

Chapter 2

Documenting Italo-Romance minority languages in the Americas: Problems and tentative solutions

Luigi Andriani^a, Jan Casalicchio^b, Francesco Cicone^c,
Roberta D'Alessandro^a, Alberto Frasson^a, Brechje van Osch^d,
Luana Sorgini^a & Silvia Terenghi^a

^aUtrecht University, UiL-OTS ^bUniversity of Palermo ^cUniversity of Insubria
^dUniversity of Tromsø

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, which aims to understand language change in contact by examining the language of Italian communities in the Americas.

1 Introduction

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, which aims to understand language change in contact by examining the language of Italian communities in the Americas (<https://microcontact.sites.uu.nl>).

The speakers involved in our study are first-generation Italians (so-called *émigrés*; henceforth: “G1”), most of whom 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 close to ideal for a study on language contact, because most of them were tendentially monolingual

speakers of an Italo-Romance dialect when they arrived in the Americas. Italian was not widely spoken in Italy until the 1960s, and hence they were mostly monolingual speakers of these varieties when they left Italy. When they arrived in the Americas, they entered into sudden intensive contact with other Romance languages: we are focusing here on French (in Quebec and in later fieldwork in Belgium¹), Spanish (in Argentina), and 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 found in the Americas and Belgium, first and foremost because the contact has been more intense, and because the communities speaking these varieties are generally larger in Italy. Finally, we also investigate Italo-Romance speakers in contact with English in the United States as control group.

The general aim of the project is to draw a predictive analysis of language change in contact by looking at multiple microcontact situations. Contact is investigated here not on a one language-to-one language basis but on a many-to-many basis: in this way, each phenomenon can be checked against multiple, minimally-varying, equivalent phenomena in the contact languages. By the end of the project, we hope to have identified the structural triggers for language change. Furthermore, we also wish to compare change in contact with diachronic change, to ascertain whether they follow similar paths, as is often claimed in the literature. 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, 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. The preliminary results are being published in a number of papers (Andriani et al. 2020, Casalicchio & Frasson 2019, Sorgini 2020, Terenghi 2020, D’Alessandro 2021, Frasson (In press), Frasson et al. (In press), and many other papers in preparation).

The varieties that were originally selected for investigation are Piedmontese, Venetan, Tuscan (Florentine and Sieneese), Coastal/Eastern Abruzzese, Neapolitan, Salentino, and Sicilian. They are displayed in the map in Figure 1 (in regular font).²

¹See below and Section 3 for details on why we additionally targeted Italian emigrants in the French-speaking part of Belgium.

²The map is retrieved from https://en.wikipedia.org/wiki/Languages_of_Italy.

2 Documenting Italo-Romance minority languages in the Americas



Figure 1: Languages of Italy and selected varieties
https://commons.wikimedia.org/wiki/File:Linguistic_map_of_Italy_-_Legend.svg CC-BY-SA-4.0 Mikima

These varieties were chosen for several reasons: they maximally instantiate the variation recorded for our target phenomena across Italo-Romance and are the most 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 documentation 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 for the last 70 years, not many people could speak Italian at the beginning of the 20th century.

The languages selected did not all prove optimal. In particular, we did not manage to find Tuscan, Salentino or Neapolitan speakers. Instead, a very large community of Calabrian and Friulian speakers was identified during fieldwork. Figure 1, in italicized font, shows where those varieties are spoken in Italy. 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 Calabrian instead.

One additional change to the original plan was in the locations in which 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 as a control: we selected the English varieties spoken in New York and Boston. However, 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.

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 geographical proximity between the two countries, the relationship of these speakers with their homeland was as severed as that of the Italians who had emigrated to South America. Moreover, no interfering additional languages (besides the target varieties and the contact language) were detected. It was therefore decided that the data collection for contact with French should be moved from Quebec to Belgium. Finally, fieldwork in Italy has not been prioritized for the reasons given above. Nonetheless, contact data from some Italian regions were collected through online questionnaires: observations on this process are outside the scope of the current study.

This article is based on the fieldwork sessions carried out by the Microcontact team. The first session targeted Argentina only (cf. Section 3 below) and took place in May 2018. This 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 study in Belgium was carried out in November/December 2019. It will be immediately clear that in these early fieldwork sessions the primary focus was not on data elicitation (although we did have a questionnaire to ascertain at least some basic facts regarding heritage syntax), but rather on checking the status of these languages and looking for speakers with the right profile, given that very little to no up-to-date information was available to us regarding these heritage speakers and their languages. Our intention was to gain an initial overview of the syntactic profile of these speakers through the first fieldwork studies, then to return to Europe, analyse the data, and formulate some hypotheses, before carrying out more extensive fieldwork later to verify them. This second, more extended, period of fieldwork was planned to take place in

2 Documenting Italo-Romance minority languages in the Americas

spring-summer 2020, but this has been impossible because of the COVID-19 pandemic, which means that our data are mostly incomplete, but we made use of an online data collection that did bring some results.

This article, however, does not present an analysis of the results of our investigation; instead, it provides a report of the organization and realization of the data collection, with specific focus on the fieldwork part.

Table 1 presents an overview of the number of speakers we managed to reach and interview during fieldwork 1 in the various locations. They are listed by generation. Table 2 provides an overview of speakers interviewed remotely during the pandemic.³ At the moment, it is not clear when fieldwork 2 will be able to take place, nor whether it will be possible to undertake fieldwork before the end of the project, which will be in June 2022.

Table 1: Number of speakers interviewed by generation

	Brazil	Argentina	Canada	US (NYC)	Belgium	Total
G1	7	50	34	32	6	129
HS 1	1	14	2	26	2	45
HS 2 and ff.	42	10	–	–	–	52
Total	50	74	36	58	8	226

Table 2: Number of speakers interviewed remotely during the pandemic

Brazil	Argentina	US	Belgium	Italy	Total
44	61	4	3	228	340

Given the age of the speakers and the conditions under which the fieldwork was planned to take place, we expected the data collection to be quite difficult: in this paper, we discuss the issues that arose during the preparation of the fieldwork and while it was underway. 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 (“where we started” subsections). Then,

³Additional data were collected from France (4), Australia (3), and Uruguay (1).

we describe what the actual situation turned out to be (“what we found/did” subsections). We conclude each section with a list of tips and warnings about what needs to be taken into account when setting up similar research.

More specifically, Sections 2 and 3 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 an elderly population.

2 Documenting Italo-Romance varieties in the Americas

2.1 Where we started

Fieldwork for Italo-Romance varieties in the Americas is somewhat 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 that are now found outside of their original environment. 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, the situation is not comparable to that of documenting a previously undocumented language from an 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 make-up of the Italo-Romance speaking communities for this section, and on the results of the syntactic research in the next section.

The initial fieldwork was preceded by a data crowdsourcing enterprise, which consisted in asking younger generation speakers to record the elderly members of the community and upload the recordings on an interactive atlas. The atlas can be found here: <https://microcontact.hum.uu.nl/#home>. While this atlas had a large response from Italy, both North and South America were almost completely unresponsive. The entries that are now visible on the atlas were mainly uploaded by our fieldworkers.

Before turning to the presentation of each fieldwork area, some general considerations regarding data protection protocols that are in place in Europe but not elsewhere are in order. 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 directives within the EU.⁴ These directives may not be entirely consistent with those of the non-EU countries. The challenge lies in checking the GDPR against the regulations of the target country, aiming for an optimal level of mutual adherence. This can be done with the support of the embassies, but responses can be slow or even unforthcoming. 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.

Regarding the data collection itself, it must be kept in mind that Italo-Romance communities in the Americas have very different characteristics. In this subsection, we review the information about Italo-Romance communities that was available in the literature before the start of the project. It will be immediately clear that the type of information available and the level of detail in the data reported in each subsection is significantly different. This is a reflection of the documentation of these varieties and their speakers: North America – particularly the US but also Canada – has a long tradition of heritage studies, mostly in the field of sociology and anthropology, but also in linguistics. Furthermore, the ethnic background of US citizens has always been meticulously recorded; we therefore know exactly how many Italians live in each state, and where they are from, while this is not the case for Argentina and Brazil. Italians in Argentina in particular have mingled with the local population and switched to Spanish much faster than any other group.

In the following subsections, we report the kind of information that was available to us before we planned the first fieldwork.

2.1.1 Argentina

Argentina was a very popular destination for Italian immigrants in the 19th and 20th centuries. According to the website of the Ministry of Foreign Affairs and International Cooperation of Italy, Argentina received 57% of the total number of Italian people who emigrated overseas between 1946 and 1955.⁵ Maurizio (2008) reports data from the Instituto Nacional de Estadística y Censos de la República Argentina, showing that a total of 2,604,447 immigrants lived in Argentina in 1960; 18% of these immigrants were from South America, and 82% from other countries, of which 31% were from Italy.

⁴The GDPR became effective after the starting date of our project. Before then, a different set of rules regulated data protection within the EU: we therefore had to modify our original protocol to make it compliant with the new regulations.

⁵These data are taken from https://www.esteri.it/mae/doc_osservatorio/rapporto_italiani_argentina_logo.pdf.

At first, Italian immigrants moved to Argentina only temporarily for economic purposes, with the aim of improving their quality of life once they were back in Italy. This form of immigration began between the 18th and the 19th century, but became a mass phenomenon in the last quarter of the 19th century. Temporary immigrants either stayed in Argentina for several years and then moved back to Italy, or they were seasonal workers, who left Italy during the local autumn/winter and came back during the spring/summer. In the period of mass immigration, this trend was accompanied by permanent immigration, where families would relocate and settle in the new country (Ferrari 2008). Geographically, the first immigrants were predominantly northern Italians; in the final years of the 19th century, immigration from the South increased, until southern Italians formed the majority of immigrants just before World War I.

The places that were most influenced by the arrival of southern Italians were cities such as Buenos Aires, Córdoba and Santa Fe. In these cities, the number of immigrants from various countries was extremely high. Spanish was not only the official language, but also the *lingua franca* for immigrants who 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 it should be noted 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 of these, in the Boca, a district of Buenos Aires, was created by immigrants from Genoa, whose dialect is described as extremely widely-used and alive from the second half of the 19th century onwards, but 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 in popular use in the 1980s (Meo Zilio 1990). This is the only information we had regarding these varieties in Argentina.⁶

With regard to the situation for Italo-Romance varieties from the south of Italy, we had only minimal details before the fieldwork took place.

⁶A reviewer suggests that we should include the exact number of speakers and the statistics regarding these varieties in south America in previous centuries. These statistics do not exist, and we are including here everything that we were able to find. Although the information we have is obviously incomplete, we still believe that it provides a useful idea of the language situation of Italian emigrants in Argentina.

2.1.2 Brazil

Brazil was one of the main destinations for Italian emigrants in the second part of the 19th century. According to Cenni (2003), 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 São Paulo, Italian cultural heritage is still alive, but heritage languages ('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 in Rio Grande do Sul is different: they settled in extremely isolated mountain areas, which allowed their varieties to resist the ban on the use of the language imposed in the 1940s and the pressure exerted by 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 multiple sociolinguistic studies conducted in Brazil. These speakers, despite being mainly third- or fourth-generation HSs, are native speakers of an Italo-Romance variety and in most cases have no knowledge of Italian. They are hence 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.

Italian immigration to Brazil started to decline at the beginning of the 20th century and came to an almost complete halt after World War II, earlier than in other American countries. This makes it very difficult to find G1 speakers who are still alive. One exception to this 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. We therefore 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 unlike that found in Argentina and Brazil. To begin with, the demographics of the Italian emigrant population are quite different from what is found in the rest of the Americas: emigration to Canada, and more specifically to Quebec, is relatively more recent than that to the other areas under investigation. Although the first records of Italian emigration to Canada date back to the last quarter of the 19th century and the flow of migration never completely stopped, the period of intense movement was fairly brief, lasting from 1951 to 1967. Due to this demographic difference, the majority of first-generation speakers in Canada were typically not (completely) illiterate when they left Italy, as they had received at least some formal education in Italian, and had also been exposed to Italian in the increasingly popular media. Therefore, we knew that Italian, or at least a non-standard variety thereof, would be a not insignificant source of interference on the dialects spoken by our informants. We based our knowledge on the available studies on the language(s) spoken by the Italian community in Montreal (focusing on their Italian: Reinke 2014 and the extensive work by Villata, e.g. Villata 2010).

Moreover, the overall migration flow to Canada was never at a level comparable to that of our other research areas: the peaks, registered in 1956, 1958, 1966, and 1967, were of roughly 28,000 people a year.⁷ We therefore expected to find fewer participants in this research area than in our other target areas, even if we did not have the exact numbers for Quebec alone.

Finally, some parts of Quebec are *de facto* bilingual areas: while French is the only official language of the province, English is widely spoken (as well as being an official language of the country), especially in Montreal. We based our decision to include the area in our study on knowledge of pro-French campaigns and policies, which were particularly prominent in the 1970s. However, 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, particularly a non-Romance one.

2.1.4 US

Italo-Romance varieties have been exported to the US since (at least) the 19th century; the 1880 census lists 81,249 Italian migrants, a number that had increased sharply to 4,114,603 by 1920 (cf. Cavaioli 2008). In the census carried out in 2000,

⁷Source: ISTAT, http://seriestoriche.istat.it/fileadmin/documenti/Tavola_2.9.1.xls.

2 Documenting Italo-Romance minority languages in the Americas

“Italian” was reported to be spoken by about 1,000,000 people in the US, with the most significant numbers concentrated in the Northeast of the country, where we carried out our fieldwork.⁸ However, according to the multi-year American Community Survey 2009-2013, there has been a decrease of about 300,000 speakers, i.e. a drop of a third in just 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 find Italo-Romance HSs after the first US-born generation.

Moreover, these statistics combine all the languages imported by Italian migrants under the umbrella term “Italian”. This is slightly inaccurate, as pre-World War II migrants mainly exported their local languages, and had minimal knowledge of Italian, or perhaps none at all. This situation changed after World War II, particularly after 1965 with the Immigration Reform Law, thanks to which the families 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 educated in Italian and were hence no longer monolingual Italo-Romance speakers as the previous migrants were (Haller 1991: 391-392; De Fina & Fellin 2010). As a result, the speakers’ competence in their local languages began to be affected by the new wave of Italian, as well as by English. Moreover, the coexistence 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. This koine has since been the focus of the majority of studies of 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, such as Sydney Italian (Bettoni 1990, 1991) and Montreal Italian (Reinke 2014).

For New York, Haller’s (1987, 1993, 1997a, 1997b, 2002) work provides a solid description of the sociolinguistic situation of the Italian community between 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’

⁸New 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>]. Note that the same level of accuracy and documentation is not available for all countries; we provide here what was available to us before prior to our fieldwork, and the information on which we based our planning.

dialectal Italian, Italianized dialect, pidginized American Italian, and archaic dialects”. With regard to 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 had been functioning as the “community language” for decades (Haller 1991, Haller 1997a: 401). A few decades later, this situation has reached a point at which the Italo-Romance varieties are increasingly fading away, and Italian is taking over or Italo-Romance is being abandoned altogether in favor of English.

2.1.5 Interim summary

Although we were aware of the social differences between the areas in which we planned to collect data, before starting our fieldwork 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 in Italian, if any (even in its regional varieties). These G1 speakers would have maintained little contact with their 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 this scenario would have been the possibility of systematically excluding 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 a good level of detail at the microvariation level too.

However, in some cases the socio-historical differences proved to be more far-reaching than expected and to have not insignificant impact on the linguistic profile of our informants. These issues are discussed in more detail in the following section, along with some practical matters to be considered when organizing fieldwork, and the solutions that we found to the various problems that arose.

2.2 What we found

2.2.1 Argentina

As discussed above, Argentina was the destination of huge numbers of immigrants from the end of the 19th century onwards. However, when we tried to contact associations of Sicilian, Neapolitan and Abruzzese speakers by email, we received a large number of Non-Delivery Reports (NDRs) that led us to conclude that they were no longer active. We also asked for information from Facebook groups dedicated to Italian descendants in Argentina, as well as from some distant relatives of ours and of other contacts with relatives in Argentina.⁹ Neither crowdsourcing nor any subsequent attempt helped to identify speakers who would be eligible to take part in our study.

Due to this lack of information, we decided to carry out a pre-fieldwork exercise, with the specific aim 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 this was where finding contacts in advance had proved most challenging. 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 first main fieldwork session, during which we targeted immigrants who had moved to Argentina after World War II, as well as the small number of their descendants who had acquired the Italo-Romance variety. As in other American countries, institutions, especially schools, played a major role in the diffusion of the monolingual Spanish model. The researchers were told multiple times that teachers explicitly suggested, or even ordered, that the parents should speak only Spanish to their children, in order to avoid “confusion” for the child.¹⁰

Unfortunately, these interventions were effective in the majority of cases, meaning that the Italo-Romance varieties were abandoned by almost all immigrants.

⁹In 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 into society, I’m not sure whether they still speak their dialects. For instance, my grandfather Francisco didn’t speak his dialect, he spoke Spanish.’]

¹⁰As 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.’]

Exceptions are found primarily when there were elderly family members (especially grandparents) who never learned Spanish and thus kept speaking their Italo-Romance L1 to their grandchildren.

Despite the geographical distance, during both fieldwork studies in Argentina we found that the local Italian community managed to keep strong bonds to their hometowns through frequent visits to Italy (particularly from the 1980s onwards) and through the countless regional and local associations. According to our informants, these associations were particularly active in the 1940s-1970s, and helped to recreate a sense of Italian community through recurrent parties and celebrations. The members of the associations have tried 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 their language, switching to Italian or to Spanish.

More generally, Italian gained ground because of marriages between people from different regions of Italy, who chose to speak Italian to their children, when they could, in order to keep a stronger bond with their home country. In the bigger cities, there are also Italian schools that some of our speakers attended. Finally, although the immigrants and their descendants feel a particular link to their region, they feel proud of Italy as a whole and identify with it, particularly when they talk to people who are not descended from Italian immigrants. As a consequence, many first-generation immigrants (as well as the subsequent generations) speak Italian alongside their local variety, and they all insisted on speaking Italian to the fieldworkers.

The various Italian associations were very useful in our search for informants, since they know most members of the community, but even they could not identify more than a couple of speakers each, as there are not many speakers left. However, some associations do offer courses in their Italo-Romance variety. These courses are attended by second- or third-generation immigrants who never developed a high proficiency in the language and wish to improve it or even learn it from scratch. In the Friulian association of Buenos Aires and in the Piedmontese association of Córdoba, for example, 8-10 people followed an Italo-Romance language course. These courses were a useful source for our search for HSs.

The informants we interviewed were found mainly through members of these associations. In Santa Fe, we found some speakers thanks to the Italian scholars at the Universidad Nacional del Litoral, who are running a project on the Italian cultural heritage of the city.¹¹ In three cases we found speakers of Cocoliche

¹¹The person responsible for the project is prof. Adriana Crolla (<http://www.fhuc.unl.edu.ar/portalingo/crear/gringa/>).

(see §2.1.1 above): unfortunately we could not interview them, because they were some of the oldest members of the community and they were not able to perform our tasks. Interestingly, they were usually no longer able to distinguish between Cocoliche and their own Italo-Romance variety, a situation that we also found in New York with the koine and the Italo-Romance varieties.

Overall, most informants were kind but somewhat suspicious at the beginning, especially in the bigger cities: they all refused to allow the fieldworker to visit them at home, unless they were accompanied by a member of the community whom they already knew. As a result, when possible, the interviews were held in emigrant association premises. However, in some cases they had to be carried out in bars, which made for a less than ideal situation, as informants could be distracted, the audio stimuli of the questionnaire were difficult to understand because we had to lower the volume, and the recordings were affected by background noises.¹²

With regard to the geographic distribution, 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 we could find very few second- or third- generation speakers with a high proficiency in the language. In the smaller towns and villages, on the other hand, the Italo-Romance varieties have virtually died out. One example is Colonia Caroya, in the province of Córdoba: until a few decades ago, Friulian was the main language (Spanish being the only official language), to the extent that, according to our informants, it was impossible to find a job there if you did not speak Friulian. The situation today, however, 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

Unlike in Argentina, Italo-Romance varieties in Brazil have survived mainly in the countryside and, to some extent, in bigger cities in southern areas of the country, as we highlighted in §2.1.2. Immigration from Italy after World War II was very limited, and it was hence challenging to find first-generation immigrants with the right profile for our research. The state of Rio Grande do Sul was the

¹²One reviewer observes that it would have been possible to train local members of the community to perform the data collection for us. This was not possible due to a lack of time, during the first round of fieldwork, given that the researchers had to travel extensively to meet the various speakers. The second round of interviews, which was performed online, was instead realized with the help of trained local community members, who knew the interviewers by then and gladly agreed to help when possible.

best option, as both northern and southern varieties are still spoken in the area by G1 and HSs.

Ten interviews were carried out in Porto Alegre, which is home to a large 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 mistrust shown by local associations of HSs towards the research; in particular, one of the local associations refused to help the researcher to find informants. A few participants were found as a result of posting on Facebook groups of descendants. The best result in Porto Alegre, however, came from the successful cooperation with the Calabrian Center of Rio Grande do Sul; this contact was established prior to the fieldwork and was helpful in finding Calabro-Lucanian G1 and HSs. The Calabrian community in Porto Alegre is made up of immigrants (and their descendants) from the town of Morano Calabro; they all therefore speak exactly the same variety, the Moranese dialect. However, the general sense of distrust was also evident here; obtaining signed informed consent was particularly problematic, since most of the participants were afraid that it could be a scam.

The interior of the state of Rio Grande do Sul is home to almost two million descendants of immigrants from northern Italy, most of whom have maintained the language of their ancestors across generations, as a result of the isolated nature of these communities. 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 isolated nature of these communities, which helped to preserve three northern varieties (Venetan, Friulian, and Eastern 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 not only made it possible to physically reach the speakers, but were also of great help in communicating with older people who, in some cases, had never left their villages and were hence not always willing to talk to a foreigner or to be recorded. Unlike the Calabro-Lucanian informants in Porto Alegre, speakers of northern varieties in Rio Grande do Sul are the descendants of immigrants from different areas in Italy. Their dialects are not exactly identical to each other, nor to the varieties of the languages spoken in Italy. Moreover, isolation from Italy has allowed the preservation of archaic features of the languages in these varieties. Obtaining written informed consent also proved problematic for informants in the Serra Gaucha and in the Fourth Colony of Immigration. The interviewer decided to opt for recorded oral consent by the informants: the whole informed consent form was read out loud during the recording and participants consequently agreed to be interviewed.

Overall, the good outcome of the data collection in Brazil was heavily dependent on the contacts the interviewer established prior to the fieldwork. The need to have trusted relationships with local members of the communities proved to be essential.

2.2.3 From Quebec to 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 the age of the speakers (younger, which in turn entailed physical and cognitive difference – cf. §3.3.1) and their 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 in Quebec had managed to achieve a reasonable level of economic security, which had led, among other things, to increased contact with Italy. 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 a year. Even those who had no direct or prolonged contact with their hometowns had been in regular contact with their relatives in Italy via long-distance communication devices over the years.

The different sociolinguistic settings of the two main areas of investigation, Montreal and Quebec City, posed a further a problem in Quebec. The latter has a small Italian community whose 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 find one HS of an Italo-Romance variety in the whole city (Venetan), compared to 9 G1 speakers of different Italo-Romance varieties, mostly Piedmontese, Venetan, and Friulian. Montreal, on the other hand, was the destination of a massive flow of emigration from Italy: the traditional Italian area, *la petite Italie* ('little Italy'), is nowadays a sheer memory of the Italian emigration. Despite being home to the Church of the Madonna della Difesa, built by the Molisano community in the late 1910s, and to some Italian shops and cafés, over the years the bulk of the Italian community has moved away from the area, 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 municipalities, especially Laval. However, this setting also proved detrimental to the preservation of the Italian dialects: the presence of emigrants from linguistically different areas of Italy and the need for mutual help and support resulted in the

development of an Italian koine, which, alongside English and French lexical borrowings, includes many regional structural and lexical features and a (broadly speaking) Italian structure, somewhat resembling that which is found in New York. This variety (“Italianese”, e.g. Villata 2010) is most readily available to our informants for daily communication within the community and, over time, has come to overshadow 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), compared to 22 G1 speakers of various target languages.

A further difference between the two areas of interest in Quebec is related 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. The widespread use of English, despite efforts (particularly via education policies) to make all new residents and their descendants use French, 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 also inside) the community, especially in Montreal.

In an attempt to find a speaker profile that more closely matched the ideal speaker for this research, we turned instead to Belgium. 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 the country and its greater geographical accessibility, Belgium had been originally excluded from the investigation areas because of its (relative) closeness to Italy, which could have led to extensive contact between the emigrants and their hometowns. However, after we had established that the Italian population in Quebec did not sufficiently match the required profile, and given the accessibility of the area, we decided to run a small-scale fieldwork exercise 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, as well as Brussels;

in total, we interviewed eight informants: six G1 and two HSs. Our original concerns regarding the possible stronger connection between this population and their homeland turned out to be unfounded: we discovered that the terms of the contracts for the mine work ensured that the miners did not go back to Italy for the whole duration, reducing the links considerably. After the contracts expired, some workers went back to Italy for good; those who stayed in Belgium continued their lives away from their home country, as they had previously.

The general speaker profile was a better match for our ideal speaker. Due to the disruptions to the school system brought about by the war, we found in our interviews that the emigrants, in most cases fairly young men, were mostly illiterate, which was no doubt beneficial to the maintenance of the dialects. Furthermore, the well-organized migration flow brought people from the same areas of Italy together, both to live and to work: this ensured 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). This situation is demonstrated by the presence, even today, of small municipalities in the south of Belgium where Italian dialects are still widely spoken among the members of the Italian community: an example is Sicilian in Morlanwelz, which was the destination of a considerable wave of emigration from Villarosa, in Sicily (Enna province).¹³ These local and linguistic clusters, together with the limited contact with Italy over many years, also resulted in a very limited knowledge of Italian. Moreover, 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, alongside their original dialects. No other contact language is 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 originally came from southern Italy, mostly from Sicily and Campania. Up to this point we have found very few speakers of northern varieties. Unfortunately, it has so far not been possible to continue this small-scale fieldwork with a more extensive data collection.

2.2.4 US

The search for Italo-Romance speakers in a metropolis such as New York City was not straightforward. This is due to the fact that, over the years, the different communities of Italians experienced disaggregation and displacement within and outside the urban area, as we discuss below. The most fruitful means of finding

¹³This is reflected, at institutional level, by the fact that the two are twin-towns: <https://en.wikipedia.org/wiki/Morlanwelz>.

our speakers was to bypass the official routes, such as the (unresponsive) Italian Cultural Institutes and Embassy, and instead to approach:

- 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, as well as one family in Jersey City, NJ. The people we interviewed were speakers of Friulian, Nònes (a Ladin-Lombard/Venetan transitional variety from Trentino), Eastern Abruzzese, Neapolitan, and Sicilian. We also interviewed speakers of varieties that were not considered in our other field trips: Eastern Campanian, Cilentano (a southern Campanian variety), Apulo-Barese (from central Apulia), and Ciociaro (an upper-southern variety from southern Lazio). Moreover, as mentioned in §2.1.4, most of these speakers had some knowledge of the spoken Italian that they exported (in the case of G1) or learnt in the US (some HSs). However, many HSs only had active knowledge of the Italo-American koine, especially those HSs whose families came from the south of Italy, as we discuss below.

G1 speakers migrated between 1940 and 1980 from different areas of Italy, where the local varieties are spoken alongside (regionally marked) Italian. Most of the G1 speakers who arrived immediately after World War II settled in the same place as the historical communities of Italians during previous migration waves. These areas include (but are not limited to) Manhattan (Rose Hill, Little Friuli in Murray Hill), Brooklyn (in and around Bensonhurst, Williamsburg-Green Point), and Queens (Astoria). However, due to the arrival of new migrant communities and the increasing cost of living in the city, most of these Italian communities were forced to move away from these neighborhoods and to relocate to more peripheral areas.¹⁴ These displacements led to the partial or total dissolution of once-compact linguistic communities; moreover, language shift to English-only happened very frequently 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 HLs proved rather challenging, as the large majority (especially from southern Italy) had been forced to switch to English-only by their families

¹⁴Mainly further East in Queens, further South in Brooklyn, and Staten Island, or outside the five boroughs of NYC.

for integration purposes, therefore completely abandoning their HLs and retaining, in the best cases, only a passive knowledge of them.

The linguistic profiles of the US-born HSs we interviewed included: (usually, highly educated) speakers with active competence of both their own local variety and Italian, which they learnt from (educated) family members, or during secondary education; speakers with different degrees of active competence in their own local variety (depending on their age, the type and length of exposure to that variety, and cohesion of their community), but little competence in Italian; and speakers with passive knowledge of the local variety of their own families, which they define as “the archaic dialect”, and active competence in a regionally influenced variety of (Americanized) Italian koine, the lingua franca, which they refer to as “the modern dialect”, or Brooklynese/Queenese Italian (alongside the local Brooklynese/Queenese English).

Unsurprisingly, the most proficient speakers are from the elder generation, i.e. over 70-75 y.o (with very few exceptions). The linguistic repertoire of these speakers is extensive, as they learnt their own HL from close-knit communities of Italian-born parents, grandparents and/or other relatives, who had emigrated in the first half of the 20th century to the specific Italian areas/neighborhoods of New York City. Many of these elderly speakers also maintained close connections with their families in Italy, allowing them to also be exposed to the home dialect and/or spoken Italian. Indeed, these HSs learnt a conservative, “frozen-in-time” variant of their own HL from their families of G1 *émigrés*; when they visit their family’s birthplaces in Italy and interact with locals there, they are told they still speak an archaic version of the relevant dialect (cf. Aalberse et al. 2019: 114). This is not only due to the fact that older HSs learnt a conservative variant of the HL imported by their families, 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 over the last 70 years.

One feature shared between some Italian-born G1 speakers who migrated to the US in their early years and US-born HSs is that they all grew up as sequential bilinguals. They first acquired their own Italo-Romance HL (the local dialect and/or the supraregional Italian koine), 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 y.o., the level of proficiency in the HL diminishes rather drastically. This is likely due to a less constant exposure to the relevant HL, or to a drop in usage on a daily basis. Moreover, in addition to the decrease in input, 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 HSs grew up learning a supra-regional variant of Italian, originally taken overseas by their families after World War II and later influenced by English, while their own local HL was mainly heard from grandparents, older relatives and/or elders in their neighborhood, but was not actively used. Indeed, from the intermediate to the new generations, the lexis, phonetics/phonology and syntax of the relevant baseline dialect appear to have blended to different extents with a spoken variant of Italian, as well as English, resulting in the Italo-American koine. In fact, speakers under 40 years old (mainly from the south) appear to show an “established confusion” about the Italo-Romance language they speak. They self-report that they are not able to speak “proper Italian” and can only speak “the dialect”, but they actually speak this Italo-American (southern-based) koine. However, the Sicilian and Nonesi communities of HSs seemed to have better preserved their local dialects, across all age groups.

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, sometimes leading to difficulties of comparison 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 the Americas 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 exit action in case of danger; furthermore, it should contain a safety protocol that you may establish with your PI.¹⁵
- 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 need help, and make your presence known to them. Establish a daily routine with your home supervisor, like sending an email or a message at a given hour every day to confirm that everything is okay and you do not need help.

¹⁵For 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/>.

2 Documenting Italo-Romance minority languages in the Americas

- Make sure emails to potential referents are short, clear and personalized. After gathering and reviewing any available sources for contacts, including outdated webpages that might suggest further connections, it is crucial 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 should be clear and concise and should include statements about (i) the purely scientific purpose of the research survey, (ii) the non-profit nature of the enterprise, and (iii) compliance with the data protection protocol (i.e. anonymity), although public institutions are usually willing to be acknowledged in research outputs. Whilst the message template should address a varied range of potential (sources of) informants, a plain, somewhat detached invitation may not engage all the intended recipients. It is worth adjusting the text in a way that demonstrates a sensitive understanding of 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). 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 spend time participating in the life of the regional associations: parties are not only entertaining, but also an excellent means to meet people in a more relaxed environment than an interview offers, and to exchange contacts. In our case, we carried out pre-fieldwork in May as the fieldwork had to take place in the following spring.¹⁶

- Involve the local communities (even more): try to train local people to carry out the interviews.

Training local people to carry out interviews when the researcher is not present is a very good idea, if time constraints allow. Ideally, this should happen for each collection of data, but this cannot always be achieved, especially if the time to be spent in one location is too short. In this respect,

¹⁶A reviewer asks how many days are recommended. We do not have a precise answer for that: the more the better, obviously, so that the researcher can get to know the people and the community a little more. In our case, the pre-fieldwork lasted about 20 days, and the limit was determined by budget considerations rather than anything else.

when calculating the budget in your project proposal, make sure that you allocate some money for your fieldwork assistant as well as for the speakers. Note that university-internal regulations might make it very difficult to transfer money overseas, so paying in local money to be reimbursed later is preferable.

- Consider technical problems in isolated/remote areas.
In the most remote areas that we visited in our fieldwork, especially in Brazil, some basic IT requirements were not met: for instance, the 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 carry multiple rechargeable batteries for laptops and recorders; alternatively, the questionnaire should be structured in a way that allows it to be carried out without a laptop. In that case, it is a good idea to print out the questionnaire, so that it is possible to ask at least some of the questions if the laptop cannot be charged. 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 might genuinely not have enough time to spend on a questionnaire (in our case: between half an hour and one hour), as “time is money” in their busy schedules, despite their interest and willingness to participate in the study. In New York, this was true for most of the informants, regardless of their age, and was also true to a lesser extent in Quebec. Moreover, in large metropolitan areas such as New York and Montreal, communities are scattered across the city, and travelling on public transport 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 transport, and are very expensive. Allow therefore plenty of time to reach your destination.
- Remember that it might be culturally inappropriate to ask speakers to sign a document they cannot read.
In this case, still complying with the legal requirements, the informed consent form can be read aloud, with speakers giving at least their oral consent to the collection and processing of their personal data. Make sure this procedure is approved in the ethical and data protection protocols before you start your research.

- Consider the mistrust of the informants.
Both in bigger cities and in more isolated villages there may be a general feeling of a lack of safety. To overcome this sense of mistrust and to make the informant feel at ease even in the presence of a stranger, it is preferable to go to interviews with another community member whenever possible. This is also beneficial for the fieldworker.
- Be prepared to interview people in unconventional places.
As much as we tried to make arrangements in advance to carry out our interviews in a quiet environment, this was not always an option (see also the feeling of mistrust mentioned above). A good solution, when private spaces are not available, would be to carry out the interviews in offices belonging to the various cultural associations. However, even this option is sometimes not available. It might then prove necessary to carry out the interview 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 be challenging).

3 Syntactic tests

3.1 Where we started

Our research consisted of three parts: the first aimed to assess the proficiency of the speakers; the second tested the presence of the syntactic phenomena under analysis; and the third gathered sociolinguistic information about the informants' language history and use. At this stage, as we had only preliminary hypotheses to test, we did not concentrate on designing the exact method for data elicitation; rather, we tried to understand whether what was reported about heritage languages and the particular phenomena under investigation was true.

In the first fieldwork trip, the main task was to ascertain the existence of subject clitics and DOM, as well as ternary demonstrative systems, and the syntactic conditions under which these phenomena might be present.

3.1.1 Assessing proficiency

We decided to make proficiency testing the first part of our research because we wanted to make sure that our informants were genuinely able to speak the

target dialect, rather than Italian with some dialectal expressions (for legitimate concerns on the issue, cf. §2.1). Once the data were collected, each component of the group listened to them. The group includes researchers who are native or highly fluent speakers of all the varieties under investigation; most of them are also native speakers of Italian.

Before beginning the fieldwork, a data crowdsourcing enterprise was carried out, over the course of several months. Through this crowdsourcing we had hoped to identify and hence pre-select speakers with the right profile, i.e. proficient speakers, to interview during fieldwork. However, we did not obtain the desired results from this type of data collection, as we did not receive sufficient responses from either North or South America. We therefore interviewed all the speakers who made themselves available, on the basis of their own linguistic self-evaluation, and we needed a more reliable method of assessing 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 proficiency assessment will be discussed more in detail in the next subsection, here we will introduce the bases upon which we decided to take proficiency into consideration.

Our fieldwork researchers were all native speakers of Italian, and were therefore able to assess whether the speakers were using some form of Italian or the Italo-Romance variety. Afterwards, the data were checked by the rest of the team, which included native or very fluent speakers of all the Italo-Romance varieties under investigation.

Spontaneous speech was chosen as the introductory task as it offers 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 at all. We therefore thought that a good way to make them feel at ease could be to allow them to talk freely: we found that some of the less proficient speakers had some issues with speaking the language to start with, but became more and more fluent while talking to us. To try and trigger the use of the dialect, we mostly asked them questions about their childhood and their arrival in 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 sociolinguistic information (year of emigration, level of education, current and past dialect usage, *etc.*) that we needed to control for, to keep our sample of informants as homogeneous as possible, without bombarding the speakers with very direct and structured questions. Thirdly, the spontaneous speech data were used to complement the

information gathered from the questionnaire, and to cross-check whether spontaneous production might reveal different linguistic patterns with respect to the elicited data gathered in the questionnaire.

In addition to using spontaneous speech, we also assessed the level of proficiency through the HALA test, developed by O’Grady et al. (2009) and widely used in heritage communities to check the level of proficiency of the 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 proficiency correlates with frequency of use in all language domains, which has direct consequences for latency times. We hence decided to include an additional task based on the HALA materials, as will be explained in more detail in §3.2.1.

3.1.2 Towards the syntactic questionnaires

The second part of our research 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 HSs. Nonetheless, we wanted to develop one unified questionnaire, so as to make our results directly comparable. We therefore 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 would be the HSs (for extensive remarks, cf. Polinsky 2018: chapter 3). Of course, the three phenomena have different instantiations in each of the Italo-Romance varieties under investigation; variation between them is however at the microlevel, so most tests were simply translated from one language to the next without losing crucial information, or without changing information structure, for instance.

We excluded grammaticality judgments and translations. Polar grammaticality judgments (Yes/No type) are not granular enough, while scalar judgments (in which each experimental item is rated on an n -point scale, e.g. Likert scales) detect finer differences among the stimuli, but ultimately also pose some problems (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 cannot be safely assumed that an n point scale is sensitive enough to faithfully represent the acceptability continuum, both in its extent 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 throughout the task, scalar grammaticality judgments are clearly demanding on the memory load side as well. Finally, grammaticality judgments prove particularly difficult for HSs, (Polinsky 2018: chapter 3, and references), which would have made it difficult to use the same questionnaire for all our participants.

Translations have been commonly used in 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: ASI_{It}). However, we were particularly concerned with avoiding the interference of any other language while performing the tasks, particularly given that the Italo-Romance varieties were not that frequently used and were hence expected 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 as 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 up 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, a flexible format of this type, with different variables to accommodate all possible needs of our speakers, would have made the translations not fully comparable, as they would have primed the informants in different ways. We therefore decided 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 number of studies on the issue (for a discussion and specific references, cf. Stadthagen-González et al. 2018), we judged that this format would be beneficial for our research in many respects, including 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. Variations on the two-alternative forced choice task described above were used to test these questions whenever possible and depending on the nature of the phenomenon: subject clitics, DOM, and, to some extent, deixis (with the support of pictures), were included; for deixis we added a semi-guided production task as well.

More information on how we paired tasks and phenomena and on how each task was designed and carried out will be provided in §3.2; in the remainder of

this section, 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 & Cordin 1989; Poletto 1993, 2000). However, Frasson (In press) found that in Brazilian Venetan subject clitics display pronominal behavior (see also Benincà & Poletto 2004). 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;¹⁷
- 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 & 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 realization of subject clitics varies substantially in this context; often, both a sentence with a subject clitic and one without a subject clitic are accepted by speakers;

¹⁷As shown in Poletto (2000), the behavior of subject clitics in coordinated structures is very nuanced. Following her analysis, we avoided testing coordinated structures with two distinct inflected verbs that have the same nominal object, as well as coordinated structures involving the same verb with different tense or aspect specifications; subject clitics may behave differently in these types of coordinated structures.

- vii. default agreement constructions: as with *vi*, realization 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 also accept 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. The main focus of our investigation with respect to deixis was 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 to 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, different demonstrative systems are observed: 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 systems of this sort 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 instance, 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). The Portuguese system is of this type, and differentiates between *este* ‘this’ and *aqui* ‘here’ (close to the speaker), *esse* ‘that’ and *aí* ‘there’ (close to the hearer), and *aquela* ‘that’ and *alá* ‘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 difference in interpretation: this is the case for

French adnominal and pronominal demonstratives (*ce* and *celui*, respectively, in their masculine singular versions).

The literature on deixis in Romance varieties highlights a high level of microvariation (see, for the most extensive overviews, Ledgeway 2015 and Ledgeway & Smith 2016): Romance varieties display all four systems and there is significant variation, 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.

3.1.2.3 Differential Object Marking. Differential Object Marking (‘DOM’; Moravcsik 1978; Bossong 1985, 1991), also known in the Romance literature as prepositional accusative (Diez 1874; Meyer-Lübke 1890–1902, 1895–1900), is the phenomenon whereby some Direct Objects (‘DOs’) are marked differently than others, depending on certain 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), while 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 (To appear), for Barese; Ledgeway 2009 for Neapolitan; Ledgeway et al. 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).

It should additionally be noted that the preposition marking DOs in these languages is the same as the preposition 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 have led 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 what we label as DOM might be referring to a range of different phenomena that just happen to share the same

superficial outcome.¹⁸ We wanted to know whether this is actually the case and, if so, whether these differences are simply the product of diachronic evolution, or if they have in fact developed from different starting points in different languages. Furthermore, in the case of contact with Argentinian Spanish, we wanted to investigate whether a possible change in the distribution of DOM in Italo-Romance varieties reflects the change found in 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 their arrival 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 of spontaneous conversation, but some of the speakers were so happy to talk to us that they talked for half an hour or even an hour. Their willingness to speak to us may also have come from their knowledge 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 speaking 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 mentioned in §3.1.1, we also performed an additional test to assess lexical proficiency on the basis of material designed 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 shown, and participants have to name the objects depicted as quickly as possible in the target language. 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 picture appearing on the screen (highlighted by an audio signal) and the participant's naming of that picture. However, for the test to be carried out successfully, it is necessary to compare the speed of lexical retrieval across the different languages in the participant's repertoire, 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, suggesting that longer times are simply due to external factors. Since the test had to be performed in different languages, we decided to only use a short part of the original HALA test: we used

¹⁸ A detailed discussion of this question is included in Luana Sorgini's ongoing PhD dissertation.

six items in the Italo-Romance variety before the questionnaire and then asked participants to repeat the test in the language they felt to be their dominant one after the questionnaire.^{19,20}

A complicating factor was the experimental setting: as already mentioned, we had to carry out interviews in unconventional locations, making it difficult to detect the signal sound that was intended to be 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 a number of 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, many of them do not have an orthography and are mainly spoken. This is not the case for all of them: some of the varieties under investigation have a long written tradition and therefore a standardized spelling convention, but even then their written systems show microvariation, mirroring the actual linguistic microvariation found across the Italo-Romance domain. Once again, presenting the speakers with a slightly

¹⁹Despite 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 meet the HALA guidelines. Ideally, the test should 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, this option was not available in our case.

²⁰An anonymous reviewer points out that “In addition to assessing reaction time, *etc.*, it seems careful attention needs to be paid to whether they are producing words in the ‘right’ variety – the regional variety whose proficiency you want to assess. This means KNOWING whether words borrowed from Italian are part of the vocabulary of the regional variety.” We do not see this as an issue at all. While fieldworkers were not native speakers of all varieties, they did know the various words in the different Italo-Romance varieties. Furthermore, the data were double-checked by native speakers after the return from fieldwork.

The reviewer also observes: “It’s also not clear to me why you could not follow the “one language a day” protocol. At the least, it’d be good to make sure the participant has been speaking in the relevant lg. for some time just before the test.” This first fieldwork session was mainly focused on checking the language profile of the speakers, rather than on eliciting data: thus, the fieldworkers did not have enough days in one location to follow the “one language a day” rule. Furthermore, after an hour of telling stories in their native Italo-Romance varieties, very often with other family members present, the speakers did not show particular problems with using the language.

different spelling system for their variety could have resulted in slight unfamiliarity with the stimuli, similar to what we might expect for varieties without conventionalized spelling mentioned above. Furthermore, the choice of one standard written variety over another might have triggered unwanted judgments on the spelling, rather than on the stimuli themselves. Another point that we had to consider is that most G1 speakers may have vision problems as well as issues with reading due to their age or to illiteracy.

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 so 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 New York, the language of interaction for the interviews was adapted to the speakers' relative confidence with the languages in question. G1 speakers who learnt English after their adolescence preferred being spoken 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 [...] when asked to switch to dialect, the informants generally continued to speak their high variety [(dialectal) Italian] 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 influenced our choice of tests. For subject clitics, a two-alternative forced choice task was the best way of identifying 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 anomalous behavior. This is shown, for instance, by the context of coordination:

(1) Friulian

- a. Al mangje e al bêf
he.SCL eat-PRS.3SG and he.SCL drink-PRS.3SG

- b. Al mangje e bêf
 he.SCL eat-PRS.3SG and drink-PRS.3SG
 ‘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 constitute pronominal behavior. The pronominal and 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 provided crucial support. Not only did it help to confirm – or otherwise – the results that we obtained through the questionnaire, but it allowed us to observe further aspects of the distribution of subject clitics that would otherwise have been 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.

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

- (2) 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 13 sentences testing DOM plus fillers, for a total of 24 sentences. Speakers of northern varieties were given 9 sentences testing DOM plus fillers for a total of 23 sentences. We made this decision because we were not expecting production of DOM on a wide range of arguments by speakers of northern varieties, as the equivalent varieties spoken within Italy are not typically considered to have DOM.

The informants were asked to choose between a sentence including DOM and one without: these stimuli were presented in random order. Although speakers

needed guidance when taking the questionnaire (e.g. sometimes they had difficulty understanding certain lexical items in the stimulus (due to microvariation in the lexical entry), and a translation had to be provided, the test worked in most cases and revealed differences in the use of DOM with respect to the homeland varieties.²¹ 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 judgments, 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 judgments, 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 with the elicited form because of the switch in the deictic center at the conversation turn. For instance, in the case of elicited imitation, the informant might switch the deictic center when repeating the sentence, e.g. *I am here* > *You are there*. While both sentences are equally grammatical, they change in their interpretive content, but this is not tested (or indeed testable) in an elicited imitation task.

To circumvent these issues, we selected a picture-sentence matching task and 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.

Our informants had to identify themselves with the speaking character and refer to the dog present in the context of the picture (Figure 2: a, b, or c) by choosing one of either two or three (depending on the system in the target variety) recorded audio stimuli associated with each picture. For instance, given Figure 2(a) 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

²¹An anonymous reviewer asks: “What if they prefer the non-DOM form because it’s more like std. Italian?”. In northern varieties, the choice of the non-DOM option is to be attributed to the absence of the phenomena in the dialect rather than to the influence of Standard Italian. We did not find a consistent preference for the non-DOM option in southern varieties.

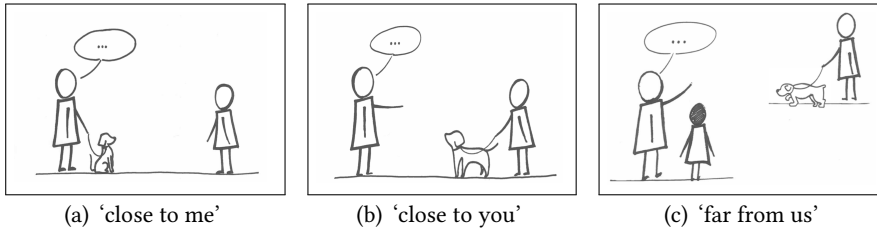


Figure 2: Picture-sentence matching task.

(close to me) is my dog', i.e. the proximal demonstrative *this*.²² The stimuli were also presented in random order for this task.

This set-up was not without problems: most importantly, some of our informants found it particularly hard to identify themselves with the speaker in the picture; similarly, some participants found it difficult to understand 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, in consideration of possible vision difficulties, the main characters on the picture were very large, 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 small distance apart. The informants sometimes explained their answer by saying: "it's still close, if it were *that* it would be something else". These size considerations, together with the identification problem, led to an overall higher rate of proximal forms even in non-proximal contexts. However, responses changed substantially when real-life situations were investigated. One such method used to elicit the (actual) distal forms 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 before the fieldwork, so the data collection was, in this respect, not uniform, and the results not completely trustworthy.

Semi-guided production proved a better test for deixis: in this case, we used three pictures of cats of different colors: black, orange, and white. These pictures

²²This was the target sentence for the pronominal context. Other syntactic contexts tested were adnominal (e.g. 'This dog is mine'), and demonstrative-reinforcer ('This here is my dog').

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 judged that the actual contrast within the context would make this task easier to perform for our speakers: they effectively needed to choose different demonstratives to make us understand which cat they meant. However, this method was also far from perfect: the most significant 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’). To try to elicit the target response, we sometimes suggested the whole set of answers in the contact language to help the informants to understand, without priming the language-specific demonstrative system.

One last issue that arose in the preparation of the deixis questionnaire was the clear difference between the tasks. 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; this kept the questionnaire to a minimum length so as not to tire our elderly informants, but still ensured the quality of our investigation. It was impossible, however, to run a comparable task for the deixis part of the questionnaire. Yet, it would have been ideal to show some sets of filler pictures targeting other phenomena alongside the images in Figure 2, which would also have had the benefit of making the task less repetitive. While designing the task, we thought that the addition of fillers would have been an online confounding factor (the informants would have had to correctly interpret multiple scenes) and would also have been time-consuming, particularly given that we were trying 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 from the participants, who thought that they were being asked the same question over and over again. Variation in the referent could have 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, who would all have spoken the exact same local varieties, and all

possible variables would have been perfectly controlled for. Moreover, all participants would carry out exactly the same task in exactly the same way. However, due to the scarcity and the heterogeneous make-up 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.²³

Regarding the characteristics of the participants, it is clear from the description presented in section 2 that they were not evenly distributed across the different varieties, host countries and generations. This must be taken into account when analyzing the results, particularly for the purposes of statistical analysis. For instance, there were eight speakers of Abruzzese in Argentina, only two in Canada and none in Brazil. Two speakers is too small a number 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 G1, and six were second-generation HSs. Again, given that two speakers cannot really constitute a separate subgroup, it was impossible to take “generation” into account as a variable in the statistical analysis. All 8 speakers were therefore treated as belonging to the same group, whereas in fact there was an important difference, namely that some of them were immigrants and others were born in the host country.

Moreover, as mentioned above, there were differences between communities in terms of 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 when analysing the results. For instance, we found certain differences between the use of SCLs by speakers of heritage Friulian in Argentina and Brazil. While an initial interpretation of a difference of this sort 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)

²³Recall, however, that this article is not concerned with these questions themselves, but is primarily reporting on the first fieldwork exercise, which had a mainly descriptive aim.

per condition (sentence type) was used, with the understanding that we would return to carry out more extensive and targeted second fieldwork. 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 the topic of investigation, which could in theory have led participants to use specific answer strategies (for instance: always picking the sentence with DOM). We chose this set-up to avoid having to stop the questionnaire half way through due to speaker tiredness.

Finally, as mentioned above, some of the interviewers had to improvise, either because 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 that some of the researchers carried out the experiment in the dialects, whereas others did so in the contact language or in standard Italian. It has been noted (Aalberse & Muysken 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 chose to present the participants with the full set of options for demonstratives in the contact language, which may have led to a higher number 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 brings with it additional issues that we considered before fieldwork; however, we found that it had a larger impact on the results than expected. Advanced age brings a number of common issues, such as 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.

A further problem is the difficulty of retaining long sentences: we therefore 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.²⁴ When it was impossible to provide 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.

3.3.2 Tips and warnings

Here are some tips to design a good questionnaire for heritage speakers:

- 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 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 speaking in their non-dominant language without a predefined topic, some (controlled) production tasks can help (and can make the collected data more comparable). Remember that spontaneous speech is very useful in assessing proficiency, too. If the elicited data contradict the spontaneously produced data, they should be excluded (at least, we chose to exclude them).
- Use the target variety when possible, to minimize the interference of other languages.
- If the questionnaire is too long and the researchers wish to test at least two phenomena, use the questions for the other phenomenon as fillers in your questionnaire.
- Agree in advance what should be done 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. It is therefore advisable to think about the

²⁴An anonymous reviewer points out that heritage speakers have difficulty in giving acceptability judgments. We are aware of that, and that is precisely why we went for forced choice and sentence completion tasks rather than the classical generative method of “is this sentence acceptable for you?” which we knew would not provide results.

main issues that may arise beforehand (e.g.: a participant does not understand the task) and to devise 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 repeating what they have heard last (or to listen to just one stimulus) or to what the fieldworker suggested as an example. It is therefore important to be careful to avoid priming wherever possible 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 overall, weaker short-term memory, lower attention span, *etc.* You should keep these factors in mind when designing your questionnaire, and specifically: use short questions, both for the short-term memory and to limit exposure time; make sure that your stimuli are fully accessible (the volume of audio stimuli must be loud enough; if written stimuli are chosen, the font size should be fairly large). It is a good idea to split a long questionnaire into two parts and test them separately.
- 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 equipment, make sure there is always a relative present, 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 company, and they especially appreciate the opportunity to speak to younger people from their home country. Having an hour-long recording of spontaneous speech

is great for your research, but may be problematic if you have scheduled another interview shortly after the current one.

4 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 relating to heritage Italo-Romance speakers in North and South America and Europe. While many of these tips and much of this information can be found in general manuals or fieldwork reports, some are specific to the Italo-Romance community.²⁵ Furthermore, we provided a description of the status of these varieties, many of which had not been documented since the 1960s. When we did have some documentation of previous stages of the languages, we compared that to what we found, and showed that the situation has changed considerably. With the exception of the northern Italian-speaking community, most Italo-Romance heritage varieties in America are close to extinction: for this reason, documenting these languages is now all the more important. While this paper only draws some practical conclusions, there is much more to the study of heritage Italo-Romance in the Americas and we hope that our remarks will be helpful to researchers willing to undertake similar investigations.

Abbreviations

PRS present SCL subject clitic SG singular

Acknowledgements

We would like to express our gratitude to the informants who participated in our study and all our contacts in the Americas and Belgium who helped us reaching the communities and supported the fieldworkers. We would also like to thank two anonymous reviewers for useful comments and suggestions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement: CoG 681959_MicroContact).

²⁵For other considerations specific to Italo-Romance varieties, leaving aside the different fieldwork setting with respect to our study, see Cornips & Poletto 2005.

References

- Aalberse, Suzanne, Ad Backus & Pieter Muysken. 2019. *Heritage languages: A language contact approach*. Amsterdam: John Benjamins.
- Aalberse, Suzanne & Pieter Muysken. 2013. Perspectives on Heritage Languages. Position paper for the Leeuwenhorst Workshop on Heritage Languages, January 23–25th 2013, 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.
- Andriani, Luigi, Roberta D’Alessandro, Alberto Frasson, Brechje van Osch, Luana Sorgini & Silvia Terenghi. 2020. *The ranking of internal and external factors in heritage language syntactic change*. Paper presented at the 43rd Generative Linguistics in the Old World (GLOW) conference, HU Berlin/online, April 2020. DOI: <https://osf.io/es53m/>.
- Bagna, Carla. 2011. America Latina. In Massimo Vedovelli (ed.), *Storia linguistica dell’emigrazione italiana nel mondo*, 305–357. Roma: Carocci.
- Belletti, Adriana. 2018. On a-marking of object topics in the Italian left periphery. In Roberto Petrosino, Pietro Cerrone & Harry van der Hulst (eds.), *From Sounds to Structures: Beyond the Veil of Maya*, 445–466. Berlin/Boston: De Gruyter. DOI: 10.1515/9781501506734-016.
- Benincà, Paola & Cecilia Poletto. 2004. Topic, Focus and V2: Defining the CP sublayers. In Luigi Rizzi (ed.), *The structure of CP and IP*, 52–75. Oxford/New York: Oxford University Press.
- 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*, vol. 2, 75–89. Amsterdam: John Benjamins.
- Bettoni, Camilla. 1991. Language shift and morphological attrition among second generation Italo-Australians. *Rivista di linguistica* 3(2). 369–387.
- Bossong, Georg. 1985. *Differentielle Objektmarkierung in den Neuiranischen Sprachen*. Tübingen: Narr Verlag.
- Bossong, Georg. 1991. Differential Object Marking in Romance and beyond. In Dieter Wanner & Douglas A. Kibbee (eds.), *New Analyses in Romance Linguistics: Selected papers from the XVIII Linguistic Symposium on Romance Languages, Urbana-Champaign (April 7–9 1988)*, 143–170. Amsterdam: John Benjamins.
- Brandi, Luciana & Patrizia Cordin. 1989. Two Italian dialects and the Null Subject Parameter. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 111–142. Dordrecht: Springer.

- Casalicchio, Jan & Alberto Frasson. 2019. Cambiamenti nell'uso dei soggetti clitici veneti: Il ruolo del contatto con l'italiano. In Annamaria Chilà & Alessandro De Angelis (eds.), *Capitoli di morfosintassi delle varietà romanze d'Italia: Teoria e dati empirici*, 193–258. Palermo: Centro di studi filologici e linguistici siciliani.
- Cavaoli, 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.
- Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques. Part I. *Lingua* 115. 939–957.
- 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. 2021. Syntactic change in contact: Romance. *Annual Review of Linguistics* 7(1). 309–328. DOI: 10.1146/annurev-linguistics-011619-030311.
- 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*, 195–205. Cambridge: Cambridge University Press.
- Diez, Friedrich Christian. 1874. *Grammaire des langues romanes*. Vol. II, III. Paris: Franck.
- Ferrari, Andrea. 2008. *Aspetti socio-culturali dell'emigrazione italiana in Argentina: Il caso di Santa Fe*. University of Turin. (MA thesis).
- Frasson, Alberto. In press. Clitics are not enough: On agreement and null subjects in Brazilian Venetan. *Linguistic Analysis* 43(1–2).
- Frasson, Alberto, Roberta D'Alessandro & Brechje van Osch. In press. Subject clitics in microcontact. A case study from heritage Friulian in Argentina and Brazil. *Heritage Language Journal*.
- Guardiano, Cristina. 2000. Note sull'oggetto diretto preposizionale nel siciliano. *L'Italia Dialettale* 51. 1–35.
- Guardiano, Cristina. 2010. L'oggetto diretto preposizionale in siciliano. Una breve rassegna e qualche domanda. In Jacopo Garzonio (ed.), *Quaderni di lavoro ASIt (Studi sui dialetti della Sicilia)*, vol. 11, 83–101.
- 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*, 401–411. London: Routledge.
- Haller, Hermann W. 1997b. Italian. In Joshua A. Fishman & Ofelia García (eds.), *The Multilingual Apple. Languages in New York City*, 119–142. Berlin/New York: Mouton De Gruyter.
- 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*, 1083–1086. Turin: UTET.
- Ledgeway, Adam. 2009. *Grammatica diacronica del napoletano*. Tübingen: Max Niemeyer Verlag.
- Ledgeway, Adam. 2015. Varieties in Italy. In Konstanze Jungbluth & Federica Da Milano (eds.), *Manual of Deixis in Romance Languages*, 146–162. Berlin: De Gruyter.
- Ledgeway, Adam, Norma Schifano & Giuseppina Silvestri. 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 & John Charles Smith. 2016. Deixis. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance Languages*, 879–896. Oxford: Oxford University Press.
- Leivada, Evelina, Roberta D'Alessandro & Kleanthes Grohmann. 2019. Eliciting big data from small, young, or non-standard languages: 10 experimental challenges. *Frontiers in Psychology* 10. 313. DOI: 10.3389/fpsyg.2019.00313.
- Leonetti, Manuel. 2004. Specificity and Differential Object Marking in Spanish. *Catalan Journal of Linguistics* 3. 75–114.
- Manzini, Maria Rita & Leonardo Maria Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
- Maurizio, Roxana. 2008. Migración y desarrollo: El caso de Argentina. In Andrés Solimano (ed.), *Migraciones internacionales en América Latina. Booms, crisis y desarrollo*, 75–184. Fondo de Cultura Económica.
- Meo Zilio, Giovanni. 1990. *Estudios hispanoamericanos*. Vol. 1: Temas lingüísticos. Roma: Bulzoni.
- Meyer-Lübke, Wilhelm. 1895–1900. *Grammaire des langues romanes*. Paris: Welter.
- Meyer-Lübke, Wilhelm. 1890–1902. *Grammatik der romanischen Sprachen*. Leipzig: Fues.

2 Documenting Italo-Romance minority languages in the Americas

- Moravcsik, Edith A. 1978. On the case marking of objects. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 4. Syntax, 249–289. Stanford: Stanford University Press.
- O’Grady, William, Amy J. Schafer, Jawee Perla, On-Soon Lee & Julia Wieting. 2009. A psycholinguistic tool for the assessment of language loss: The HALA Project. *Language Documentation & Conservation* 3. 100–112.
- Pascual y Cabo, Diego. 2013. *Agreement reflexes of emerging optionality in heritage speaker Spanish*. University of Florida. (Doctoral dissertation).
- 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: Oxford University Press.
- Polinsky, Maria. 2018. *Heritage Languages and their Speakers*. Cambridge: Cambridge University Press.
- 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*, 149–167. Berlin: De Gruyter.
- Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In Osvaldo Jaeggli & Carmen Silva-Corvalán (eds.), *Studies in Romance Linguistics*, 391–420. Dordrecht: Foris.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar. Handbook in generative syntax*, 281–337. Dordrecht: Springer.
- Saab, Andres. 2018. Case conflicts and A-movement: DOM, doubling and clitics in Rioplatense Spanish. Lecture series held at Utrecht University (December 2018).
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert W. M. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Sorgini, Luana. 2020. *Differential object marking in heritage Romance varieties*. Paper presented at the 1st Heritage Language Syntax workshop, Utrecht/online, 12–16 October 2020. DOI: <https://osf.io/g7k2n/>.
- Stadthagen-González, Hans, Luis López, M. Carmen Parafita Couto & 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: 10.1075/lab.16030.sta.

Andriani et al.

Terenghi, Silvia. 2020. *Indexicality is not affected by contact. Ternary demonstrative systems in Italo-Romance HLs*. Paper presented at the 1st Heritage Language Syntax workshop, Utrecht/online, 12–16 October 2020. DOI: [10.17605/osf.io/5tusx](https://doi.org/10.17605/osf.io/5tusx).

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