Documenting Italo-Romance heritage languages in the Americas

Luigi Andriani[1], Jan Casalicchio[2], Francesco Ciconte[3], Roberta D'Alessandro[1], Alberto Frasson[1], Brechje van Osch[4], Luana Sorgini[1], Silvia Terenghi[1]
1=Utrecht University, UiL-OTS; 2=University of Palermo; 3=University of Insubria; 4=University of Tromsoe

1st draft – September 2020

This article describes the process of preparation and implementation of a data collection enterprise targeting Italo-Romance emigrant languages in North and South America. This data collection is part of the ERC *Microcontact* project, aimed at understanding language change in contact by examining the language of the first-generation Italian emigrants in America (G1 henceforth, https://microcontact.sites.uu.nl/).

The speakers involved in our study are first-generation Italians (so-called *émigrés*), mostly emigrated to North and South America between the 1940s and the 1960s, and second and third generation speakers (heritage speakers, 'HS'). The population of Italian emigrants is next to ideal in a study on language contact, because most of them are tendentially monolingual speakers of an Italo-Romance variety other than Italian (traditionally referred to as Italian "dialect") at the moment of their arrival in America: their heritage/minority languages are not taught in Italy, therefore the fact that these speakers are monolingual reflects the fact that they did not get any education. At their arrival, they entered into sudden intensive contact with other Romance languages: the ones we focus on are French (in Quebec[1]), Spanish (in Argentina), Portuguese (in Brazil). We also consider these varieties in contact with Italian (in Italy), bearing in mind that this contact is very different from that which is found in America, first and foremost because it has carried out for much longer, and because there are more speakers in Italy. Finally, we also investigate Italo-Romance speakers in contact with English in the United States as control group.

The project follows the evolution of these contact situations, by focusing on three language phenomena in seven Italo-Romance varieties. The phenomena that we selected are: differential object marking ('DOM'), deixis and demonstratives, and subject clitics ('SCLs') / null subjects. Other language features, such as topicalization and unaccusativity, are also taken into account to a lesser extent. These phenomena have been selected because they are well documented for the languages at issue and their diachronic evolution can be tracked rather straightforwardly. For each of these phenomena we checked whether they are preserved in the various contact situations, and in which syntactic contexts. Furthermore, we were interested in the influence of the contact languages on the languages of first-generation (and subsequent generations) speakers.

The varieties that were originally selected for investigation are Piedmontese, Venetan, Tuscan (Florentine and Sienese), Eastern/Coastal Abruzzese, Neapolitan, Salentino, and Sicilian. These varieties were chosen for several reasons: they maximally instantiate the variation recorded for our target phenomena across Italo-Romance and are the mostly spoken by Italian emigrants in the Americas. Moreover, they all have a long literary tradition, with the exception of Abruzzese, which was selected because of the wide documentation on the language available to the PI. This was crucial for us to be able to compare the diachronic evolution of the phenomena that we are considering with their change in contact. While Italian and Italo-Romance languages have been in extensive contact in the last 70 years, not many

---

[1] See below and Section 3 for details on why, in a subsequent fieldwork, we targeted Italian emigrants in the French-speaking part of Belgium.

people could speak Italian in the beginning of the previous century. This means that their development from Early Romance to the 20th century is mainly endogenous, and not due to contact.

The languages selected turned out to be not all optimal. In particular, it was not possible to find Tuscan, Salentino and Neapolitan speakers. A very large community of Calabrese and Friulian speakers was identified instead during fieldwork. In order to have a large and consistent set of data, it was decided to exclude the varieties with very few speakers and introduce Friulian and Calabrese instead.

An additional change of plan that we made regards the places where we carried out our fieldwork. Recall that the locations and languages that were originally selected were Argentina/Argentinian Spanish, Brazil/Brazilian Portuguese, Quebec/Quebecois French, and Italy/Italian. English was also included for control: we selected the English varieties spoken in New York and Boston. As a matter of fact, fieldwork research showed that the Canadian situation was rather different from what we had envisaged. Speakers of this area were in fact mostly Italo-Romance/French/English trilingual. English in particular was very perceptible in their spoken language, and therefore constituted an interference that was difficult to overcome. Instead, after some research, it became clear that Italo-Romance speakers in French-speaking Belgium present a profile that can be compared to that of our target population in Argentina and Brazil. In Belgium, we found speakers who had left Italy in the 1940s-1960s. Despite the geographic proximity between the two countries, their relationships with their homeland were as severed as those of the Italians who emigrated to South America. Moreover, no interfering additional languages (besides the target varieties and the contact language) were detected. Therefore, it was decided to move the data collection for contact with French from Quebec to Belgium. Finally, contact data from Italy were left as a backup, for the reasons given above.

This article is based on the several fieldwork sessions carried out by the Microcontact team. The first one targeted Argentina only (cf. Section 3 below) and took place in May 2018. It was followed by three parallel fieldwork sessions (March/April 2019) completed in Argentina, Brazil, and Quebec. The control fieldwork took place in New York City between October 2019 and January 2020, while a pilot fieldwork in Belgium was carried out in November/December 2019. The analyses that were performed over this first set of collected data were meant to be checked against more data in subsequent fieldworks; this has however been prevented by the current COVID-19 pandemic.

Because of the age of the speakers and the conditions in which the fieldwork was going to take place, we expected data collection to be quite difficult: in this paper, we discuss the issues arisen during fieldwork preparation and performance. Each section focuses on a specific stage of the data collection and is structured as follows: first, we introduce the background, i.e. the information that is already available in the literature and how we planned to use it to carry out our data collection. Then, we describe what the actual situation turned out to be. We conclude every section with a list of tips and warnings about what needs to be taken into account when performing a similar task.

More specifically, Sections 2 to 4 address issues related to our fieldwork, with a special focus on its practical (Section 2) and theoretical (Section 3) preparation, and on the main problems encountered when working with elderly population.

2. Fieldwork

2.1 Where we started
Fieldwork for Italo-Romance varieties in the Americas is rather unique and different from other kinds of fieldwork, in that it targets varieties that have been known, spoken and in many cases

also written for centuries, but in an environment which is not their original one. Furthermore, these languages have undergone contact with other Romance varieties for a considerable amount of time, and are therefore rather difficult to understand even for native speakers of the baseline varieties in Italy. On the one hand, this experience in not comparable to that of documenting a previously undocumented language of uncertain family; on the other hand, it is not as simple as carrying out a dialectological inquiry in Italy, where people share a common language (Italian) and can understand instructions and translations into Italian, and share at least one language with the interviewer.

In what follows, we report our fieldwork experience, focusing on the actual setup of Italo-Romance speaking communities for this section, and on the results of the syntactic inquiry in the next one.

Before turning to the presentation of each fieldwork area, some general consideration regarding data protection protocols that hold in Europe but not elsewhere. No fieldwork can start without a certified ethical clearance and an approved data protection protocol in compliance with the latest GDPR (General Data Protection Regulation 2016/679; see Leivada et al. 2019), which enforces strict, in some parts problematic, directives within the EU. Of course, these directives may not be entirely consistent with those of the non-EU countries. A challenging task is to check the GDPR against the regulations of the target country, seeking for an optimal level of mutual adherence. This can be done with the support of the embassies, which, however, can be slow, or even unresponsive. An effective alternative is to invite universities in the target countries to co-supervise the fieldwork, thus ensuring that data collection and storage comply with the regulations of both the EU and the target-country.

In addition to the GDPR guidelines, each research institution may have its own internal procedures for data protection and privacy impact assessment as well as different instructions for the fieldworker's conduct and safety. It is important to establish immediate contact with the offices in charge and provide a detailed description of the intended survey, restating, at cost of sounding redundant, the non-private nature of the questions and tasks involving the informants. Furthermore, it is worth bearing in mind that, at administrative level, all kinds of fieldwork proposal tend to be checked uniformly for ethical compliance, resulting at times in non-discrete evaluations of the potential infringements and risks, which instead vary significantly among science fields.

Regarding the actual data collection, it needs to be considered that Italo-Romance communities in the Americas have very different characteristics. In this subsection, we review the information about Italo-Romance communities available in the literature before the beginning of the project.

2.1.1 Argentina
Argentina was a very popular destination for Italian immigrants in the 19th and 20th century. According to the website of the Ministry of Foreign Affairs, Argentina received 57% of the total amount of Italian people who emigrated overseas between 1946 and 1955[2]. At first, Italian immigrants moved there only temporarily to have a better income, so as to later improve their quality of life in Italy. This form of immigration started between the 18th and the 19th century, but it became a mass phenomenon in the last quarter of the 19th century. Temporary immigrants either stayed in Argentina for some years and then settled back to Italy, or they were seasonal workers, leaving Italy in autumn/winter and coming back in spring/summer. In the period of mass immigration, this trend was flanked by permanent immigration movements, where families would move altogether and settle in the new country (Ferrari 2008). Geographically, the first immigrants were predominantly northern Italians; in the last years of the 19th century,

---

[2] These data are taken from https://www.esteri.it/mae/doc_osservatorio/rapporto_italiani_argentina_logo.pdf.

immigration from the South grew more and more, and it became majoritarian before World War I.

The areas most interested by the arrival of Southern Italians were cities like Buenos Aires, Córdoba and Santa Fe. In these cities, the number of immigrants from various countries was extremely high, and Spanish was not only the official language, but also the lingua franca for immigrants that had different first languages ('L1s'). This was the optimal condition for the emergence of hybrid varieties like Cocoliche (see a.o. Bagna 2011), a contact variety often described by contemporary sources as a mix of Spanish and Italian, although we should point out that the Italian elements often came from Italo-Romance varieties rather than from Italian.

In some areas of Argentina, however, linguistically homogeneous communities arose, creating linguistic islands, such as in the Boca and Colonya Caroya. The first one, in the Boca, a district of Buenos Aires, was generated by Genoan immigrants, whose dialect is described as extremely popular and alive since the second half of the 19th century, but it was reported as essentially dead during the 1980s. In contrast, Colonya Caroya, a town in the province of Córdoba, was home to immigrants from Friuli, and the language was still alive and popular in the 1980s (Meo Zilio 1988).

As for Southern Italo-Romance varieties, we only had a rough idea of the situation before the fieldwork took place.

2.1.2 Brazil

Brazil used to be one of the main destinations of Italian emigrants in the second part of the 19th century. The main areas of Italian immigration were the states of São Paulo (especially for immigrants from southern Italy) and Rio Grande do Sul, the southernmost Brazilian state (especially for immigrants from northern Italy).

In the case of São Paulo, Italian cultural heritage is still alive, but HLs died out fast, as the communities assimilated to the Portuguese-speaking majority. Moreover, Brazilian authorities carried out campaigns against the use of foreign languages (including Italian and Italo-Romance varieties) in the 1930s, which ultimately led to a ban on their use in the 1940s. Consequently, there are no traces left of southern Italo-Romance varieties in São Paulo, nor of the "Paulistano" Italian, a koine variety of Italian strongly influenced by Portuguese that was used in the city at the turn of the 19th century (Cenni 2003).

The case of northern immigrants to Rio Grande do Sul is different: they settled in extremely isolated mountain areas, a condition that allowed their varieties to resist the ban on the use of the language imposed in the 1940s and pressure of Portuguese in recent years. Almost half a million people, descendants of the original settlers, still speak a northern Italo-Romance variety in the area, a phenomenon that has been the focus of many sociolinguistic studies conducted in Brazil. These speakers, despite being mainly third or fourth generation HSs, are native speakers of an Italo-Romance variety and they do not have any knowledge of Italian in most cases. Therefore, they are good candidates for the study of contact with Portuguese, their second language. Since these communities are particularly isolated and not very easy to reach, during the year prior to the fieldwork, relationships were developed with the Venetan Association of Rio Grande do Sul and the Federal University of Santa Maria, as well as a few other contacts in the area, with the goal of developing trusted local contacts.

The number of Italian immigrants to Brazil started to decline at the beginning of the 20th century and almost came to a stop after World War II, thus earlier than in other American countries. This makes it very difficult to find G1 speakers who are still alive. An exception to this situation is the city of Porto Alegre, the capital of the state of Rio Grande do Sul, to which immigration from southern Italy continued after World War II. Therefore, we selected the state of Rio Grande do Sul as our target area for the fieldwork in Brazil, as it is home to both G1 and HSs of both northern and southern varieties.

2.1.3 Quebec

The situation in Quebec was expected to be yet dissimilar. To begin with, the demographics of the Italian emigrant population is quite different than in the rest of America: emigration to Canada, and more specifically to Quebec, is relatively more recent than that to the other areas under investigation. Although the first records for the Italian emigration to Canada date back to the last quarter of the 19th century and the migration flow never completely stopped, the movement intensified only between 1951 and 1967. Due to this demographic difference, the majority of speakers of the first generation in Canada were typically not (completely) illiterate when they left Italy, as they had received at least some formal education in Italian, besides being exposed to the national variety in the increasingly popular medias. Therefore, we knew that Italian, or at least a non-standard variety thereof, was not going to be a completely negligible source of interference on the dialects of our informants. We based our knowledge on available studies on the language(s) spoken by the Italian community of Montreal (focusing on their Italian: Reinke 2014 and the extensive production by Villata, e.g. Villata 2010).

Moreover, the overall migration flow to Canada was never comparable in range to that of our other research areas: its peaks, registered in 1956, 1958, 1966, and 1967, were of roughly 28,000 people a year.[3] Therefore, we expected to find less participants in this research area than elsewhere even if we do not have the exact numbers for Quebec only.

Finally, some parts of Quebec are *de facto* bilingual areas: while French is the only official language of the province, English is widely spoken, especially in Montreal. However, we relied on general knowledge of pro-French campaigns and policies, particularly strong during the 1970s, for the inclusion of the area in our study. Still, we were prepared to find some (reduced) instances of speakers also proficient in English, at least to some extent: those would have ideally been excluded from our study, to avoid the confounding factor of an additional variety, let alone a non-Romance one.

3.1.4 US

Italo-Romance varieties have been exported to the US since (at least) the 19th century; the census from 1880 recorded 81,249 Italian migrants, and this steeply increased up to 4,114,603 by 1920 (cf. Cavaioli 2008). In the year-2000 census, "Italian" was reported to be spoken by about a 1,000,000 people in the US, with the most significant numbers concentrated in the Northeast of the country,[4] where we carried out our fieldwork. However, according to the multi-year American Community Survey 2009-2013, there has been a decrease of about 300,000 speakers, i.e. a third less in over 10 years. This is most likely due to a rapid language shift to English only, which typically occurs within the third generation in Italian communities abroad (see De Fina 2014; for NYC, see Haller 1987, 1993). For this reason, the vitality of Italo-Romance varieties spoken in the US is endangered, as it is virtually impossible to come across Italo-Romance HSs past the first US-born generation.

Moreover, these statistics lump together the languages imported by Italian migrants under the umbrella term "Italian". This is largely inaccurate, as pre-WWII migrants mainly exported their local languages, rather than Italian, as we know. This situation changed after WWII, particularly after 1965 with the Immigration Reform Law, thanks to which the families

---

[3] Source: ISTAT, http://seriestoriche.istat.it/fileadmin/documenti/Tavola_2.9.1.xls.

[4] New York State (294,271), New Jersey (116,365), Pennsylvania (70,434), Massachusetts (59,811), Connecticut (50,891), Maryland (13,798), Rhode Island (13,759), and Virginia (10,099) [https://www.census.gov/population/cen2000/phc-t20/tab05.pdf]. Notice that not for all countries can we find the same level of accuracy and documentation regarding the population census. We report here what was available to us before we went on fieldwork, and the information on which we based our planning.

of Italian migrants were allowed to legally live and work in the US. These more recent waves of migrants had an 'Italianizing' impact on the local Italo-Romance languages, as speakers were more schooled (in Italian) and, hence, no longer Italo-Romance monolinguals like the previous migrants (Haller 1991: 391-392; De Fina & Fellin 2010). Thus, the speakers' competence in their local languages started being affected – to differing degrees – by the imported Italian, as well as English. Viceversa, the Italo-Romance substrate, alongside English, also permeated into the regional Italian of these migrants. Moreover, the co-presence of more-or-less-intelligible local varieties brought about the need for a linguistic koine, i.e. a shared Italo-Romance variety intelligible to everyone, and this has been the focus of most studies on Italian communities in English-speaking countries. Indeed, this situation is documented for New York by Haller (1987 *et seq.*) and is common to other urban contexts with Italo-Romance–English contact, e.g. Sidney Italian (Bettoni 1990, 1991) and Montreal Italian (Reinke 2014).

As far as New York is concerned, Haller's (1987, 1993, 1997a, 1997b, 2002) work provides a solid description of the (socio)linguistic situation of the Italian community between the years 1980 and 2000. He proposes a multilingual continuum for Italo-Romance varieties, which are "used, besides English, with various degrees of competence, according to generation, time of emigration, and education" (Haller 1987: 396): "*'Standard' dialectal Italian*, *Italianized dialect*, *pidginized American Italian*, and *archaic dialects*". As for the koine variety, Haller confirms that "[t]he migration from the depressed South to Rome and Northern Italy and the emigration to the United States both acted as 'Schools of Italianization', exposing individuals for the first time to other dialects and languages and forcing them to develop a *lingua franca* in order to be able to communicate with each other" (Haller 1987: 393). Hence, while the Italo-Romance local dialects would be employed within the family and closer-knit circles of fellow countrymen, this Italian koine has been functioning as the 'community language' for decades (Haller 1991; 1997a: 401). A few decades later, this situation is currently at its peak, with Italo-Romance varieties slowly fading away and Italian taking over (if at all), next to English.

2.1.5 Interim summary

Despite our awareness of the social differences between the different areas where we intended to collect data, before starting our fieldworks we were working under the hypothesis that the most relevant socio-historical conditions were comparable for all the targeted countries: we expected to find G1 speakers with very low competence, if any, in Italian (even in its regional varieties), who maintained little contact with the respective communities of origin in Italy. Moreover, we expected the varieties under analysis to be faithfully preserved by the communities abroad, at least as home languages, and as such passed on to the following generation(s). The advantages of such scenario would have been the possibility to systematically exclude the influence of external factors (e.g. competence, language exposure, age of bilingualism onset, *etc.*, as well as socio-historical variables) on the development of the phenomena under analysis, and to assess, for each HS, the input language (i.e. the language spoken by their parents, the G1) with an outstanding level of detail also at the microvariation level.

However, in some cases the socio-historical differences proved to be more far-reaching than expected and to have a non-negligible bearing on the linguistic profile of our informants. These issues are discussed in more detail in the following section, along with some practical matters that should be considered when organizing a fieldwork, and followed by the solutions we found to the various issues that arose.

2.2 What we found

2.2.1 Argentina

As we discussed above, Argentina was the destination of huge immigration waves since the last decades of the 19th century. When trying to reach out for associations of Sicilian, Neapolitan and Abruzzese people through emails, however, we received an enormous amount of Non-Delivery Reports (NDRs) that led us to think that they were mostly not active anymore. At that point, we asked information to Facebook groups dedicated to Italians' descendants in Argentina, and to distant relatives[5]. Neither crowdsourcing nor any subsequent attempts helped identify speakers who could qualify for our inquiry.

Due to the lack of information, we decided to set a pre-fieldwork, with the specific goals of checking the current situation within the Italo-Romance communities and establishing a network of informants for the following fieldwork. This pre-fieldwork was carried out only in Argentina as that was the area where least contacts could be found before the actual fieldwork. Once in Argentina, our researcher was able to establish a good network after visiting the presidents of some associations and some colleagues at local universities.

The pre-fieldwork was followed by the actual fiedwork, during which we targeted immigrants who moved to Argentina after World War II, as well as their few descendants who had acquired the Italo-Romance variety. As in other American countries, the institutions, especially schools, played a major role in the diffusion of the monolingual Spanish model. The researchers were told various times that teachers explicitly suggested, or even ordered, the parents to speak only Spanish to their children, in order to avoid "confusion" in the child's language ability[6].

Unfortunately, these interventions were effective most of the times, so that the Italo-Romance varieties were abandoned by almost all immigrants. Exceptions are found especially when there were elder family members (especially grandparents) who never managed to learn Spanish and thus kept speaking their Italo-Romance L1 to their grandchildren.

Despite the geographical distance, during the Argentinian fieldworks we found that the local Italian community managed to keep strong bonds to their hometowns by frequent visits to Italy (especially starting from the 1980s) and through the countless regional and local associations. According to our informants, the associations were particularly active in the 1940s-1970s in recreating a sense of Italian community through recurrent parties and celebrations. The members of the associations tried and still try to maintain the traditions of their home regions, such as religious celebrations, typical food and even clothing. Curiously enough, the only thing they usually did not maintain is the local language, switching to Italian or even, and most commonly among younger members, to Spanish.

More generally, Italian gained ground because of marriages between people originating from different Italian regions, who choose to speak Italian to their children to keep a stronger bond with their home country. In the bigger cities, there are also schools of the Italian community, in which some subjects are taught in Italian. Finally, although the immigrants and their descendants feel a particular link to their region, they feel proud of Italy as a whole and

---

[5] In the case of Abruzzese, one of our contacts wrote to us: "La Argentina ha recibido inmigrantes de todo el mundo que han traído sus idiomas y dialectos, pero al haberse mezclado con toda la sociedad no sabría si continúan hablando el dialecto. Por ejemplo mi abuelo Francisco no hablaba su dialecto, hablaba español." ['Argentina has received immigrants from all over the world, who brought their languages and dialects with them, but as they have integrated in society, I'm not sure whether they still speak their dialects. For instance, my grandfather Francisco didn't speak his dialect, he spoke Spanish.']

[6] As one of our speakers told us during the interview: "In taule a si fevelave simpri furlan, ai vut tancj di chei problems ta scuele, parcè che i disevi peraules in furlan, e an clamat a me mari, che no si feveli plui il furlan parcè che no si podeve." ['We would always speak Friulian at home, I had so many problems at school because I would say words in Friulian, and they called my mum so that we would not speak Friulian anymore because it was not allowed.']

identify with it, especially when they talk to people who do not have an Italian emigration background. As a consequence, most first-generation immigrants (let alone the subsequent generations) speak Italian besides their local variety, and they all insisted in speaking Italian to the fieldworkers.

The associations were very useful in our search for informants, since they know most members of the community. However, even they could not indicate more than two or three people each, because there are not many speakers left. On the other hand, some associations offer courses in their regional Italo-Romance variety, which are followed by second or third generation immigrants who never developed a high proficiency in the language and wish to improve it (or even to learn it from scratch). In the Friulian association of Buenos Aires and in the Piedmontese association of Córdoba, for example, 8-10 people followed the language course. They were a useful source for our search for HSs.

The informants we interviewed were found mainly through members of the associations and, in Santa Fe, thanks to the Italianists of the Universidad Nacional del Litoral, who are running a project on the Italian cultural heritage in this city.[7] In three cases we found speakers of Cocoliche (see 2.1.1 above): unfortunately we could not interview them, because they belong to the oldest members of the community and were unable to perform our tasks. In addition, they usually could not distinguish anymore between Cocoliche and their own Italo-Romance variety.

Overall, most informants were kind but rather suspicious at the beginning, especially in the bigger cities: due to the general unsafety, they all refused to receive the fieldworker at home, unless he was accompanied by a member of the community they already knew. As a result, when possible, the interviews were held in the rooms of the emigrants' associations. However, in some cases they had to be carried out in bars, a less than ideal place, as informants could be distracted, the audiostimuli of the questionnaire were difficult to understand because we had to lower the volume, and the recordings were affected by background noises.

As far as the geographic distribution is concerned, we observed that the Italo-Romance varieties are still found in the main cities. Nowadays, however, their use is limited to the family group, and there are very few speakers with a high proficiency in the second and third generation. In the smaller centres, on the other hand, the Italo-Romance varieties have virtually died out. One significant example is Colonia Caroya, in the province of Córdoba: until some decades ago, Friulian was the main language (Spanish being the official language only), to the extent that, according to our informants, it was impossible to find a job there if you did not speak Friulian. Nowadays, however, the situation has radically changed: we could only find three informants (all above 70 years old), while the rest of the community speaks neither Friulian nor Italian.

2.2.2 Brazil

In Brazil, Italo-Romance varieties survived mainly in the countryside and, to some extent, in bigger cities in southern areas of the country, as we highlightedabove, in section 2.1.2. Immigration from Italy after World War II was very limited, therefore it has been quite complicated to find first-generation immigrants with the right profile for our fieldwork. The state of Rio Grande do Sul was the best option, as both northern and southern varieties are still spoken in the area by G1 and HSs.

10 interviews were carried out in Porto Alegre, which is home to a big community of Calabro-Lucanian speakers, as well as smaller Sicilian and Abruzzese communities. The language used by the interviewer was mainly Portuguese. The first problem encountered in Porto Alegre was the general sense of distrust shown by local associations of HSs towards the

---

[7] The responsible of the project is prof. Adriana Crolla (http://www.fhuc.unl.edu.ar/portalgringo/crear/gringa/).

research; in particular, one of the local associations, the Abruzzese Association of Rio Grande do Sul, refused to help in finding informants for the interview. A few participants were found as a result of active posting on Facebook groups of descendants. The best result accomplished in Porto Alegre, however, was the successful cooperation with the Calabrese Centre of Rio Grande do Sul; this contact was established prior to the fieldwork and helped in finding Calabro-Lucanian G1 and HSs. The Calabrese community in Porto Alegre is formed by immigrants (and descendants) from the town of Morano Calabro and they therefore speak exactly the same variety, the Moranese dialect. However, also in this case, the general sense of distrust was evident; getting a written informed consent was particularly problematic, since most of the participants were afraid it could be a scam.

The inner part of the state of Rio Grande do Sul is home to almost two millions of descendants of immigrants from northern Italy; most of them maintained the language of their ancestors across generations, as a result of the conditions of isolation in which these communities live. 42 interviews were carried out in the area of the Serra Gaucha and in the Fourth Colony of Italian Immigration. The languages used by the interviewer were Venetan and Portuguese. The condition of isolation of these communities, which helped in the preservation of three northern varieties (Venetan, Friulian and Lombard), also represented the main obstacle in reaching the speakers. In some cases, the only way to get to the villages was by car. The contacts we established in the area made it possible not only to physically reach the speakers, but also to help the communication with old people who, in some cases, never left their villages and were therefore not always willing to talk to a foreigner or be recorded. Unlike the Calabro-Lucanian informants in Porto Alegre, speakers of northern varieties in Rio Grande do Sul are the descendants of immigrants that came from different areas in Italy. Their dialects are not exactly identical to each other, nor to the varieties of the languages spoken in Italy. Besides, isolation from Italy made it possible for archaic features of the languages to be preserved in these varieties. The written informed consent was problematic for informants in the Serra Gaucha and in the Fourth Colony of Immigration as well. The interviewer decided to opt for a recorded oral consent by the informants: the whole informed consent form was read out loud during the recording and participants consequently accepted to be interviewed.

Overall, the good outcome of the data collection in Brazil strongly depended on the contacts the interviewer established prior to the fieldwork. The necessity of having trusted relationships with local members of the communities proved to be essential in consideration of the geographical and social structure of the country

2.2.3 Quebec and Belgium
In Quebec, the socio-historical issues highlighted above (younger emigration, bilingualism) turned out to be significant, as they substantially altered the speakers' profiles. The main differences, with respect to the speakers we found in Brazil and Argentina, were indeed the age of the speakers (younger, with physical and cognitive correlates – cf. 3.3.1) and the knowledge of other languages: most of our speakers had full knowledge of (regional) Italian and English, in some cases at the expense of French (the target contact variety for the area). Moreover, the Italian immigrants managed to achieve a reasonable level of economic well-being, which translates, among other things, into increased contacts with Italy. Therefore, most of our informants had been to Italy several times since they had left the country, and some of them regularly spent their holidays there, even for several months each year. Even those who, instead, did not have a direct and prolonged contact with their hometowns had been in regular contact with their relatives in Italy via long-distance communication devices over the years.

Another issue that we faced in Quebec were the different sociolinguistic settings of the two main areas of investigation, Montreal and Quebec City. The latter has a small Italian community and its members mostly switched to French on a daily basis because of marriage

with their local partners: on the whole, this had a negative impact on the transmission of their languages to the subsequent generation(s), as attested by the fact that we could only retrace one HS of an Italo-Romance variety in the whole city (Venetan), against 9 G1 speakers of different Italo-Romance varieties, mostly Piedmontese, Venetan, and Friulian. Montreal, on the other hand, was the destination of a massive emigration flow from Italy: the traditional Italian area, *la petite Italie* ('little Italy'), is nowadays a sheer memory of the Italian emigration. While it is home to the Church of the Madonna della Difesa, built by the Molisano community in the late 1910s, and it still shelters some Italian shops and cafés, over the years the bulk of the Italian community has moved away from it, to the outskirts of the city (e.g. Saint-Léonard and Rivière-des-Prairies, in the northern part of the island) as well as to its neighboring municiplities, especially Laval. However, this setting, too, proved to be detrimental to the preservation of the Italian dialects: the presence of emigrants from linguistically different areas of Italy and the necessity of mutual help and support resulted in the development of an Italian koine, which, besides English and French lexical borrowings, adds many regional structural and lexical features to a (broadly speaking) Italian structure. This variety ('Italianese', e.g. Villata 2010) is the most readily available one to our informants for daily communications within the community and, over time, has overshadowed the original dialectal richness of the city: as a result, once again, we only managed to interview one HS of one of the target Italo-Romance varieties (Sicilian), whereas 22 G1 speakers of various target languages participated in our study.

A further difference between the two areas of interest in Quebec is linked to the contact varieties, as mentioned above. While French is the contact variety for the Italians who emigrated to Quebec City, English is the most widely spoken local language for the Italian community in Montreal. This latter outcome, despite the efforts to make all new residents and their descendants converge on French (above all: via education policies), is due to the prestige of English in the wider North American context. Due to these differences, our informants in Quebec were very far away from our ideal speaker profile: instead of illiterate, dialectal speakers with a working knowledge of the contact variety, we found quite literate informants, with a good knowledge of some variety of Italian (a koine one, with specific regionalisms, if not a variety closer to the standard), a passive to good knowledge of French and a good general knowledge of English, the most widely used language outside of (but sometimes: inside, as well) the community, especially in Montreal.

In the attempt to find a speaker profile that matched more closely the ideal speaker for this research, we targeted Belgium. The Italian emigration to Belgium reached its peak right after World War II, with the bilateral agreements between the two nations. First signed in 1946, but effective until 1956, these agreements regulated the transfer of Italian workers to the Belgian mines. Despite the well-known presence of Italians in its territory and its greater geographical accessibility, Belgium had been originally excluded from the investigation areas because of its (relative) closeness to Italy, that could have led to many contacts between the emigrants and their hometowns. However, after we assessed that the Quebec Italian population did not match the required profile sufficiently, and given the accessibility of the area, we decided to run a small-scale fieldwork in the French-speaking part of Belgium.

This pre-test on the general feasibility of a more in-depth study of the Italian emigrant community in Belgium, taking into account our original socio-historical and sociolinguistic conditions, proved extremely fruitful. We decided to target the southern mine region around Charleroi and La Louvière, besides Bruxelles; in total, we interviewed eight informants: six G1 and two HSs. Our original worry of a population more connected to the homeland turned out to be unfounded: as we were told, the mines working contracts ensured that the miners did not go back to Italy for the whole contract duration, reducing the links considerably. After the

contracts expired, some workers went back to Italy for good; those who instead stayed in Belgium, continued their lives away from their home country, as in the previous period.

Moreover, the general speaker profile matched our ideal speaker better. Due to the disruptions to the school system brought about by the war, the emigrants, in most cases rather young men, were indeed quite illiterate, which was no doubt beneficial to the maintenance of the dialects. Furthermore, the well-organized migration flow brought people from the same area of origin in Italy to live and work together: this made sure that the dialects, as the most accessible varieties known to the miners, were preserved in daily life and even passed on to the following generation(s). A witness to this state of affairs is the presence, still today, of little municipalities in the south of Belgium where Italian dialects are still widely spoken among the members of the Italian community: an example is Morlanwelz, that was the destination of a considerable emigration wave from Villarosa, in Sicily (Enna province). These local and linguistic clusters, together with the limited contact with Italy over many years, resulted in a very low knowledge of Italian, too. Furthermore, and again in contrast to what we found in Quebec, French (less so its local Walloon dialect) is the major language spoken by these emigrants, besides their original dialects. No other contact language is instead attested, making the sociolinguistic context of Belgium overall more suitable for our study. The only disadvantage of the research in Belgium is the very limited number of varieties available: the majority of the miners came originally from southern Italy, and mostly from Sicily and Campania. Up to this point we have found few to no speakers of northern varieties. Unfortunately, it has been so far impossible to continue this small-scale fieldwork with a more extensive data collection.

2.2.4 US
In New York, the search for speakers whose community experienced displacement within the urban area was by no means trivial. The most fruitful way to find our speakers was to by-pass the official routes, such as the (unresponsive) Italian Cultural Institutes and Embassy, and to resort to:
- clubs, societies and cultural associations of local communities;
- shops run by Italians (pizzerias, restaurants, tailors, barber shops, etc.);
- individual contacts (internal or external to these communities) who acted as mediators between the researcher and the Italian communities, e.g. Endangered Language Alliance.

A total of 58 speakers (G1: 32; HSs: 26) from different heritage communities were interviewed in Manhattan, Brooklyn, and Queens, besides one family in Jersey City, NJ. The people we interviewed were speakers of Central and Western Friulian, Nònes (a Ladin-Lombard/Venetan transitional variety from Trentino), Eastern Abruzzese, Neapolitan, and Sicilian. Moreover, we interviewed speakers of varieties that were not considered for other fieldworks: Eastern Campanian, Cilentano (a southern Campanian variety), Apulo-Barese (from Apulia), and Ciociaro (an upper-southern variety from southern Lazio).

G1 speakers migrated between 1940 and 1980 from the different areas of Italy, where the local varieties are spoken alongside their respective (regionally marked) Italian. Most of the G1 speakers who arrived right after WWII settled down where the historical communities of Italians had settled during previous migration waves. These areas include (but are not limited to) Manhattan (Little Italy, Little Friuli, Murray Hill), Brooklyn (Bensonhurst, Williamsburg), Queens (Astoria). However, due to the increasing cost of living in the city, most of these communities were forced to move away from these neighborhoods and relocate to more peripheral areas.[8] Such displacements led to the partial or total dissolution of once-compact

---

[8] Mainly Queens, Brooklyn, and Staten Island, or outside NYC.

linguistic communities, but a language shift to English only also took place due to the generational change of 'community leaders' in clubs and associations. In fact, finding US-born HSs who are (fully) proficient in their local language proved rather challenging, as the large majority (especially from southern Italy) was forced to switch to English only by their families for integration purposes, therefore completely abandoning their heritage languages ('HLs') and retaining, in the best cases, only a passive knowledge of them. The profiles of the HSs we interviewed included: speakers with different degrees of active competence of their own local variety (depending on their age, the type and length of exposure to that variety, and cohesion of their community), speakers (usually highly educated) with active competence of their own local variety and Italian (learnt from (educated) family members, or during secondary education), and speakers with passive knowledge of their own local variety, which they define as 'the archaic dialect' of their families, and active competence in a regionally influenced variety of (Americanized) Italian, which they refer to as 'the modern dialect', or *Brooklynese/Queenese* Italian (next to the local Brooklynese/Queenese English).

Unsurprisingly, the most proficient speakers are the elders, i.e. above 70-75 y.o. The linguistic repertoire of these speakers is extensive, as they learnt the HL from close-knit communities of Italian-born parents, grandparents and/or other relatives, who emigrated within the first half of the 20th century to the specific Italian areas/neighbourhoods of New York City. Many of these speakers also kept tight connections with their families in Italy, allowing them to be exposed to the home dialect and/or spoken Italian. However, whenever they visit their family's birthplaces in Italy and interact with locals there, they are told they speak an 'archaic' variety of the language (cf. Aalberse *et al.* 2019: 114). Not only is this due to the fact that G1 imported a conservative variant, but also that these speakers were *not* exposed – as much as later generations in the US have been – to the growing linguistic pressure of Italian during the past 70 years.

A common point between Italian-born G1 speakers, who migrated to the US in the first years of their lives, and US-born HSs is that they all grew up as sequential bilinguals. They first acquired their own Italo-Romance HL (either the local dialect and/or the supraregional Italian koiné), and later English as 'Child L2', i.e. bilinguals who acquired a second language between 4 y.o. and puberty. (cf. Aalberse *et al.* 2019: 117).

For speakers younger than 70 years of age, the level of proficiency in the HL diminishes rather drastically. This is likely due to a less constant exposure to the relevant , or a drop in usage on a daily basis. Moreover, apart from the decrease in input, it is also likely that the quality of that input changed as an effect of attrition and cross-linguistic influence from English, resulting from the long-term decreased activation of the HL (cf. Pascual y Cabo 2013). Indeed, these speakers grew up learning a supraregional variant of Italian, originally brought overseas by their families after WWII (which became, in turn, influenced by English), while the HL was only used with grandparents, older relatives and/or elders in their neighbourhood. From the intermediate to the new generations, the dialectal lexis, phonology and syntax of the relevant baseline language seem to have blended to different extents with a spoken variant of Italian, as well as English. In fact, speakers below 40 years old (mainly from the south) appear to show an 'established confusion' about the languages they speak. They self-report that they are not able to speak 'proper' Italian and can only speak the dialect, but they actually speak the Italo-American (southern-based) koine. The communities of HSs which seemed to have better preserved their local dialects were Sicilians and Nonesi.

As a (partial) result of the issues highlighted above for each of the contact areas under investigation, the sample of speakers we could find is more varied than we were hoping for, leading sometimes to comparability issues that will be addressed in more detail in 3.3.

2.3 Tips and warnings

Here are some things to consider if you wish to set up fieldwork outside of Europe, and in America in particular:

- **Before you go, set up a risk assessment plan with your university**.
  The plan should contain information about the potential risks and an extraction strategy in case of danger; furthermore, it should contain a safety protocol that you may fix with your PI[9].

- **Make sure you stick to the safety rules agreed upon with your university/PI.**
  As soon as you arrive in a new place, contact or identify the relevant consulate in case you needed help, and make your presence known to them.
  Establish a daily routine with your referent, like sending an email or a message at a given hour every day to confirm that everything is okay and you don't need help.

- **Make sure emails to potential referents are short, clear and personalized.**
  After gathering and reviewing any available source for contacts, including outdated webpages which may bring to mind further connections, it is key to approach the possible (source of) informants with an email/message that is at the same time catchy and trustworthy, especially in the subject line. The text content, clear and concise, should include statements about (i) the purely scientific purpose of the research survey, (ii) the **non-profit** nature of the enterprise, and (iii) the compliance with the data protection protocol (i.e. anonymity), although public institutions are usually willing to be acknowledged in the research outputs. Whilst the message template should address a varied range of recipients, these may not feel engaged by a plain, somewhat detached, invitation. It is worth adjusting the text in a way that is informed of, or in fact sensitive to, the contact's role within the targeted local community, highlighting the shared benefits of the cooperation (academic for universities, socio-cultural for associations, motivating for individuals, etc.).

- **Allow plenty of time to network with the local community** (and to build a relationship of mutual trust). **A pre-fieldwork could prove extremely useful for this purpose, too.**
  Be it a secluded community in the Brazilian mountains or a dynamic community in Brooklyn, it is easier to access the speakers from within the local community. In particular, it is advisable to invest time participating to the regional associations' life: parties are not only entertaining, but also an excellent means to meet people in a more relaxed environment than the interview one, and to exchange contacts.

- **Involve the local communities (even more): try to instruct local people to carry out the interviews** even when the researcher is not present is a very good idea, if time constraints allow it. Ideally, this should happen for every data collection, however it is not always possible to make it happen, especially if the time to be spent in one location is too short.

- **Consider technical problems in isolated/remote areas.**
  In the most remote areas touched by our fieldwork, especially in Brazil, it has been the case that basic IT requirements were not met: for instance, internet connection was sometimes unavailable, but even more importantly power sockets were missing. This is a problem when performing a computer-based questionnaire. One suggestion might be to structure the questionnaire so that it can also be carried out in the absence of a

---

[9] For a taxonomy of possible risks, a useful tool is this Advisory note by the *International Science Council*: https://council.science/publications/advisory-note-responsibilities-for-preventing-avoiding-and-mitigating-harm-to-researchers-undertaking-fieldwork-in-risky-settings/.

laptop. In that case, it is a good idea to print out the questionnaire, so as to be able to ask the informants at least a part of the items when there is not a way to charge the laptop. Moreover, in some particularly isolated areas, it is advisable for the interviewer to take some water and food, in case they are stranded somewhere, and no transportation is immediately available.

- **Consider problems working with communities in big, busy cities, too.**
  Some informants will genuinely not have enough time (between half an hour and one hour) to spend on a questionnaire, as "time is gold" in their busy schedules, despite their interest and will to participate in the study. In New York, this was true for most of the informants, regardless of their age; in Quebec, this was also the case, although to a lesser extent. Moreover, consider that in large metropolitan areas such as that of New York and Montreal, communities are scattered all over the city, and travelling with public transportation is extremely time-consuming (for instance, between 1 and 3 hours without leaving the metropolitan area of NYC); likewise, taxis might take as long as public transportation, and are very expensive; therefore, allow plenty of time to reach your destination in advance.

- **Remember that especially in the case of illiterate speakers, it is culturally inappropriate to ask to sign a document they cannot read.**
  In this case, still complying with the legal requirements, it is possible to read out aloud the informed consent form and make sure the speakers give at least their oral consent to the collection and processing of their personal data.

- **Consider the mistrust of the informants.**
  Both in bigger cities and in more isolated villages there may be a general feeling of unsafety; this can especially be true if the fieldworker is a young male. To overcome the mistrust and make the informant feel at ease even in the presence of a stranger, and whenever possible, it is beneficial to go to interviews with another community member. This is also beneficial for the fieldworker.

- **Be prepared to interview people in unconventional places.**
  As much as we tried to make arrangements well beforehand to carry out our interviews in a quiet environment, this was not always an option (see also the feeling of mistrust introduced above). A good solution, when private spaces are not available, would be to carry out the interviews in the associations' offices. However, also this option is sometimes not available. It can then happen that the interview has to be carried out in parks, cafés, restaurants, shops, offices, waiting halls, *etc.*, where a quiet environment is not an option. It is advisable to keep this in mind while designing the questionnaire, as this can affect its feasibility (e.g. audio stimuli are more problematic in such contexts, and the use of a laptop might turn out to be non ideal).

3. Syntactic tests: the inquiry

3.1 Where we started
Our inquiry consisted of two parts, the first one aiming to assess the proficiency of the speakers and th second one testing the syntactic phenomena under analysis. We also gathered socio-linguistic information about the informants' language history and use.

3.1.1 Assessing proficiency
We decided to test proficiency as a first part of our questionnaire because we wanted to make sure that our informants were really able to speak the target dialect, rather than an italianized version of it (for legitimate concerns on the issue, cf. Section 2.1). Ideally, we would have only

met proficient speakers during our fieldwork, thanks to the pre-selection made possible by the preceding crowdsourcing phase (see D'Alessandro et al., in prep.). However, we did not get the desired results from this type of data collection: therefore, we interviewed all the speakers who made themselves available basing ourselves on their own linguistic self-evaluation, and were in need of a more reliable means to assess their proficiency in the dialect. We ultimately used this part of the interview as a pre-selection method to skim the questionnaires. Proficiency was tested in two ways: by means of spontaneous speech (first) and by performing a specific lexical decision task (afterwards). While the actual set-up of the procifiency assessment will be discussed more in detail in the next subsection, here we will introduce the considerations from which we moved from when we decided to take proficiency into consideration.

As it has been widely demonstrated in L2 acquisition and heritage linguistics research, spontaneous speech can be a reliable metric for language proficiency (cf. Polinsky 2018: 110 ff., and references therein). Different measures, such as the rate of speech (relative to a chosen unit: be it words, morphemes, *etc.* per minute) or the length of utterances, can be employed to comparatively assess the proficiency of a speaker in the different languages in their repertoire. By letting our informants speak, we were impressionistically able to understand even before the questionnaire whether they were able to speak the Italo-Romance variety they claimed to master or not: it was sometimes the case that we were only spoken to in an italianized dialect, or in regional Italian, by speakers who claimed that they spoke the dialect, and conversely speakers who self-assessed their dialect knowledge as very low could sometimes speak the dialect consistently and fluently. Moreover, the aforementioned research has assessed that there is a significant correlation between speech rate and grammatical proficiency tested by both production and comprehension tasks. Therefore, our raw data can be used for more accurate measures.

Besides this, spontaneous speech has been chosen as the introductory task for its multiple additional advantages: firstly, we were expecting some of our informants not to speak their recessive language on a daily basis, and sometimes in fact not to speak it anymore. Therefore, we thought that a good way to make them feel at ease could be to let them talk freely: we noticed that, while at the beginning some of the less proficient speakers had some issues with speaking the language, they became more and more fluent while talking with us. To try and trigger the use of the dialect, we mostly asked them questions about their childhood and their arrival to their new country: we hoped that, by asking them to talk about a time in which they spoke the dialect on a regular basis, they could be further encouraged to reproduce it. Secondly, spontaneous speech allowed us to gather socio-linguistic information (year of emigration, level of education, current and past dialect usage, *etc.*) that we needed to control for so as to keep our sample of informants the most homogeneous possible, but without boring them with very direct and structured questions. Thirdly, the spontaneous speech data were used to complement the information gathered from the questionnaire, and to check whether spontaneous production might reveal different linguistic patterns than the judgment data gathered through the questionnaire.

Another measure of proficiency is provided by lexical decision tasks, i.e. tasks designed to assess the lexical proficiency of speakers. These tests can be designed both as translation of lists of words and as picture-naming tasks. Here, again, it has been shown that lexical knowledge robustly correlates with grammatical knowledge (cf. Polinsky 2018: 107 ff., and references therein): for instance, it is well-known that, in normally developing children, the lexical and grammatical development progress in a directly proportional fashion. A similar correlation between lexical knowledge and grammar (in production tasks) has been detected by Polinsky in heritage Russian speakers and later replicated in subsequent studies on other HLs.

O'Grady et al. (2009) discovered a comparable correlation by developing and executing their HALA (Hawai'i Assessment of Language Access) test on a group of Heritage Korean speakers. The HALA test is a publicly available picture-naming set of tasks aimed at measuring the accessibility times of items and structures in the different languages in the repertoire of multilingual speakers, so as to assess the relative dominance hierarchy. The overall idea is that, in all language domains, proficiency correlates with frequency of use, which has direct consequences on latency times. Given these premises, we chose to have an additional task based on the HALA materials, as will be explained in more details in 3.2.1.

3.1.2 Towards the syntactic questionnaires

The second part of our inquiry consisted of a questionnaire testing the different syntactic phenomena in the contact varieties under analysis (for a definition of the phenomena and of the varieties, see Section 1). Our population is extremely varied in nature: G1 of different ages and levels of education, affected by attrition to different extents, and ~~HL~~ HSs~~speakerss~~; nonetheless, we wanted to develop one and the same questionnaire, so as to make our results directly comparable. Therefore, we had to take into account the specific challenges posed by each different group of speakers while designing the questionnaire. According to the literature, the most difficult group to test is the HSs one (for extensive remarks, cf. Polinsky 2018: chapter 3).

Considering that we were going to test mostly elderly speakers, we decided to restrict our choice to offline tasks, so as not to tax the participants' processing resources too much and to avoid the time pressure. Moreover, some online experiments request equipment that would have been hardly conciliable with the setting of our fieldwork (cf. the discussion in 2, and mostly the remarks on the unconventional interview places made in 2.3).

We then had to exclude grammaticality judgements and translations. Polar grammaticality judgements (Yes/No type) are not granular enough, while scalar ones (in which each experimental item is rated on an *n*-point scale, e.g. the Likert scales) detect finer differences among the stimuli, but are ultimately questionable under some respects, too (for a wider discussion, cf. Stadthagen-González et al. 2018 and references therein). The major issue is related to the very concept of scale: it is too trivial an assumption that an *n*-point scale is sensitive enough to faithfully represent the acceptability continuum, both in its extension and in its actual match with the speaker's own continuum representation). Moreover, and assuming that informants are consistent in applying one and the same scale all through the task, scalar grammaticality judgements are clearly demanding on the mnemonic side as well. Finally, grammaticality judgements prove particularly difficult for HSs, (Polinsky 2018: ch. 3, and references), which would have made it difficult to design one and the same questionnaire for all our participants.

Translations, instead, have been commonly chosen to carry out research on Italo-Romance varieties on a large scale (cf. for instance the traditional atlases that document Italo-Romance: AIS, ALI; and in more recent times: ASIt). However, one of our main concerns was to try to avoid the interference of any other language while performing the tasks, especially considering that, due to the frequency of use of the Italo-Romance varieties, we were expecting these to be the non-dominant languages of our informants. Moreover, our research targets different areas and different types of speakers, so different varieties would have had to be chosen to be the starting point of a translation task: the translations could have been performed from the local contact varieties (Spanish, Portuguese, French, and English) or from Italian, leaving the choice to the speakers (for instance, elderly speakers who migrated as adults might have preferred Italian, while HSs might have had a preference for the local contact variety). However, such a flexible format, with different variables to accommodate all our speakers'

imaginable needs, would have made the translations not fully comparable, as they would have primed the informants in different ways. We decided therefore to exclude translations.

Instead, we decided to structure our questionnaire as a two-alternative forced choice task. In this setup, the informants are asked to compare the acceptability of (a list of) pairs of stimuli by choosing, within each pair, the most acceptable item. Following a considerable amount of studies on the issue (for a discussion and specific references, cf. Stadthagen-González *et al.* 2018), we reckoned that such a format would have been beneficial for our research under many respects, among which the fact that it is less demanding to compare two items than to rate them on a predefined and consistent scale.

For each phenomenon, we identified a number of research questions on the basis of a preliminary review of the available literature. These questions were tested using (variations on) the two-alternative forced choice task described above, whenever possible and depending on the nature of the phenomenon (subject cliticss and DOM, and partly deixis, with the support of pictures); an alternative, additional semi-guided production task was added for deixis, besides a sentence completion task for auxiliary selection, about which we will not say much as this was an extra task. More information on how we paired tasks and phenomena and on how each task was designed and worked will be provided in 3.2; in the remainder of this section, instead, the specific conditions that were tested for each phenomenon will be explained in more detail.

3.1.2.1 Subject clitics. Subject clitics are found in most northern Italo-Romance and Rhaeto-Romance varieties. They differ from regular tonic subject pronouns in that they are syntactically deficient elements; they are inflectional heads, on a par with verbal agreement endings (Rizzi 1986; Brandi and Cordin 1989; Poletto 1993, 2000). However, Frasson (2020) shows that in Brazilian Venetan subject clitics display pronominal behavior. In our questionnaire we firstly tested the agreement-like or pronominal nature of subject clitics checking:

i. doubling contexts: an agreement marker obligatorily doubles an overt lexical subject, while a pronoun cannot occur when the subject is already expressed;

ii. coordinated structures: an agreement marker is obligatorily realized in both conjuncts in coordinated structures, while a pronoun is generally realized only in the first;

iii. negation: agreement markers normally follow the preverbal negation marker, while pronouns precede it;

iv. interpolation: agreement markers cannot be separated from the verb by non-clitic material, while pronouns can.

In addition, we checked for three more contexts that generally display some instability in younger speakers of Venetan in Italy (see Casalicchio and Frasson 2019). More precisely, these are:

v. interrogative sentences: subject clitics are normally enclitic in interrogative sentences, but there is a tendency to realize them proclitically, on a par with declarative sentences;

vi. impersonal constructions with meteorological verbs: the realisation of subject clitics varies substantially with respect to this context; often, both a sentence with a subject clitic and one without a subject clitic are accepted by speakers;

vii. default agreement constructions: as in the case of vi, realisation varies substantially in this context; sentences with post-verbal subjects and restrictive relative clauses normally require a default third person singular agreement on the verb (and no subject clitic); however, speakers often accept also full agreement (with a subject clitic) in these contexts.

Despite not being strictly related to the agreement-like or pronominal nature of subject clitics, the contexts in v-vii were added in order to test the stability of heritage varieties with respect to varieties spoken in Italy.

3.1.2.2. Deixis. In the case of deixis, the main focus of our inquiry was on the number of deictic contrasts encoded in demonstrative systems. Demonstrative forms and spatial adverbs anchor an object or an area in the external world to (one of) the discourse participants, by defining them in terms of the distance from the speaker and/or the hearer (e.g. *this* means that something is close the speaker, *there* means that an area is far away from the speaker).

Depending on how many discourse participants are available as possible anchoring points, we get different demonstrative systems: if the only relevant reference point is the speaker, then we typically have a system that encodes a two-way contrast (an object or an area close to the speaker as opposed to an object or an area far from the speaker; e.g. Italian *questo* 'this' and *qui* 'here' as opposed to *quello* 'that' and *là* 'there'). Since such systems have two forms, they can be referred to as binary systems. However, it is also possible for the first term of a binary system to jointly refer to an object or an area close to both discourse participants, without any further specification as to who is closer to the referent, and conversely for the second term of a binary system to refer to an object or an area that is far from both discourse participants at the same time: this is the case, for instace, of Catalan *aquest* 'this (close to the speaker and/or to the hearer)', *aquell* 'that (far away from the speaker and the hearer)'. If, instead, the hearer is also relevant in the spatial relations, the resulting system will encode a three-way contrast (an object or an area close to the speaker; an object or an area close to the hearer, or an object or an area far from both the speaker and the hearer). An example of such as system is Portuguese, which differentiates between *este* 'this' and *aqui* 'here' (close to the speaker)', *ese* 'that' and *ai* 'there' (close to the hearer)', and *aquele* 'that' and *ala* 'there' (far from both). These systems display three contrastive forms and can therefore be defined as ternary systems. In the Romance domain there are also systems that do not encode any deictic contrast, i.e. that only display one form that can be used in different deictic contexts without yielding any interpretation difference: this is the case for the adnominal and pronominal demonstratives of French (*ce* and *celui*, respectively, in their masculine singular versions).

The literature on Romance varieties on deixis highlights an extensive level of microvariation (see, for the most extensive overviews, Ledgeway 2015 and Ledgeway & Smith 2016): Romance varieties display all four systems and the variation level is pervasive, especially in the southern Italo-Romance domain. Therefore, we chose to investigate which deictic contrasts are encoded in Italo-Romance varieties in microcontact, by eliciting material related to the three possible deictic domains (close to the speaker, close to the hearer, far from the speech act participants), to test how these systems behave in contact and, ultimately, to better understand how these forms are encoded in the grammar.

3.2.1.3. DOM. DOM, acronym of Differential Object Marking (Moravcsik 1978, Bossong, 1985, 1991), also known in the Romance literature as prepositional accusative (Diez 1874; Meyer-Lubke 1890, 1895) is the phenomenon whereby some Direct Objects ('DOs') are marked differently than others, depending on some semantic and pragmatic features of the object. The phenomenon has different distribution patterns in Romance: some languages only display DOM with pronouns (e.g. some Eastern Abruzzese varieties, as in Manzini & Savoia 2005), other varieties only display DOM with a subset of pronouns (e.g. Ariellese, as in D'Alessandro 2017). In other cases, DOM is only possible in clitic doubling contexts (e.g. in Piedmontese, Manzini & Savoia 2005), whereas it is linked to specificity and definiteness in Southern Italo-Romance varieties (see Andriani, in press, for Barese; Ledgeway, 2009 for Neapolitan and Ledgeway, Schifano and Silvestri, 2019 for Calabrian; Guardiano 2000, 2010

for some Sicilian varieties) as well as in Peninsular Spanish (Leonetti 2004). Conversely, in Argentinian Spanish DOM is strictly linked to Case (Saab 2018) and in Standard Italian it mostly marks Object Experiencers (see Belletti 2018 for a recent overview and discussion).

In addition to this, we should also consider that the preposition marking DOs in these languages is the same one that introduces Indirect Objects ('IOs'), namely *a* (notice the contrast between Spanish DO *Veo **a** Juan* 'I see Juan' and the IO *Le doy el libro **a** Juan* 'I gave the book to Juan'). These differences gave way to a lively discussion in the literature on what really triggers DOM, and whether DOM objects are true accusatives or datives.

The starting point of our investigation was that the data we know from the literature suggest that what we label as DOM might be referring to a range of different phenomena that only happen to share the same superficial outcome. We wanted to know whether this is actually the case and if so, whether these differences are only the product of diachronic evolution, or if they really develop in the languages starting from different points. Furthermore, in the case of contact with Argentinian Spanish, we wanted to investigate if a possible change in the distribution of DOM in Italo-Romance varieties reflects the one of the contact variety.

3.2 What we did

3.2.1 Assessing proficiency

The production tasks worked well with most informants: we asked them to tell us about how they arrived in the Americas and what they found there (if they were G1), or in general about their childhood, parents and links to Italy. Our plan was to collect at least 5-10 minutes, but some of them were so happy that they talked for half an hour or even an hour. This is also due to the fact that they knew that their recordings would be published (strictly anonymized) on the project's atlas, so they were happy that their story would reach a larger audience. Still, in some cases the informants felt awkward to speak the dialect when the fieldworker was not a speaker of the same language. In these cases, they often mixed it with Italian or with the language of their new country.

As anticipated in 3.1.1, we also performed an additional test to assess lexical proficiency on the basis of material elaborated by the HALA research group. In the HALA test, three sets of pictures (body parts, natural elements, and general pictures to create short sentences) are showed, which participants have to name in the target language as quickly as possible. Not only does this give an indication of vocabulary size, but also of speed of lexical access, both of which are indicators of language proficiency. The speed is measured by calculating the time lapse between the moment in which the picture appears on the screen (highlighted by an audio signal) and the moment in which the participant names that picture. However, for the test to be carried out successfully, it is necessary to compare the speed of lexical retrieval across different languages in the repertoire of the participant, to comparatively assess whether the specific times for a given language are linked to a genuine delay in retrieval (and hence to lower proficiency) or whether they are in line with the access times in other varieties and long times are simply due to external factors. Since the test has to be performed in different languages, we decided to only use a short part of the original HALA test: we asked six items in the Italo-Romance variety before the questionnaire and then asked to repeat the test in the language they felt as their dominant one after the questionnaire.[10]

---

[10] Note that, despite our efforts to comply with the test requirements and with the non lab-based nature of our data collection, this set-up still does not match the HALA guidelines. Ideally, in fact, the test shoud be performed in one language a day, and after having started the conversation in that specific language, so that the informant is in the 'right' language mode. Clearly, such an option was not available in our case.

So far, the data have been left in their raw version, that is: we still have not calculated the different response times across varieties. A complicating factor was the experimental setting: as already mentioned, we had to carry out interviews in unconventional locations, which makes it difficult for us to detect the signal sound that we should take as the starting point in the calculation of the response times.

3.2.2 Designing and running the syntactic questionnaires
In the design of our questionnaires, we had to consider many constraints related to the status of the varieties under analysis, the specific differences among the syntactic phenomena considered in our study, and the type of population that we were targeting.

The first issue we faced in the design of our questionnaire was the fact that Italo-Romance languages are not standardized and, as such, most of them do not have an orthography and are mainly spoken. This made it virtually impossible for us to use written stimuli in our interviews: the non-familiarity with the written variant of a dialect would have created a barrier between the task(s) and the speakers. Moreover, we would have had to define, for each variety that has no standard writing, a consistent written representation, at the cost of making it not fully intelligible to our informants. We considered using the standard Italian writing system for all varieties, but were concerned that it could have led to more normative and less naturalistic judgments; a similar issue would have resulted from the usage of the writing system of the contact vareities.

Some of the varieties under investigation have a long written tradition and therefore a standardized spelling convention. Still, the written systems show microvariation, mirroring the actual linguistic microvariation found across the Italo-Romance domian. Once again, presenting the speakers with a slightly different spelling system for their variety could have resulted in slight unfamiliarity with the presented stimuli, in a way reminiscent of the case of the varieties that do not have a conventionalized spelling discussed above. Furthermore, the choice of one standard written variety over another might have triggered unwanted judgements on the spelling, besides or rather than on the stimuli. Another point that we had to consider is that most G1 speakers are elderly people who may have problems in visualizing and reading due to their age or who might be altogether illiterate.

Having ruled out the possibility of a written questionnaire, we were left with two possible options for an oral questionnaire: to lead the interview personally, or to use pre-recorded stimuli. Given that every interviewer had to test speakers of all varieties involved in our study, it would have been difficult if not impossible for the fieldworker to perform the interviews in all target varieties; an attempt to do it would have led to biased data. Therefore, we decided to have native speakers of each target Italo-Romance variety pre-record a set of stimuli in their variety and present our informants with those auditory stimuli. Nonetheless, when possible and whenever the interviewer and the informant spoke the same variety, that specific language was used throughout the whole interview.

Moreover, in the case of New York, the language of instruction for the interviews was adapted to the speakers' relative 'confidence' with the languages in question. G1 speakers who learnt English past their teenagehood preferred being talked to in Italian, while the remaining G1 speakers and all HSs preferred English. Whenever possible, the interviewer also used the relevant Italo-Romance variety to encourage the speaker not to switch to English or Italian. However, this strategy was not always successful, as Haller (1987: 394) also reports: "even though the interviews were conducted by Italian-Americans accepted in the community [... w]hen asked to switch to dialect, the informants generally continued to speak their high variety [(*dialectal*) *Italian*; L.A.] after uttering a few dialect words, even if the interviewer was somewhat fluent in the specific dialect".

Some issues related to the phenomena under analysis further oriented us in the choice of tests. In the case of subject clitics, a two-alternative forced choice task was the best way to identify the agreement-like or pronominal behavior. Participants had to choose between two proposed sentences: one with a well-behaved agreement-like subject clitic and one without the clitic or with a clitic displaying an anomalous behavior. This is shown, for instance, by the context of coordination:

(1) Friulian
(a) Al   mangje e   al   bêf.
    he.SCL eats   and he. SCL drinks.
(b) Al   mangje e   bêf.
    he. SCL eats   and drinks.
 'He is eating and drinking.'

The sentence in (1a) shows that the subject clitic is repeated in both conjuncts in a coordinated structure; this is expected, as subject clitics are obligatory agreement markers realized every time a finite verb appears; the sentence in (1b) shows that the marker is realized only in the first conjunct, which is taken to be a pronominal behavior. The pronominal or agreement-like behavior were presented in a random order. Speakers heard the two stimuli one after the other and in random order, and had to choose which one they preferred.

The forced choice task proved successful for subject clitics in most cases: informants understood the task correctly. However, the spontaneous production task was a fundamental support to the questionnaire. Not only did it help (dis)confirming the results we obtained with the questionnaire, but it made it possible for us to notice further aspects of the distribution of subject clitics that would otherwise be left unnoticed. The most relevant example in this respect is the tendency to realize more overt pronominal subjects in heritage northern varieties in comparison to heritage southern varieties.

As for DOM, the forced choice task targeted the following range of direct objects, to determine whether they would trigger DOM:

1st person pronoun > 3rd person pronouns [+human] > kinship > [+human][+animate][-definite] > [-human][+animate][+definite] > [-human][+animate][-definite]

This order reflects Silverstein's (1976) animacy scale, since the general understanding of DOM in Italo-Romance varieties is that the higher the object is on the scale, the more likely it is to be marked.

These objects were tested both in situ and in fronted topic position (Rizzi 1997). Speakers of the southern and northern groups had two slightly different questionnaires. Informants of the southern varieties had thirteen sentences testing DOM plus fillers, for a total of twenty-four sentences. Speakers of northern varieties were given nine sentences testing DOM plus fillers for a total of twenty-three sentences. This was decided given that we were not expecting production of DOM on a wide range of arguments by speakers of northern varieties, since most of these are typically considered non-DOM in their homeland counterparts.

The informants were asked to choose between the sentence that included DOM and the one that did not: also these stimuli were presented in random order. The informants were asked to choose one of them by either saying 'a/b' or 'the first/the second one'. They could also just repeat the option they liked the most, and in this case we marked down any changes that have been produced (e.g. if the informants turned an indefinite object into a definite one when they repeated the sentence).

Although speakers needed guidance when taking the questionnaire (e.g. sometimes they had troubles understanding certain lexical items in the variety of the stimulus, and translation had to be provided), the test worked in most cases. In some cases, when informants deemed the first sentence correct, they confirmed it before listening to the second sentence. In these cases we had to ask informants to wait until they heard both sentences before deciding between the two options.

For deixis, we decided to avoid grammaticality judgements, sentence completion and elicited imitation, as demonstratives heavily rely on the context in which the conversation takes place. In fact, demonstratives are always grammatical, but they carry semantic differences that make them more or less suitable for a given context: different forms are used in different contexts, and this choice may depend on other indexical properties of the sentence as well. In grammaticality judgements, it is rather difficult to recreate such a context.

Although sentence completion and elicited imitation are typically not bound to any context, they raise other issues for investigating deixis. In both these task types, the target form can show a mismatch from the elicited one because of the switch in the deictic centre at the conversation turn. For instance, in the case of elicited imitation, the informant might switch the deictic centre when repeating the sentence, e.g. 'I am here' > 'You are there'. While both sentences are equally grammatical, they change in their interpretive content, which is however not tested (nor testable) in an elicited imitation task.

To circumvent these issues, we opted for a picture-sentence matching task and for a semi-spontaneous production task. For the former, we presented our informants with some pictures of dog owners and their dogs; one of the dog owners was marked as the speaker thanks to a balloon:
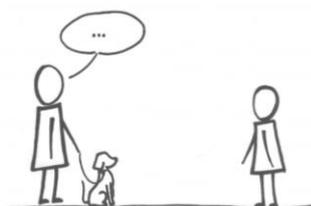
Figure 1: 'close to me'   Figure 2: 'close to you'   Figure 3: 'far from us'

Our informants had to identify themselves with the speaking character and refer to the dog present in the context of the picture (1, 2, or 3) by choosing one of either two or three (depending on the system in the target variety) recorded audio stimuli associated with each picture sentences. For instance, given Figure 1 with a dog owner, holding their dog, and another person (the hearer) on the other side of the picture, and given the dialectal audio stimuli for 'This (close to me) is my dog', 'That (close to you) is my dog' (if available in the target variety), and 'That (far from us) is my dog', the target item would have been 'This (close to me) is my dog', i.e. the proximal demonstrative *this*.[11] Also in this case, the stimuli were presented in random order.

We found that this setting was not flawless: most importantly, some of our informants found it particularly hard to identify themselves with the speaker in the picture; likewise, some participants had problems understanding that the speaker actually had an interlocutor inside the picture itself. Instead, some informants selected one of the audio stimuli on the basis of where the dog was in relation to them: given that the stimuli were presented on a laptop screen and that the screen was within their arm reach, they tended to point at or touch the dog and identify it as 'this', in any context, even the distal one. Moreover, considering potential sight issues, the

---

[11] This was the target sentence for the pronominal context. Other syntactic contexts tested were: the adnominal one (e.g. 'This dog is mine'), and the demonstrative-reinforcer one ('This here is my dog').

main characters on the picture were depicted rather big in size, which resulted in the picture itself being quite cramped and the distance between the characters to be overall too reduced: specifically, the 'close to you' space could easily be reduced to the 'close to me' one, as the speaker and the hearer were only a little space apart. The informants could sometimes motivate their answer by saying: "it's still close, if it were 'that' it would be something else". Such size considerations, together with the identification problem, led to an overall higher rate of proximal forms even for non-proximal contexts. However, when further asked with real-life situations, their responses would change substantially. One such ploy used to elicit the (actual) distal was the question 'Would you still use 'this' if the dog that you see was on the other side of the street?'. Still, no specific protocol for these cases was agreed upon before the fieldwork, so the data collection was, in this respect, not uniform, and its results not completely trustworthy.

Semi-guided production proved a better test for deixis: in this case, we used three pictures of cats in different colors: a black one, an orange one, and a white one. These pictures were placed either near the informant (the speaker), near the interviewer (the hearer), or far from both. Our informants were then asked where each cat was in the context, to which they had to reply with a demonstrative form or with a spatial adverb. We reckoned that the factor that made this task easier to perform for our speakers was the actual contrastivity within the context: they effectively needed to choose different demonstratives to let us understand to which cat they were referring. However, in this case, too, things were far from perfect: the most important issue that we encountered was how to elicit the demonstrative or spatial adverb, rather than a description of the image or of its location with respect to other objects in the room (e.g. 'the one on the chair', rather than 'that one'). In order to do so, we sometimes suggested the whole set of answers in the contact language, for the informants to understand what the target response was, avoiding the priming of the language-specific demonstrative system.

One last issue with which we were faced in the preparation of the deixis questionnaire was the clear difference in tasks. While for SCLs and DOM the task was comparable and we could use the sentences targeting SCLs as fillers for those targeting DOM and vice versa, thus keeping the questionnaire to a minimum length so as not to tire our elderly informants while ensuring for a better quality of our investigation, it was impossible to do the same for the deixis part of the questionnaire. For instance, it would have been ideal to show, besides the pitures in Figg. 1-3, some sets of filler pictures targeting other phenonema, also to avoid the repetitivity of the task. While designing the task, we reckoned that the addition of fillers would have been both an online confounding factor (the informants would have had to correctly interpret multiple scenes) and time-consuming, especially given our attempt to design a questionnaire that targeted all phenomena at once, while still being of a manageable size. However, upon testing, we realized that the absence of variation in the referent (always a 'dog', although in different positions in the picture and in different syntactic contexts: pronominal, adnominal, demonstrative-reinforcer construction) made the test extremely repetitive, which resulted in complaints on the informants' side, who thought that they were asked the same question over and over again. Variation in the referent would have perhaps been beneficial to the task.

3.3 General issues concerning experimental design and statistics
In an ideal world, all our participant groups would have had an equal number of participants, spoken the exact same local varieties, and all possible variables would have been perfectly controlled for. Moreover, all participants would take exactly the same task in exactly the same way. However, due to the scarcity and the heterogeneous composition of our target populations, as well as problems that arose during the fieldwork, this was not possible. While we must accept that no research is perfect, it is important to be aware of the possible consequences of these issues for the interpretation of the results.

Regarding the characteristics of the participants, what can be concluded from the description presented in section 2, is that our participants were not evenly distributed across the different varieties, host countries and generations. This has to be taken into consideration when analyzing the results, especially if one performs statistical analyses. For instance, there were eight speakers of Abruzzese in Argentina, only two in Canada and none in Brazil. Two speakers are too few to be able to perform any statistical analyses, so for this variety, we were only able to statistically model the linguistic behavior of the speakers in Argentina. In addition, of these eight speakers of Abruzzese in Argentina, two were ~~first-generation immigrants~~G1, and six were second-generation HSs. Again, given that two speakers are very little to form a separate subgroup, it was impossible to take 'generation' as a variable into account in the statistical analysis, and all 8 speakers were treated as pertaining to the same group, while in fact there was an important difference between some of the speakers, namely some of them were immigrants and others were born in the host country. Also, as mentioned above, there were differences between communities regarding literacy, education level, exposure to other languages, *etc*. While it is impossible to completely control for these variables in this type of study, it is important to keep their impact in mind during the analysis. For instance, we found certain differences between the use of SCLs by speakers of heritage Friulian in Argentina and Brazil. While a first interpretation of such a difference might be that there is an effect of the contact language (Spanish and Brazilian Portuguese differ in terms of their configuration of the pro-drop parameter), there were other differences between the communities. First, as mentioned, the communities in Brazil tend to be more isolated and the HLs therefore tend to be better preserved. Moreover, HSs in Argentina were mostly second-generation speakers while those in Brazil were almost exclusively third generation.

The design and the execution of our tasks was less than ideal from the perspective of experimental validity. The materials, i.e. the specific sentences for each of the phenomena, were selected with specific research questions in mind. In order to reduce the length of the questionnaire, in most cases, only one sentence (pair) per condition (sentence type) was used. For instance, the questionnaire for SCLs contained one sentence pair containing doubling, one containing coordination, one containing negation, etc. The disadvantage of this is the fact that one sentence may not be representative of the entire condition. There may be other elements in the sentence – i.e. the semantics of the verb, a specific word order – that may affect the acceptability of a subject clitic in a given sentence. Furthermore, a low number of items per condition complicates the statistical analysis. Another issue that should be taken into account is that, for some of the phenomena, all the sentences were presented together, without filler/distractor items. This may have made the participants aware of topic of investigation, which, in theory, may lead participants to resort to certain answer strategies (for instance: always picking the sentence with DOM).

Finally, as has been mentioned above, some of the interviewers had to improvise, either due to the fact that the informants did not understand the task, or because they did not have enough time to perform the complete questionnaire. This affected the uniformity of the study in various ways. For instance, not all participants answered an equal number of questions for each of the phenomena, reducing comparability across participants and/or groups. Another issue is the fact that some of the researchers carried out the experiment in the dialects, whereas other did so in the contact language or in standard Italian. It has been noted (Aalberse 2013) that the specific language spoken by the researcher may affect the respondents' linguistic behavior. The task type was also sometimes adapted on-the-go by the researcher. For instance, for those respondents who did not understand the forced choice task, it was sometimes (orally) adapted to a translation task. Similarly, in the guided production task used for deixis, some of the researchers, but not all, opted to present the participants with the full set of options for

demonstratives in the contact language, which may have led to a higher instances of target responses for those participants.

3.3.1 Interviewing the elderly

The main target speakers of this project are first-generation emigrants, who are quite elderly. The average age of G1 speakers was around 75. This brought about additional issues that we considered before fieldwork, but that in some cases had a larger impact on the results than expected. Advanced age brings about number of common issues, like partial or complete loss of hearing and sight, which we tried to take into account when designing the questionnaire, while still respecting the constraints imposed on us by the different phenomena.

An additional issue is the difficulty of retaining long sentences: therefore, we tried to keep the stimuli as short as possible. Furthermore, while this was true of many younger speakers too, many elderly speakers had clear difficulties with the very concept of choosing between two options: rather, they would approve of the very first stimulus out loud, regardless of its grammaticality and without listening to the second one. When it was not possible to give them more instructions that would help them to complete the task as originally planned, they were given sentences in the contact language, and they were asked to translate them into the dialect.

3.4 Tips and warnings

Here are some tips to design a good questionnaire for heritage speakers
- **When designing a questionnaire, keep statistical analysis in mind.**
  Define the dependent and independent variables of the research and select the tested conditions carefully; this also implies avoiding variability within those conditions to exclude influence from confounding factors. Even if you do not plan to do statistics with your data, you could change your mind at the end of the fieldwork, so it is better to collect data that can be used in case. Also remember that you will need at least 6 items per condition in order to be able to run significant statistical tests with the data elicited; likewise, you will need to add fillers.
- **Make sure the stimuli are culturally appropriate.**
  Verbs like *kissing* (a man or a woman) or *liking someone* may create uneasiness, especially among elderly speakers. Moreover, it is not a good idea to record sentences that imply any act of *killing*, as the speakers could be afraid that a recording containing such a statement could be taken to reflect the reality.
- **Make sure you have some spontaneous data.**
  It is always a good idea to compare the questionnaire responses to some spontaneous data. If that is not possible, e.g. because the speakers are not comfortable talking in their non-dominant language without any predefined topic, some (controlled) production tasks can help (and make the collected data more comparable). Keep in mind that spontaneous speech is very useful to assess proficiency, too.
- **Use the target variety when possible, to minimize the interference of other languages.**
- **If the questionnaire is too long and it wishes to test at least two phenomena, use the questions for the other phenomenon as fillers in your questionnaire.**
- **Make agreements between different interviewers on what to do in if the task does not go as expected.**
  If you are running several parallel fieldwork sessions, it is not always possible to exchange experiences with your colleagues and solve unexpected situations in a uniform way. Therefore, it is advisable to think about the main issues that may arise

beforehand (e.g.: a participant does not understand the task) and to define a protocol on how to proceed in those cases, in order to limit the degree of unwanted variation.

- **Carry out a pilot study when possible.**
  Before starting your fieldwork, it is a good idea to perform a pilot version of your questionnaire/experiment with speakers who are comparable in age and other sociolinguistic factors to those of your target population. This might highlight some issues that can be improved upon before the actual fieldwork.

- **Try to avoid priming.**
  While this is true for all speakers, elderly informants seem to be more prone to just repeat what they have heard last (or just to listen to one stimulus) or what the fieldworker suggested as example. Therefore, be very careful especially while explaining the task.

- **Pay extra attention to the design of your stimuli if you are planning to interview elderly people.**
  Elderly people may present some challenges that are linked to their age: hearing and sight issues, longer processing times and more expensive processing altogether, weaker short-term memory, lower attention span, *etc*. You should keep them in mind when designing your questionnaire, and specifically: use short questions, both for the short-term memory and to limit the time of exposure; make sure that your stimuli are fully accessible (the volume of audio stimuli has to be loud enough; if written stimuli are chosen, the font size has to be rather big). If feasible, it is a good idea to split a long questionnaire into two parts and test them at different moments.

- **Be ready to get more involved with the community, especially when testing elderly speakers.**
  As elderly speakers can be suspicious, especially when using modern technology such as recording material, make sure there is always a relative around, if possible, or another member of the community who can assist the person and reassure them that you are not doing anything inappropriate. Also, be ready to spend more time with your informants than you were planning to: some of them are lonely and really enjoy some company and they especially appreciate the opportunity to speak to younger people from their home country. Having a free production of one hour is great for your research, but may be problematic if you have scheduled another interview shortly after the current one.

Closing remarks

In this article, we have tried to highlight all the information we collected and all the things we learned when setting up and carrying out fieldwork of heritage Italo-Romance speakers in North and South America. While many of these tips and information can be found in general manuals or fieldwork reports, some are specific to the Italo-Romance community. Furthermore, we provided a description of the status of these varieties, many of which had not been documented since the '60s. In the case in which we did have some documentation of previous stages of the languages, we compared that to what we found, and shown that the situation has changed considerably. With the exception of the northern Italian/Talian-speaking community, most Italo-Romance heritage varieties in America are close to extinction. Given their decline also in Italy, it does not seem to us that they will be revived; the feeling we have is that the process that will bring them to full extinction within one generation cannot be stopped.

References
Aalberse, Suzanne, & Muysken, Pieter. (2013). "Perspectives on Heritage Languages". Position paper, Nijmegen: Radboud University.

Andriani, Luigi (to appear), Differential Object Marking in Barese: Old and new challenges. In Monica-Alexandrina Irimia & Alexandru Mardale (eds), *Differential Object Marking in Romance: Towards Microvariation*. Amsterdam: John Benjamins.

Bagna, Carla (2011). America Latina, in Massimo Vedovelli (ed) *Storia linguistica dell'emigrazione italiana nel mondo*. Roma: Carocci, 305-357.

Belletti, Adriana (2018). On a-marking of object topics in the Italian left periphery. In Petrosino, Roberto, Pietro Cerrone & H. van der Hulst (eds), *From sounds to structures*. Berlin, Boston: De Gruyter Mouton. doi: https://doi.org/10.1515/9781501506734-016

Bettoni, Camilla (1990). Italian Language Attrition in Sydney: The Role of Dialect. In: Michael Halliday, John Gibbons & Nicholas Howard (eds), *Learning, Keeping and Using Language*. Amsterdam/Philadelphia: John Benjamins, Volume 2, 75-89.

Bettoni, Camilla (1991). Language Shift and Morphological Attrition Among Second Generation Italo-Australians. *Rivista di linguistica* 3 (2), 369-387.

Brandi, Luciana & Patrizia Cordin (1989). Two Italian Dialects and the Null Subject Parameter. In O. A. Jaeggli and K. J. Safir (eds), *The Null Subject Parameter*. Dordrecht: Springer. 111-142.

Cavaioli, Frank J. (2008). Patterns of Italian Immigration to the United States. *The Catholic Social Science Review* 13, 213-229.

Cenni, Franco (2003). *Italianos no Brasil: "andiamo in Merica."* São Paulo: EdUSP.

D'Alessandro, Roberta (2017). When you have too many features: Auxiliaries, agreement and clitics in Italian varieties. *Glossa: A Journal of General Linguistics*, 2(1), 50. DOI: http://doi.org/10.5334/gjgl.102

D'Alessandro, Roberta, Francesco Ciconte, Luigi Andriani & Brechje van Osch (in prep), Citizen science and crowdsourcing for linguistic data. To appear in Gutierrez, E. & I. Ortega-Santos, *Language Contact,* Special issue of *Glossa.*

De Fina, Anna (2014). Language and identities in US communities of Italian origin. *Forum Italicum* 48(2), 253–267.

De Fina, Anna & Luciana Fellin (2010). Italian in the USA. In: Kim Potowski (ed.), *Language diversity in the USA*. New York: Cambridge University Press, 195-205.

Diez F. 1874 - Grammaire des langues romanes, II et III, Paris, Franck.

Ferrari, Andrea (2008). *Aspetti socio-culturali dell'emigrazione italiana in Argentina: il caso di Santa Fe*. Unpublished MA Thesis, University of Turin.

Haller, Hermann W. (1987). Italian Speech Varieties in the United States and the Italian-American Lingua Franca. *Italica* 64 (3), 393-409.

Haller, Hermann W. (1991). Atteggiamenti linguistici nelle comunità italo-americane. *Rivista di Linguistica* 3 (2), 389-405.

Haller, Hermann W. (1993). *Una lingua perduta e ritrovata: l'italiano degli italo-americani*. Florence/Rome: La Nuova Italia.

Haller, Hermann W. (1997a). Dialects Abroad. In: Martin Maiden & Mair Parry (eds), *The Dialects of Italy*. London: Routledge, 401-411.

Haller, Hermann W. (1997b) Italian. In: Joshua A. Fishman & Ofelia Garcìa (eds), *The Multilingual Apple. Languages in New York City*. Berlin/New York: Mouton De Gruyter, 119-142.

Haller, Hermann W. (2002) I dialetti italiani negli USA. In: Manlio Cortelazzo, Gianrenzo Clivio, Nicola De Blasi & Carla Marcato (eds), *Dialetti italiani: storia, struttura, usi* Turin: UTET, 1083-1086.

Ledgeway, Adam (2009). *Grammatica diacronica del napoletano* (Beihefte zur Zeitschrift für romanische Philologie Band 350). Tübingen: Max Niemeyer Verlag.

Ledgeway, Adam (2015). Varieties in Italy. In K. Jungbluth and F. Da Milano (eds), *Manual of Deixis in Romance Languages.* Berlin: De Gruyter, 146-162..

Ledgeway, Adam, Schifano, Norma and Silvestri, Giuseppina (2019). 'Differential object marking and the properties of D in the dialects of the extreme south of Italy', *Glossa: a journal of general linguistics* 4(1):51.1-25.

Ledgeway, Adam, J. C. Smith (2016). Deixis. In A. Ledgeway and M. Maiden (eds), *The Oxford Guide to the Romance Languages*, 879-896. Oxford: OUP.

Leivada, E., R. D'Alessandro, K. Grohmann (2019). Eliciting Big Data From Small, Young, or Non-standard Languages: 10 Experimental Challenges. *Front. Psychol.* (https://doi.org/10.3389/fpsyg.2019.00313).

Manzini, Maria Rita & L. M. Savoia (2005). *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.

Meyer-Lübke, Wilhelm (1895-1900), *Grammaire des langues romanes*, Paris, Welter (1890-1902), *Grammatik der romanischen Sprachen*, Leipzig, Fues

Moravcsik, Edith A. 1978. On the case marking of objects. In Joseph H Greenberg (ed.), *Universals of human language*, *vol. 4: Syntax*. Stanford: Stanford University Press, 249–289.

O'Grady *et al*. (2009). A Psycholinguistic Tool for the Assessment of Language Loss: The HALA Project. *Language Documentation & Conservation* 3: 100-112.

Poletto, Cecilia (1993). *La sintassi del soggetto nei dialetti italiani settentrionali*. Padova: Unipress.

Poletto, Cecilia (2000). *The higher functional field: evidence from Northern Italian Dialects*. Oxford University Press.

Polinsky, Maria (2018). *Heritage Languages and Their Speakers*. Cambridge: CUP.

Reinke, Kristin (2014). Language contact in a multilingual setting. The attractive force of Italo-romance dialects on Italian in Montreal. In: Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder & Achim Rabus (eds), *Congruence in Contact-induced Language Change Language*. Berlin/Boston: De Gruyter. 149-167.

Rizzi, L. 1986. On the Status of Subject Clitics in Romance. In O. Jaeggli and C. Silva-Corvalan (eds) *Studies in Romance Linguistics*. Dordrecht: Foris, 391-419.

Stadthagen-González, Hans, Luis López, M. Carmen Parafita Couto and C. Alejandro Párraga (2018). Using two-alternative forced choice tasks and Thurstone's law of comparative judgments for code-switching research. *Linguistic Approaches to Bilingualism*, 8: 67-97. DOI: https://doi.org/10.1075/lab.16030.sta

Toso, Fiorenzo (2011). Comunità dialettofone italiane in America Latina. Tra storia e attualità. Raffaella Bombi & Vincenzo Orioles (eds.), *Nuovi valori dell'italianità nel mondo. Tra identità e imprenditorialità.* Udine: Forum, 165-176.

Villata, Bruno (1990). *L'italiano a Contatto con il Francese e con l'Inglese.* Montréal: Montfort & Villeroy Éditeurs.

Villata, Bruno (1999). Il destino delle lingue minoritarie. In: Bruno Villata (ed.), *L'Italiano a Montréal nel Duemila. Atti del Colloquio Montréal 19 Marzo 1999.* Montréal: Lòsna & Tron, 51-58.

Villata, Bruno (2010). *L'italianese – L'italiano comune parlato a Montreal*. Montreal: Lòsna & Tron.