, VR#) ←		CRV		→ + beginning (VRV, #RV)	
(Fricative A), Fricative U, Deletion, Vocalization		Retroflex, Tap A, Approx A, Trill A, Trill U			Tap U, Approximant U

A final observation applies to the presence on the Cartesian plane of single subjects that seem to gather into three groups:

- a) A group on the left exclusively composed of Italian speakers (CHI, STE, BIB, CRI);
- b) A central mixed group, composed of bilinguals and German dialect speakers who speak a variety of dialect with alveolar rhoticity (HEL, AND, MAT, MAN)¹⁵;
- c) A group on the right composed of German dialect speakers (CHR, JAS, ULL, VEW, VER) and early bilingual speakers (ALE).

To sum up, the group (a) of Italian speakers adopts alveolar /r/, the group (b) of German and bilingual speakers also produces alveolar rhotics (although for different reasons), group (c) of German speakers creates uvular /r/ instead.

Thanks to the subdivision carried out on the map in Figure 3 we can differentiate the sample into three subsamples, each pertaining to a different language variety, and then conduct three separate analysis with the purpose to describe more precisely the contextual distribution of the different allophones.

In doing so, we are able to answer the first two questions of our whole investigation and we can therefore affirm that linguistic variables, the allophones of /r/ identified in the sample, organize themselves so as to detect at least three distinct language varieties to be identified

¹⁴ What is of great interest is the unmarked central position of the alveolar vibrant shared both by Italian and early Bilingual speakers. This position would strengthen the idea of a substantial perceptive ambiguity pertaining the place of articulation of the vibrant, (on the so called perceptual bridge see Engstrand et al., 2007).

¹⁵ In South-Tyrol dialects the /r/ is not always uvular but distributed in a scattered way along the many valleys, (Klein, Schmitt, 1969; Spreafico, Vietti, 2010).

thanks to the interaction with linguistic and social variables.

To answer the third question we need to analyze each of these three varieties.

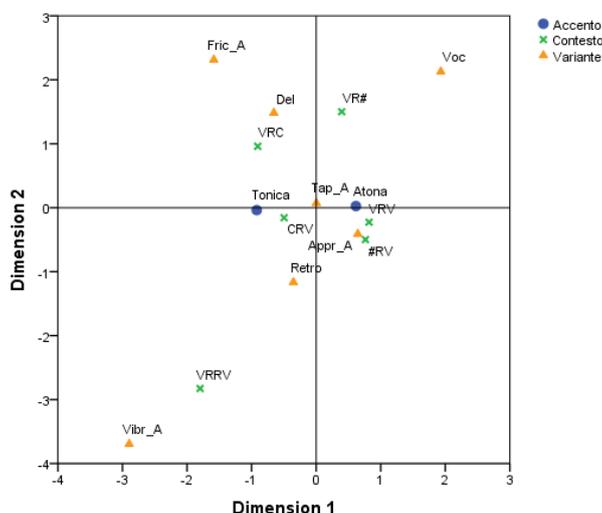


Figure 4. Map of Italian spoken by Italian speakers (a)

By observing the productions of group (a) of Italian speakers in Figure 4 we can assess which variants are used by speakers as well as how these are distributed on the basis of the contexts.

The allophones of /r/ clearly indicate the marginal and peripheral presence of trills, fricatives, vocalizations and deletions and the weak associations with the pre-consonantic contexts and word-endings, above in the picture, and the intervocalic geminated, below on the left in the picture. For reasons of method and entirety of the descriptive overview, these variants were kept in the analysis, but on a purely statistical level they should not have been taken into account as the sum of each of them does not reach twenty occurrences against the four hundred occurrences of the taps or the two hundred occurrences of the approximants. Indeed these latter two represent the core of the allophonic system of /r/ of Italian as spoken by Italian speakers: on one hand the central position of the taps on the Cartesian plane highlights a substantial lack of relationships with the contexts (hence defining a type of non marked allophone), on the other the position of the approximant shows an association with the intervocalic contexts and word beginnings.

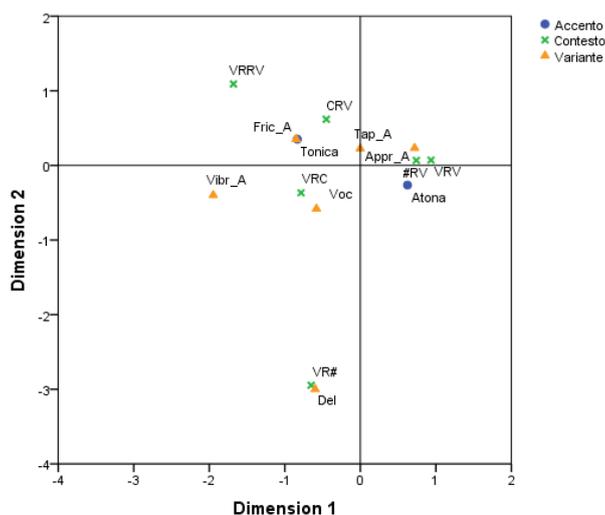


Figure 5. Map of Italian spoken by “alveolar German speakers” (b)

From the examination of the variety (b) of the central group in Figure 3 represented by the alveolar German speakers, we can determine at once if this variety is similar to that of group (a). Because of the alveolars being used, one would expect that allophones and their distribution would follow the “Italian” pattern.

At first glance, the answer seems to be negative as the disposition of the points on the plane is very different from the previous one. The main difference results from the great dispersal of allophones on the plane and a more regular association with the various contexts. Looking only at the most evident cases, for instance, it is possible to detect and observe cases of vocalization at the end of the syllable (VRC), of deletions at the end of the word (VR#) and in particular of alveolar fricatives in the CRV context. These traits seem to follow the same pattern of Tyrol dialect phonology and could therefore be understood as possible expressions of contact between phonologic systems: the same allophones of Italian but a different sound distribution as in accordance with Tyrolean dialect rules.

In other terms, alveolar German speakers do not duplicate Italian phonology and, although they adopt the same place of rhotics articulation, they distribute the sounds according to different rules.

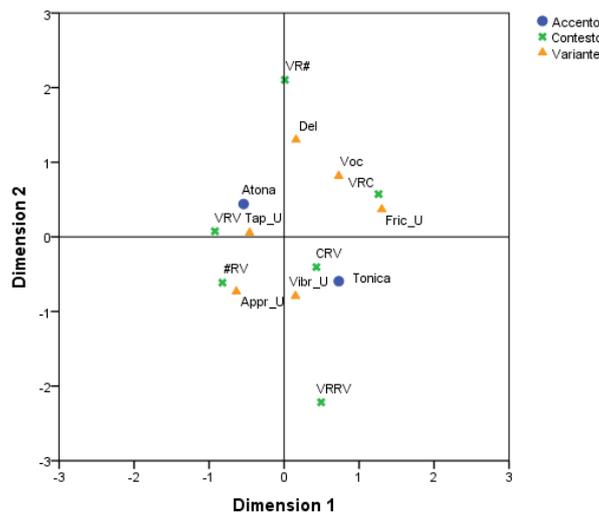


Figure 6. Map of Italian spoken by “uvular German speakers” (c)

Finally, we have carried out the analysis in Figure 6 according to the Italian variety spoken by “uvular German speakers”, group (c). The presence of an uvular tap (Tap_U) should first be noted as it is a rarely attested sound among languages around the world. This could be seen as the result of a phonological contact process, which is the origin of a new sound as a consequence of the transfer of the manner of articulation manner and not of its place.¹⁶

In addition, it must be pointed out how, despite it being an articulatory marked sound, the tap frequency is high and its distribution, similar to other Italian varieties, does not seem to be conditioned by phonetic contexts, if considering its central position.

Secondly, and similar to variety (b), allophones distribute on the Cartesian plane and create associations with contexts. Specifically, there are evident relationships among fricatives, vocalization and deletion at the end of the syllable (VRC) and between uvular approximant and word beginning (#VR), just as in (b).

4. Conclusions

The present study focused on the analysis of the variation of /r/ in a corpus of Italian variety spoken in Bozen, starting from the assumption that in order to determine the phonology of a linguistic system the speaker must necessarily rely (also) on non-linguistic information available in the communicative context. This postulation, theorized in sociolinguistics and recently in sociophonetics, is in itself little more than a truism. In our investigation we tried to show the empirical consequences of such an assumption, representing by means of a multivariate analysis technique other possible “worlds”, that is data representations which take into consideration linguistic and socio-contextual variabilities. In this way it was possible to verify the soundness of our representations according to their ability to explain

¹⁶ Due to a lack of significant experimental documentation on the rhotics in South Tyrol dialects, it cannot be excluded the pre-existence of this sound in the source language and to conclude that the development of an uvular tap can be considered as a contact effect.

data variability.

Through the multiple correspondence analysis based on given data, we highlighted which associations can emerge when social variables are to be excluded. In the map (Figure 1) relations between variants and contexts, when they emerge, appear as weak associations among many shapes per each context. The phonological plane appears unrealistic, so opaque and marked that the speaker would find it difficult to deal with. The answer to our first question, whether allophones alternate within one or more grammar systems, is that we are facing not just one grammar but probably multiple grammars.

Figure 3 allows us to answer our second question and broadly understand on which assumptions the partition of allophones is subdivided into subsets. The introduction to non-linguistic data – in this case, a simple statement on the first language spoken in the family – changes the configuration of points, clearly distinguishing Italian alveolars and German uvulars. This representation stands as indirect proof of the integration of linguistic and non-linguistic data and of the influence that the latter exerts in defining the border of varieties' grammars.¹⁷

Once the varieties are identified by means of socio-linguistic integration, it is possible to answer our third question related to the distribution of allophones within each single variety (see Figure 4, Figure 6). Not only do the forms/shapes in the three varieties not match entirely, but allophones do not share distributional properties. In conclusion, it can be assumed that it is not so much the identification of the single sounds or their articulatory place that help to identify speakers as “Italian” or “German”, as the entire knowledge of the sociophonetic variation of /r/, made of intrinsic elements (sound features), distributional properties and contextual factors.

Lastly, a remark on the methodological type concerning the statistical technique we adopted should be made. The multiple correspondence analysis appears as a technique simply aiming at exploring data rather than modelling them or defining hypotheses. Being a tool of simple representation of complex data matrixes has its own major advantages and consequently its undeniable limits. By using it in this study we were able to:

- Represent a multidimensional phenomenon and reduce the number of variables to two axes;
- Express relationships among the values of the variables and, in some cases, delete those values and variables appearing as redundant.

In addition, the result of the analysis represents an excellent starting point to further carry out analysis of classification and grouping or merely to make hypotheses to be examined with more powerful techniques such as generalized linear models (GLM).

Acknowledgement

The research is financed by the Provincia Autonoma di Bolzano - Alto Adige, Ripartizione

¹⁷ Showing, at an experimental level, the implementation of two data sources in the complex knowledge of the linguistic variation is undoubtedly a complex task.

allo studio, Università e ricerca scientifica, project grant “The articulatory sociophonetics of bilinguals in South- Tyrol: The Ultrasound Tongue Imaging potential”, 2013-2016; and by: Free University of Bozen-Bolzano, BW2050 SCUTI.

References

- Bybee, J. (2001). *Phonology and language use*. Cambridge, Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511612886>
- Celata, C., & Vietti, A. (2011) Prospettive sociofonetiche sulla variazione. *Studi Italiani di Linguistica Teorica e Applicata*, 40(2), 333-346.
- Cutugno, F. (2007). Criteri per la definizione delle mappe, esempi di mappe e di vignette per il gioco delle differenze. [Online] Available: <http://www.clips.unina.it>
- De Masi, S. (2007). Criteri per le liste di lettura. [Online] Available: <http://www.clips.unina.it>.
- Di Franco, G. (2006). *Corrispondenze multiple e altre tecniche multivariate per variabili categoriali*. Milano: FrancoAngeli.
- Docherty, G., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*. 142. 42-56. <http://dx.doi.org/10.1016/j.lingua.2013.01.011>
- Engstrand, O. Frid, F., & Lindblom, B. (2007). A perceptual bridge between coronal and dorsal /r/. In M-J. Solé P. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 175-191). Oxford: Oxford University Press.
- Ervin-Tripp, S. (1971). *Social dialects in developmental sociolinguistics*. In R. Shuy (Eds.), *Sociolinguistics: a cross-disciplinary perspective* (pp. 35-64). Washington, D.C.: Center for Applied Linguistics.
- Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 409-438. <http://dx.doi.org/10.1016/j.wocn.2005.08.002>
- Foulkes, P., Docherty, G, & D. Watt (2005). Phonological variation in child directed speech. *Language*, 81, 177-206, 2005. <http://dx.doi.org/10.1353/lan.2005.0018>
- Greenacre, M. (2007). *Correspondence analysis in practice*. Boca Raton, FL: Chapman & Hall.
- Hymes, D. H. (1972). On communicative competence. In J. Pride, & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). London: Penguin.
- Johnson, K. (2005). Decisions and Mechanisms in Exemplar-based Phonology. *UC Berkeley Phonology Lab Annual Report*. 289-311.
- Khattab, G. (2002). /r/ production in English and Arabic bilingual and monolingual speakers. *Leeds Working Papers in Linguistics & Phonetics*, 9(1). 91–129.

- Klein, K., & Schmitt, L. (1969). *Tirolischer Sprachatlas*. Innsbruck: Tyrolia-Verlag.
- Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, 34, 500-515. <http://dx.doi.org/10.1016/j.wocn.2006.05.002>
- Labov, W., Ash S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4), 431-463. <http://dx.doi.org/10.1111/j.1467-9841.2011.00504.x>
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*, Malden: Blackwell.
- Mioni, A. (1990). L'italiano nelle tre comunità linguistiche tirolesi (con particolare riguardo per la pronuncia). In K. Egger, & L. Lanthaler (Eds.), *Die deutsche Sprache in Südtirol: Einheitssprache und regionale Vielfalt* (pp. 65-76). Wien: Folio.
- Pierrehumbert, J. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge, MA: MIT Press.
- Spreafico, L., & Vietti, A. (2010), Sistemi fonetici in contatto. La variabilità di /r/ nell'italiano di tedesofoni altoatesini. In M. Pettorino, A. Giannini, F. Dovetto (Eds.), *La comunicazione parlata 3* (pp. 127-136). Napoli: Università degli Studi di Napoli L'Orientale.
- Vietti, A. (2008). Contesti d'uso in repertori linguistici complessi. Tentativi di descrizione multidimensionale dei dati della Survey Ladins. *Mondo Ladino*, 31, 239-266.
- Vietti, A. (2014). Alcune riflessioni sulla teoria degli esemplari e la variazione linguistica. *Rivista Italiana di Dialettologia*, 36, 7-22.
- Vietti, A. Language contact and sociophonetic variation. In S. Calamai Silvia, C. Celata, & L. Ciucci (Eds.), *Sociophonetics, at the crossroads of speech variation, processing and communication*. (pp. 67-69). Pisa: Edizioni della Normale.
- Vietti, A., & Spreafico, L. (2008). Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian. In P. Warren (Eds.), *Laboratory Phonology 11* (pp. 145-146). Wellington: Victoria University Press.
- Vietti, A., Spreafico, L., & Galatà V. (2014) An ultrasound study on the phonetic allophony of Tyrolean. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. [Online] Available: <http://www.icphs2015.info/pdfs/Papers/ICPHS0779.pdf>.
- Vihman, M., & Keren-Portnoy, T. (2013). *The emergence of phonology*. Cambridge: Cambridge University Press.
- Wolfram, W. (2006). Variation and Language: Overview. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 333-341). Oxford: Elsevier. <http://dx.doi.org/10.1016/b0-08-044854-2/04256-5>

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).