

# The ‘HELLO Campania’ project and dataset: documenting language practices of linguistic minorities in Campania, Italy

Francesca R. Moro – University of Naples L’Orientale

[francesca.moro@unior.it](mailto:francesca.moro@unior.it)

Margherita Di Salvo – University of Naples Federico II

[margherita.disalvo@unina.it](mailto:margherita.disalvo@unina.it)

## ABSTRACT

HELLO Campania (Heritage Languages and the Languages of the Others in Campania) is a project focused on immigrant minority languages and their speakers in the Campania region of Italy. In this paper, we present the scientific background of the project and highlight its innovative aspects. We also describe the methodology used for data collection, and we provide an overview of the dataset and the structure of the HELLO Campania Collection, which has been archived and is available for consultation in the Eurac Research CLARIN Centre repository. The dataset consists of sociolinguistic interviews, language portraits, and linguistic tasks collected from speakers belonging to six communities: Bangladesh, Ghana, the Philippines, Senegal, Sri Lanka, and Ukraine. The aim of this paper is twofold. First, it serves as a reference point for users interested in the HELLO Campania Collection, offering insights into the project’s objectives and methodological framework. Second, it contributes to a broader reflection on methodological approaches in research on multilingualism and heritage languages, with particular attention to the advantages and limitations of sociolinguistic techniques, such as questionnaires, and elicited data, such as linguistic tasks.

**Keywords:** *immigrant language, heritage language, sociolinguistics, archive, Bangladesh, Ghana, Philippines, Senegal, Sri Lanka, Ukraine*

**DOI:** 10.23814/ethn.22.26.mor-dis

## 1. Introduction: HELLO Campania

HELLO Campania (Heritage Languages and the Languages of the Others in Campania) is a project about immigrant minority languages and their speakers in Campania (<https://sites.google.com/view/prinpnrrhelllocampania/home-page>). This research has been financially supported by the Italian Ministry of University and Research (MUR) as a Research Project of Significant National Interest in line with the objectives outlined in the National Recovery and Resilience Plan<sup>1</sup> (PRIN PNRR 2022, P2022WJ8YF).

The project has three primary objectives: (i) to conduct the first sociolinguistic study of immigrant communities in Campania; (ii) to examine migrants’ language practices and linguistic choices, exploring how multilingual speakers strategically use their linguistic resources and how these practices correlate with external sociolinguistic variables such as gender, educational background, ethnic origin, type of social network, and degree of segregation; and (iii) to provide the first combined qualitative and quantitative description of a heritage language in Italy, focusing on a single case study—the Filipino community in Campania—in order to analyze processes of contact-induced language change.

<sup>1</sup> This plan is part of the Next Generation Eu Program ([www.next-generation-eu.europa.eu/index\\_en](http://www.next-generation-eu.europa.eu/index_en)).

The research of the HELLO Campania project is based at the University of Naples L’Orientale and at the University of Naples Federico II, both universities are located in Naples, Campania Region, Italy. The team is composed as follows:

- University of Naples L’Orientale: Francesca Romana Moro (PI), Gina Russo (postdoctoral researcher), Michelle Corotan (research assistant), Antonia Soriente, and Marta Maffia. This research unit focuses on the Filipino community and Tagalog<sup>2</sup> as a heritage language (objective iii).
- University of Naples Federico II: Margherita di Salvo (Associate PI), Violetta Cataldo (postdoctoral researcher), MariaPaola Noschese (research assistant). This research unit carries out sociolinguistic research on the other communities (Bangladeshi, Ghanaian, Senegalese, Sri Lankan, Ukrainian; objectives i and ii).

The aim of this paper is to present the scientific background of the HELLO Campania project, highlight its innovative features, and offer insights into its objectives and methodological framework. Using the HELLO Campania project as a case study, we critically reflect on strategies for collecting linguistic data from diverse migrant communities. This aim stems from the idea put forward by Austin (2013) regarding the need for metadocumentation, namely the sharing of the goals, processes, methods, and structures of language-documentation projects. This is precisely what we intend to do in this article: we argue that the FAIR principles—especially those of accessibility and reusability—should apply not only to project results but also to project objectives and methodologies.

This article does not discuss the scientific results of the project. Some findings have already been published in academic venues (see, for example, Moro & Russo 2024; Di Salvo 2025; Di Salvo & Di Massa 2025; Buonomo, Di Salvo & Maffia 2025; Cataldo 2025). Additional project results have been disseminated through conference presentations, which can be accessed on the project website under the *Events* and *Outcomes* sections: <https://sites.google.com/view/prinpnrrhelllocampania/home-page?authuser=0>.

Instead, our focus here is to provide an overview of the data collected and, most importantly, to present one of the project’s key outcomes: the HELLO Campania Collection. By describing this resource, we aim to make the materials accessible to linguists, scholars from other disciplines, and members of the communities involved. The HELLO Campania Collection is stored in the repository of the Eurac Research CLARIN Centre and can be accessed at <https://clarin.eurac.edu/repository/xmlui/handle/20.500.12084/94>.

Another additional outcome is the creation of the CLARIN Knowledge (K) Centre *Heritage Languages in Europe* (<https://k-centreherlan.unior.it/>). The K-Centre supports new scholars interested in exploring heritage languages by providing expert guidance, and to foster connections among those already engaged in this research. We offer resources, advice, and networking opportunities to enhance the understanding of heritage languages and their speakers.

In the sections that follow, we outline the scientific background of the project (§2), describe the methodology employed for data collection (§3), and present the HELLO Campania (§4) along with the structure of its archival collection (§5).

---

<sup>2</sup> The national language of the Philippines, ‘Filipino’, is largely based on Tagalog, so the two terms are sometimes used synonymously. In the project and in this paper, we prefer to use the term Tagalog, even though we refer to the national language, because this reflects the preferred usage of the participants.

## 2. Brief scientific background of the project

Immigration to Campania has grown exponentially in recent years (de Filippo & Strozza 2015; Buonomo et al. 2025), resulting in a higher proportion of foreigners in the region compared to the national average: between 2015 and 2016, foreigners accounted for 6.7% of the population, compared to the national average of 2.1%. Campania has experienced diverse waves of immigration, with different groups arriving over time (Buonomo & Strozza 2021). In the 1970s, communities of Eritreans, Filipinos, Cape Verdeans, and Sri Lankans began to settle in the region. During the 1980s, the first African street vendors—primarily Senegalese—became visible in Naples and its provincial markets (de Filippo et al. 2010). Since the 1990s, the migrant population has continued to grow, characterized both by the consolidation of established groups (Filipinos, Sri Lankans, Senegalese, and Moroccans) and by the arrival of new groups, particularly from Poland and Albania (see Table 1).

**Table 1:** The first 10 foreign nationalities in Naples (Istat, 2022)

Country of origin	Total population	% of foreign born population
Sri Lanka	14.291	26.70%
Ukraine	7.053	13.20%
China	4.412	8.30%
Pakistan	3.098	5.80%
Romania	2.251	4.20%
Bangladesh	1.960	3.70%
Philippines	1.708	3.20%
Nigeria	1.456	2.70%
Senegal	1.174	2.20%
Dominican Republic	1.000	1.90%
Other countries	15.036	28.10%
<b>Total</b>	<b>53.440</b>	<b>100%</b>

Although immigration to Campania is a well-established phenomenon, few studies have examined it from sociolinguistic and linguistic perspectives. Consequently, knowledge of the language use of immigrants and their descendants remains limited. The available linguistic research has primarily adopted an acquisitional perspective, focusing on the varieties of Italian spoken by migrants (De Meo & Maffia 2016; Giuliano et al. 2014; Maffia 2020, 2023; Maturi 2016, Milano 2020). The only sociolinguistic study to address immigration in Campania is that of Della Putta (2022), which examines the linguistic behavior of a group of ten Ukrainian women in Naples.

Building on this gap in the existing literature, the HELLO Campania project aims to advance research on immigrant minority languages in the region by examining the intergenerational transmission of heritage languages within six communities: Bangladeshi, Filipino, Ghanaian, Senegalese, Sri Lankan and Ukrainian.

The project adopts a novel bottom-up approach that draws on multiple subfields of linguistic research, including sociolinguistics, language acquisition, and language contact studies (see §3). Specifically, for the Filipino community, integrating sociolinguistic enquiry with a linguistic analysis of the Tagalog variety spoken by second-generation speakers will provide a better understanding of contact-induced change in diaspora

communities and of the linguistic practices of second-generation heritage speakers—a population that remains largely under-described within the Italian context.

## 2.1. Objectives and selected groups

To address objectives (i) and (ii) and respond to the research questions concerning the sociolinguistic dynamics of immigrant communities in Campania we selected a sample of six groups. These groups were chosen based on specific criteria related to their demographic size, migration history, and geographical origin. The aim was to include populations that the literature identifies as relatively resistant to language shift (e.g., Bangladesh, Pakistan) alongside others considered more prone to shift (e.g., Eastern European groups). This contrast is reflected, for example, in the differing linguistic behaviors of Ukrainian and Bangladeshi women: Ukrainian women tend to invest more in acquiring and using Italian, even within the family domain, whereas Bangladeshi women often prioritize the maintenance of their heritage language, using it exclusively with their children due to its strong association with religious and ethnic identity (Pozzi 2014; Chini & Andorno 2018).

The project was initially designed to collect data from Ukrainian, Romanian, Filipino, and Bangladeshi communities. However, it soon became clear that incorporating additional groups—particularly from Africa—would result in a more representative and comprehensive sample.

The Sri Lankan and Ukrainian communities were selected because they are the most numerous immigrant groups in Campania. The Filipino community was included both because it is one of the oldest migrant groups in the region and because its members speak Austronesian languages, an area of expertise of the project's PI. The Bangladeshi community was chosen due to its well-documented tendency to maintain heritage languages robustly in diaspora contexts. The Senegalese and Ghanaian groups, by contrast, represent more recent and numerically smaller migrant communities. Their inclusion allows for the analysis of a migration pattern particularly relevant for comparative purposes, especially when examined alongside groups with longer-established ties to the city of Naples.

Moreover, the selected groups are different in terms of gender composition and stage of settlement. Sri Lankan and Filipino migration, having a long-standing presence in the Neapolitan context, has shifted from predominantly labor-driven mobility to processes of family reunification and long-term settlement. This transition has resulted in the emergence of a sizeable second generation that plays an increasingly visible role in the city's sociolinguistic landscape (for the Sri Lankan community, see Guadagno 2022). In contrast, the more recently established Bangladeshi, Ghanaian, and Senegalese communities still reflect predominantly labor-oriented migration patterns, with inflows largely consisting of young adult males of working age. A distinct case is represented by Ukrainian migration, which—despite its long-standing presence in Italy—remains markedly gendered, with women accounting for approximately 82% of the population (Buonomo et al. 2025). As this flow continues to be shaped by circular and employment-related mobility rather than settlement dynamics, the Ukrainian second generation remains comparatively limited in size when contrasted with that of the Sri Lankan and Filipino communities.

In terms of settlement patterns, the residential distribution of the selected groups reflects broader sociohistorical dynamics. Ukrainian and Filipinos tend to reside in rich neighborhoods, in the eastern part of the city, where the use of the local dialect is relatively

limited (Dedalus 2023). Sri Lankans, by contrast, are concentrated in working-class districts that function as “strongholds of dialect maintenance” (De Blasi 2013). Bangladeshi and Senegalese are mainly located in economically deprived areas, although no recent data on language use in these neighborhoods are available. No reliable data currently exist on the residential distribution of Ghanaian migrants in Campania or Naples.

This diversity in migration histories, settlement patterns, length of residence (LOR) in Italy allows the research team to examine how these variables shapes both heritage-language transmission and processes of language contact.

To address objective (iii), the project investigates Tagalog, an Austronesian language, in contact with Italian as the dominant majority language—a language-pair combination that has not yet been systematically explored. Most studies on heritage languages to date have focused on contexts where English, Dutch, or German are the majority languages. A further innovation of the project is its systematic comparison of data from second-generation speakers with that of first-generation speakers, a design rarely adopted in Italy, where studies of immigrant varieties typically focus solely on first-generation data (see Perotto 2009 for Russian, Vietti 2010 for Peruvian Spanish, Cohal 2014 for Romanian; a notable exception is D’Anna 2017 for Arabic).

For the linguistic analysis of Tagalog as a heritage language, we focused on two specific domains known to be vulnerable in heritage language contexts, while adopting a data mining approach for the remainder of the dataset to identify patterns of variation and uncover innovations.

The two targeted domains are word order and voice. In transitive clauses, Tagalog exhibits flexible word order with a preference for verb-initial structures, whereas Italian, although not as rigid as English, generally favors verb-medial order. Based on this typological contrast, we expect heritage speakers to show a shift toward verb-medial order in transitive sentences, moving away from the verb-initial preference, a pattern already observed among English-Tagalog bilinguals in the United States (Ronquillo 2015; Tanaka et al. 2019). With regard to voice, this domain represents one of the most distinctive features of Philippine languages, making it a particularly compelling area for study from a language contact perspective. The Tagalog voice system differs significantly from that of Italian and English, featuring a symmetrical system with five voices: actor, patient, locative, benefactive, and conveyance. Notably, previous research suggests that voice tends to be more stable than word order in contact settings. For instance, Tanaka et al. (2019) found that while heritage speakers of Tagalog in the United States exhibited English influence on word order, the voice system remained largely unaffected. This could be due either to English-dominant speakers not being fully sensitive to voice distinctions, or to the Tagalog voice system being sufficiently distinct from English to resist contact-induced change. Investigating the Tagalog voice system in an Italian-dominant environment will offer important new insights into the stability and vulnerability of this feature in heritage language contexts. In addition, other areas that may prove fruitful for analysis from a language contact perspective, once the data are fully examined, include lexical semantics (Moro under review), adjective intensification (Umbal, p.c.), and give-constructions.

### **3. Methodology for data collection**

To collect data addressing the three main objectives of the project, participants from six different communities were interviewed (for participant details, see §4). Each interview involved three components: responding to a sociolinguistic questionnaire (see §3.2),

completing a language portrait (see §3.3), and, for the Filipino community only, describing two sets of visual stimuli (for materials, see §3.4). Given the need for different types of linguistic and sociolinguistic data aligned with the study’s aims, we adopted a common procedure across all groups—the sociolinguistic questionnaire and the language portrait—while administering additional specific linguistic tasks solely to Filipino participants, on whom the analysis of linguistic innovations and changes is focused.

All the interviews were audio-recorded and subsequently transcribed (for details of the workflow, see § 3.5). The interviews were conducted primarily in Italian, with a few exceptions: eight interviews were conducted in English with Sri Lankan participants; two interviews were conducted entirely in French with Senegalese participants; four interviews with Senegalese participants were conducted using both Italian and French; and the six homeland Filipino interviews were conducted entirely in Tagalog. For the linguistic tasks (administered only to the Filipino participants), we consistently asked participants to speak in Tagalog.<sup>3</sup>

The data from the sociolinguistic interviews and the language portraits were analyzed in relation to objectives (i) and (ii), i.e. to investigate language choices, language use and language attitudes and to assess whether and to what extent the segregation index and other sociolinguistic variables influence language use. The sociolinguistic questionnaire included more than 30 open-ended items (see Appendix), such as “Do you think that living in Italy could make your children forget their heritage languages and traditions?”. The questionnaire served as the basis for sociolinguistic interviews, which also incorporated the language portrait task. During this activity, participants were invited to comment on the languages they included, explain why they positioned them in specific parts of the body, and reflect on what those languages mean to them (see §3.2 and 3.3). The linguistic data elicited through the visual stimuli tasks were analyzed to investigate contact-induced changes in heritage Tagalog (objective iii). As much as possible, the research adopts a cross-generational family design, comparing the language and linguistic practices of second-generation heritage Tagalog speakers with those of first-generation migrants and with homeland speakers, who serve as the baseline for comparison (cf. Aalberse, Muysken & Backus 2019: 116).

Prior to the interview, participants were given a consent form, written in a language they knew and in terms they could fully understand (drafted in accordance with current European data protection regulations – see G.D.P.R.), explaining the general aim of the project and asking for their permission for their data to be recorded, archived and made available (once anonymized). The methods of data collection (including the questionnaire and the consent form) were previously approved by the Ethic Committee of University of Naples Federico II.

Participants were recruited using the “friend of a friend” approach, starting with a few key people in the community who had a lot of contacts, then we tried to balance the sample as much as possible in terms of gender. In the case of the Filipino community, Gina Russo benefited from the help of the Filipino Women’s Council, the Immigration Desk in Salerno (Sportello TAM TAM), and from the Church Santa Maria della Verità in Naples, where Masses are celebrated in Tagalog and which therefore served as a meeting place to approach some of the participants. For the homeland participants, Michelle

---

<sup>3</sup> One reviewer pointed out that conducting interviews in a fully multilingual mode—embracing all the languages spoken by participants—would have allowed us to capture translingual practices more comprehensively. We agree with this observation, however, such an approach would have required the involvement of interviewers from within each community. Given the two-year duration of the project, it was not feasible to hire multiple researchers to cover all language backgrounds.

Corotan interviewed members of her own social network during a short trip to the Philippines. For the other groups, the research unit of Federico II benefit from the collaboration with the Italian Language School La Scuola di Pace, from Caritas Migrantes, the Senegalese Association in Naples Senaso, and the Ukrainian School<sup>4</sup>.

The criterion for inclusion was that the participant identified with the ethnic group because they were either born and raised in the county of origin or their parents were. Following Nagy (2015: 314), for the Tagalog speakers we avoid proficiency or fluency tests as we wanted to describe the range of performance of all types of heritage speakers.

In this project, generation is defined in a sociolinguistic and acquisitional dimension. First generation participants are those who immigrated to Italy in adolescence or adulthood (after the age of 12). Having completed their primary and lower secondary education in the country of origin, they have the language of origin as their dominant language and Italian as a second language, so they can be considered late bilinguals. Second generation participants are those who have been born in Italy or immigrated in early childhood (before the age of 8).

### 3.1. Methodological considerations and complementary ethnographic insights

One reviewer rightly noted that our methodological approach could have been strengthened by incorporating participant observation, recordings of spontaneous interaction, or language-use diaries, as these methods would have aligned the study more closely with its stated objectives.

Unfortunately, this was not possible for the HELLO Campania project. As a two-year initiative, the project faced strict time constraints that prevented us from collecting ethnolinguistic data across all communities. We therefore had to rely primarily on sociolinguistic interviews. Nonetheless, two spin-off projects emerged from the original study, as will be discussed at the end of this section.

This comment also offers an opportunity to situate our methodological choices within a longstanding debate in sociolinguistics regarding the relative merits of quantitative and qualitative approaches, each of which entails distinct advantages and limitations.

Data collection techniques in linguistic research face the challenge of balancing two key criteria: ecological validity—the extent to which the data reflect real-life language use—and external validity—the extent to which the findings can be generalized beyond the study's sample or context.

Sociolinguistic interviews allow researchers to gather information from a relatively large number of speakers within a short period of time. In our project, this method made it possible to collect data from many participants—for example, 66 individuals in the Filipino case study—thus ensuring high external validity and increasing the generalizability of the findings.

However, because interviews rely on self-reported behavior, they primarily capture what speakers *believe* they do—or what they consider *appropriate* to do with language—rather than what they actually do in everyday practice. For this reason, sociolinguistic interviews are low in ecological validity: they do not document spontaneous linguistic

---

<sup>4</sup> We would like to thank all the people who helped us with the data collection. Among others, we would like to mention Charito Basa (Filipino Women's Council), Fra' Primi (Church Santa Maria della Verità, Naples), Corrado Maffia (Scuola di Pace), Giancamillo Trani (Caritas Migrantes), Pierre Preira (Senaso), Mishel Gayaan Fernando Christeguge. We also thank the Sri Lankan Consul in Naples, Avv. Capasso. For the Ukrainian community, the anthropologist Tamara Mykhaylak (University Federico II) gave us great support during the fieldwork.

behavior and take place in a structured, researcher-led context that may not fully reflect natural language use.

Conversely, participant observation and linguistic ethnography capture naturalistic language use, but they are not free from bias. Like all qualitative research, they are shaped by the researcher’s perspective, by the situations in which observation occurs, and by the interpretive decisions made throughout fieldwork. Moreover, ethnographic approaches present the following challenges: (i) **Limited generalizability:** ethnographic studies are deeply embedded in local contexts and typically involve small samples, making it difficult to generalize findings beyond the specific communities examined; (ii) **Problems of systematization and replicability:** ethnographic data—often derived from informal conversations, unstructured interviews, and participant observation—are difficult to systematize and even harder to replicate. This raises questions about whether findings reflect broader patterns or are shaped by the dynamics of researcher–participant interaction; (iii) **Practical constraints:** ethnography is time-intensive and highly localized. It requires sustained immersion in the field, which limits the number of cases that can be studied and poses challenges for replication and large-scale comparative work.

For these reasons, we opted for sociolinguistic interviews as the most feasible method for collecting comparable data across multiple communities within the project’s time and resource constraints. However, we appreciate the reviewer’s suggestion and fully agree that combining survey-based methods with an ethnographic approach represents an ideal research design.

In this regard, it is worth noting that two “spin-off” projects have emerged from the HELLO Campania initiative. One of these is the PhD research of Elsie Marie Or, entitled *An Ethnographic Study of Translingual Practices of Filipino Migrants in Italy*. In her doctoral project, Or investigates the language practices of Filipinos in Campania through an ethnolinguistic approach, collecting naturalistic data by attending community events, shared meals, and informal gatherings, and by engaging in everyday conversations with community members. Interestingly, the preliminary findings of Or’s research closely align with the results obtained through our sociolinguistic interviews. Moro and Russo (2024) demonstrated that Filipinos in Campania often mix Italian and English with Tagalog, and that the use and intergenerational transmission of Tagalog as a heritage language appear to be relatively limited. Or’s conversational data reveal similar patterns, further confirming the low prevalence of sustained Tagalog-only practices in everyday interaction (Or 2025).

The second “spin-off” project, entitled *Being a Woman between Language and Culture* and funded by the University of Naples L’Orientale (PI: Marta Maffia; researchers: Margherita Di Salvo and Rosa Gatti), examines the experiences of second-generation Sri Lankans who, according to the sociolinguistic data collected through the HELLO Campania project, display a notably high rate of heritage language maintenance. Through qualitative interviews and life stories, the project seeks to explore the complex interplay between gender, heritage language practices, and identity formation. Preliminary findings corroborate patterns already observed in the HELLO Campania data, particularly regarding the greater vulnerability experienced by young women (Di Salvo in press).

### **3.2. Sociolinguistic questionnaire**

The sociolinguistic questionnaire was prepared by Francesca Moro and Margherita Di Salvo on the basis of other questionnaires previously used by the authors (Di Salvo 2019, Moro & Russo 2024), by Nagy (2009), and by Turchetta & Vedovelli (2018). The

questionnaire aims at investigating the migrant biography (age, gender, level of education, ...), their migration to Italy (age at arrival, kind of job, ...), their ethnic and residential patterns, type of social network, language use, the composition of migrants’ repertoire and their sense of belonging.

The questionnaire is divided into three main sections.

The first section focuses on socio-biographical information (questions 1–17), the individual’s migration experience, processes of integration in Campania (questions 18–39), and relationships with the country of origin (questions 40–45). For second-generation migrants, an additional subsection collects biographical data about their parents (questions 46–56).

The second section investigates migrants’ language attitudes and self-reported competence in Standard Italian, Italian dialects, and their heritage language(s), as well as in any other languages they may know (questions 57–74).

The final section (questions 75–85) addresses patterns of language choice across a range of formal and informal domains (e.g., with family members, with parents, with children, with Italian and non-Italian friends). For the complete set of questions, please refer to the Appendix.

We decided to make our questionnaire publicly available because, although we are fully aware of its limitations—both as a survey-based technique and as a tool that can still be improved—we strongly believe in the FAIR principles (Findable, Accessible, Interoperable, Reusable). These principles guide not only our approach to data management (see § 5), but also our broader project and methodological design. Our methodology is far from perfect; however, by sharing it and critically reflecting on it, we hope that future projects and researchers will be able to use, adapt, and enhance it.

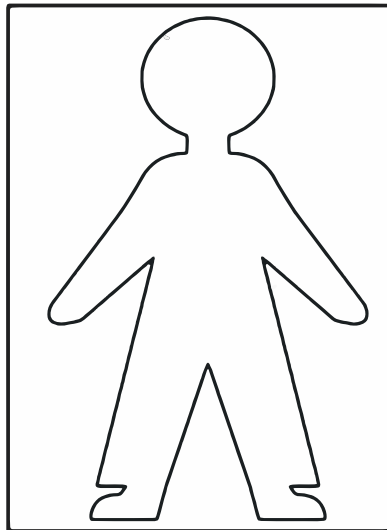
Through the questionnaire, the researchers collected spontaneous spoken texts (mainly in Italian, but in some cases also in English and French in the case of speakers with extremely low competence in Italian). The questionnaire was used to collect data with all speakers involved in the study.

Thanks to the data collected through the questionnaire, the researchers were able to test the speakers’ Italian language skills and assess their overall level of competence according to the Common European Framework of Reference for Languages. Furthermore, the data also allowed us to study the multilingual practices of first and second generation speakers. It was therefore particularly useful to collect not only self-reported evaluations but also spontaneously spoken texts.

### **3.3. Language portrait**

Language portraits are a tool for understanding language experiences, because they show how languages are connected to the body. Language portraits have been widely used in linguistic research, especially with participants whose L1 is not the dominant language of the country (Prasad 2014, Yoshida & Nichols 2024).

The procedure is as follows: after distributing the picture (see Figure 1), the activity is presented. The participants are asked to imagine that the picture represents their own body, then they are asked to think of all the languages and dialects in their lives (even if they know them very little) and to place them in their body (inside or outside) using pens, pencils and colors.



**Figure 1:** Language portrait

After the participant had completed the task, the researcher asked the participant for an oral explanation of the language portrait and recorded it. The data from the language portrait helped us to find converging (or diverging) evidence with respect to the answers to the questions about language attitudes that the participant later gave in the sociolinguistic interviews.

### **3.4. Frog story and videoclips**

To collect Tagalog language data, we used two elicitation stimuli: the picture book “Frog, where are you?” (Mayer 1969) and a list of video clips compiled specifically for the purposes of the project. We decided to use visual stimuli to elicit the data due to time constraints and to allow for comparisons between speakers and between groups. Communicative events recorded using visual stimuli (toys, picture books, movies) provide data that are less naturalistic than those collected using other methods (observation of real communicative events or communicative events staged for the linguist), but allow the researcher to control the vocabulary and grammatical features that occur in the communicative event (Himmelmann 1998:185-86). Thus, they provide a useful dataset for a first exploratory data mining research.

The picture book “Frog, where are you?” is a wordless picture book with 32 pages and 24 pictures that tells the story of a boy and a dog who are looking for a little frog that has gone missing. First they look for it in the boy’s house, then they go to the forest where they meet many animals and have many adventures. An example of two pages of the Frog story is shown in Figure 2.

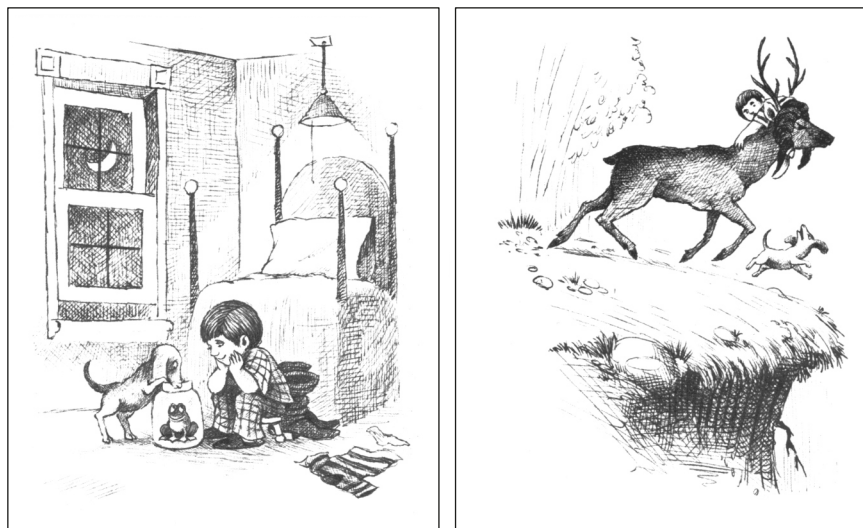






















Figure 2: Two pages of the story “Frog, where are you?” (Mayer 1969).

The Frog story has a long history in language research and has been used extensively in both language acquisition studies and language documentation. It has many advantages and some disadvantages, for a detailed discussion and examples see Moro & Klamer (2021). In terms of naturalness and spontaneity, the Frog story is considered to elicit data that may be more or less natural, depending on the culture, but definitely spontaneous. In our protocol, its production was quite spontaneous because the speakers told it on the spot without much time to prepare it. For the Filipino participants the use of pictures as a source of story telling is culturally common for those who grew up in cities or whose families have a high educational level. In villages of mountainous provinces, books are not present in the houses and are not used as a source of story telling. All in all, the Frog story allowed us to collect data that is relatively natural and spontaneous, but at the same time can be used for cross-speaker comparisons.

The list of video clips was compiled by Francesca Moro and Gina Russo from existing elicitation material specifically designed for linguistic research. The list contains 21 visual items: 14 set of pictures and 7 short videos. The 14 picture sets consist of a context picture and a target picture. The context picture introduces one or both of the characters in the target picture, while the target picture shows an action that the characters are performing.

Of the 14 sets, ten were designed by Bondoc et al. (2018) to elicit voice and word order in Tagalog (see Table 2), while four sets are distractors included by us. The symmetrical voice is a peculiarity of Philippine-type languages, like Tagalog, so it is worth studying it in relation to a language such as Italian, which does not have this feature (see § 2).

**Table 2:** Ten items designed by Bondoc et al. (2018) to elicit voice and word order in Tagalog.

	Patient-voice targeted (5 items)		Actor-voice targeted (5 items)	
Agent	animate, definite		animate, definite	
Patient	animate, definite		inanimate, indefinite	
	Items:		Items:	
	Context	Target	Context	Target
1.				
2.				
3.				
4.				
5.				

In Bondoc et al. (2018) items, animacy and referentiality are manipulated to trigger the production of the actor voice vs. patient voice in Tagalog: five items have animate agents and animate patients, while five items have animate agents and inanimate themes. In the former condition, the patient is highly prominent because is animate and definite (introduced in the context picture). This is a preferred context for patient voice. In the

latter condition, only the agent is animate and highly referential (introduced in the context picture), while the theme is inanimate and unseen previously. This is a preferred context for actor voice. The design is summarized in Table 2 above.

The procedure was as follows: the images were presented in a PowerPoint presentation, with participants describing one image at a time. The speakers were expected to say something along the following lines:





- 1) only one girl: *May isang babae* (There is a girl);
- 2) only one boy: *May isang lalake* (There is a boy);
- 3) a girl on the left, a boy on the right: *May isang babae at isang lalake* (There is a girl and a boy);
- 4) a boy on the left, a girl on the right: *May isang lalake at isang babae* (There is a boy and a girl).

After describing the context, the interviewer showed the target picture and asked in Italian or Tagalog the equivalent for “What is happening here?”.

The Bondoc et al. (2018) items were chosen because they have been used in other studies on Tagalog heritage speakers (Tanaka et al. 2019), which found English influence on word order but not on voice. In our view, using the same methodology as Tanaka on a new group of Tagalog heritage speakers, those in Italy, could really provide new insights into the issue of vulnerability of linguistic domains in different language contact situations.

The seven short videos included in the video clip list consist of three video clips from the Cut & Break set of the Max Planck Institute, designed to elicit resultative events (Bohnenmeyer, Bowerman, & Brown 2001, clips available for download at <http://fieldmanuals.mpi.nl>), and four video clips from the Event and Position list, designed to elicit give events (Moro & Fricke 2020, clips available for download at <https://vici.marianklamer.org/media.html>, see Table 3). These were included because previous work by the PI Francesca Moro on other Austronesian languages had shown that resultative and give events are susceptible to crosslinguistic influence (Moro & Klamer 2015, Moro & Irizarri van Suchtelen 2017, Villerius, Moro & Klamer 2019, Moro & Fricke 2020).

**Table 3:** Video clips used for Tagalog data elicitation

Target event	Screenshot
Give event (giving flower)	
Give event (giving banana)	
Give event (giving pen)	
Give event (fetching and giving a cup)	
Resultative event (snapping of a wooden stick)	
Resultative event (tearing up of a piece of cloth)	
Resultative event (slicing of a carrot)	

The 21 visual items (10 items by Bondoc et al., 7 videos and 4 distractor items) acted as distractors for each other and were presented in four randomized orders. In two orders, the Bondoc et al. (2018) items were flipped horizontally so that the first character in the context picture changed from a girl to a boy and vice versa.

### 3.5. Recording, data handling and workflow

The sociolinguistic interviews and linguistic tasks have been conducted and recorded in various locations, either in the participants' homes, in the homes of friends or relatives, at the Italian Language School *La Scuola di Pace*, in the church of Santa Maria della Verità in Naples, at the Senegalese Association in Naples or in public places such as ethnic shops. Data from the Filipino group were collected using a TASCAM DR-22WL portable recorder mounted on a Manfrotto tripod. The data from the Bangladeshi, Ghanian, Senegalese, Sri Lankan, and Ukrainian groups were recorded using a ZOOM H4nPro portable recorder.

The procedure was as follows: at the beginning of the session, the general aim of the project was explained and the consent form was read out. After the participant agreed to sign the consent form, the researcher started the recording. The first task was the language portrait, which was a simple task with the dual purpose of breaking the ice and getting the person to focus on their language repertoire in preparation for the sociolinguistic interview. Once the language portrait task was completed, the sociolinguistic interview began with the participant explaining the language portrait and then the researcher began asking the questions of the sociolinguistic questionnaire. For the Filipino group only, the sociolinguistic interviews were followed by the elicitation of the Frog Story narrative and the video clips. Before beginning the recording, the researcher showed the Frog story to the speaker, by giving an example in Italian of how to tell the story. The participant was then recorded retelling the story while leafing through the book. Once the Frog story had been recorded, the speaker was asked to sit in front of a laptop in order to watch and describe the videoclips on a PowerPoint presentation.

After the recording session, the data was stored on an external password-protected hard drive and anonymized by inserting a beep sound to cover sensitive information such as personal names, addresses, etc. The wav file was then converted to flac format for subsequent archiving. The metadata file was compiled and the recording was renamed using the following convention, which is adopted in the dataset: the first two letters stand for the group (BA: Bangladesh, FI: Filipino, GH: Ghana, SE: Senegal, SL: Sri Lanka, UK: Ukraine), the following number indicates the generation (1 for the first, 2 for the second, 0 for the homeland), then the gender (F for female, M for male), then a progressive number; for example the speaker SL2F05 is a Sri Lankan, second generation, female participant (number 5). For the Filipino group the file name also includes a specification of its content, whether it is a sociolinguistic interview, the Frog story or the elicited videoclips.

The language portrait was scanned as soon as possible after the recording, renamed according to the same system which was used when renaming the recordings, filed together with its recording.

The sociolinguistic as well as the linguistic data were transcribed with the software ELAN<sup>5</sup>, an open source software which allows to annotate video or audio recordings.

---

<sup>5</sup> ELAN (Version 6.8) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>.

Transcription in Italian/dialect were based on CLIPS norms (Savy 2006). Tagalog language data were also translated into English.

#### 4. The dataset

The HELLO Campania dataset contains data from a total of 182 participants, distributed as follows: 66 Filipinos, 48 Sri Lankans, 29 Ukrainians, 16 Senegalese, 12 Ghanaians and 11 Bangladeshis. Table 4 shows the total number of participants and their distribution by group and generation.

**Table 4:** Dataset of all participants included in the HELLO Campania dataset, divided by group and generation

<b>Groups</b>	<b>First generation</b>	<b>Second generation</b>	<b>Homeland speakers</b>	<b>Total</b>
Filipino	32	28	6	66
Sri Lankan	30	18	-	48
Ukrainian	28	1	-	29
Senegalese	16	-	-	16
Ghanaian	4	8	-	12
Bangladeshi	11	-	-	11
<b>Total</b>				<b>182</b>

The dataset consists of 309 audio recordings<sup>6</sup> for a total of 98 (h)ours 30 (m)inutes and 42 (s)econds. Table 5 summarizes the total hours of recording and the language portraits by group.

**Table 5:** Hours of recording and portraits divided by group

<b>Groups</b>	<b>Participants</b>	<b>Hours of recording</b>	<b>Language portraits</b>
Filipino	66	53h 53min 20s	40
Sri Lankan	48	20h 38min 49s	18
Ukrainian	29	11h 50min 43s	15
Senegalese	16	3h 47min 54s	0
Ghanian	12	3h 24min 65s	8
Bangladeshi	11	4h 55min 01s	9
<b>Total</b>	<b>182</b>	<b>98h 30min 42s</b>	<b>90</b>

The dataset consists of 75h 11m 32s of spoken Italian and dialect, 9h 28min of spoken English/French. For the Filipino, the dataset includes a total of 13h 51min 9s of spoken Tagalog (the linguistic tasks from all the groups, plus the sociolinguistic interviews from the homeland speakers only). A summary of the hours of recording for the Filipino group divided by task is given in Table 6.

<sup>6</sup> Note that the number of audio recordings for the Filipino group is not 198 (66 x 3 tasks) as expected, but 193. This is due to the fact that in one case the sociolinguistic interview was not recorded but completed in writing, and in four cases the second generation heritage speakers were unable to complete the linguistic task in Tagalog.

**Table 6:** Hours of recording for the Filipino group divided by task

Filipino group	Sociolinguistic Interview	Linguistic tasks		Tot linguistic tasks
		Videoclip	Frog Story	
<b>G1 (N=32)</b>	22h 44min 5s	3h 5min 28s	3h 20min 9s	6h 25min 37s
<b>G2 (N=28)</b>	17h 18min 6s	3h 1min 50s	2h 14min 30s	5h 16min 20s
<b>Homeland (N=6)</b>	1h 6min 6s	39 min 28s	23min 38s	1h 3min 6s

The following subsections describe the datasets of the participants for each group and provide some more detail on their biographical information.

#### 4.1. The Filipino group

For the Filipino group, the HELLO Campania dataset includes data from two sources: 26 speakers are from the “Heritage Tagalog in Italy” project (2021-2022, University of Insubria); 40 speakers are from the HELLO Campania project (2023-2025, University L’Orientale). The first project served as a kind of pilot study for the HELLO Campania project, allowing us to make a first comparison between generations (Moro & Russo 2024) and to test the methodology that we then decided to extend to other groups (of different nationalities) in the second project. Full continuity between the two projects is ensured by the presence of the same PI (Francesca Moro), the same postdoc as data collector (Gina Russo) and the same research assistant for transcriptions and translations (Michelle Corotan), and the same methodology for data collection in both projects. Furthermore, the data were collected in the Campania region: in Salerno for the “Heritage Tagalog in Italy” project, and in Naples for the HELLO Campania project. For this reason, we decided to subsume all the data of the Filipino group under the HELLO Campania Collection.

The Filipino data were collected between December 2022 and May 2024. A summary of the Filipino participants is presented in Table 7.

**Table 7:** Dataset of the Filipino participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	21	11	45.3	19.3 years	32
<b>Second generation</b>	14	14	20.6	-	28
<b>Homeland</b>	3	3	40.5	-	6

The sample of first generation speakers includes 21 females and 11 males (tot=32), with an average age of 45.3 years. The first generation speakers are classified as such because they all arrived in Italy after the onset of puberty; three speakers arrived between the ages of 12 and 14, while the remaining 29 all arrived after the age of 18. Their average length of residence (LOR) in Italy is 19.3 years. The educational level of the first generation of speakers is quite high, with 16 having attended or completed university, 11 having attended or completed high school, four having attended only middle school, and for one the educational level is unknown. They are mostly employed as domestic workers (19), or in restaurants or shops (cook or waitress) (5), one is a priest, one is a singing teacher, one runs a Filipino grocery store, one works in business administration, one is a B&B

owner, one is a condominium janitor, and two are not working or looking for a job. In terms of geographical origin, 25 first generation participants are from Luzon (16 from Central Luzon, five from Southeast Luzon, four from Northern Luzon), three from the Visayas, two from Mindanao and two from Mindoro.

The sample of second generation speakers consists of 14 women and 14 men (tot=28), with an average age of 20.6 years. They were either born in Italy (19) or arrived in Italy before the age of 10 (four arrived between the ages of 3 and 4, five between the ages of 8 and 10). Of the second generation participants, 11 have completed or are currently studying at university level, 12 have completed high school and five have completed primary or secondary school. Given the younger age of the second generation, 15 are still studying and only 13 have a job: five work in restaurants or shops (cook or waitress), two are babysitters and after-school teachers, one is a lawyer, one is a programmer, one is a hostess, one works in a Filipino grocery store, one works in a B&B, and one is a secretary.

We aimed for a cross-generational family design to make the dataset homogeneous and collect data from the same households. Many family and personal ties exist between our participants. There are eight parent-child relationships between first and second generation participants, two parent-child relationships between first generation participants, four couples, two sibling relationships, one grandparent-nephew relationship and one cousin relationship.

The sample of homeland speakers consists of three women and three men (tot=6), with an average age of 40.5 years, which is comparable to that of the first generation speakers. In terms of education, five have completed university and one has completed secondary school. Two are practitioners, two are teachers, one is a school caretaker and one is retired. We tried to match the geographical origin of homeland speakers to those of first generations speakers: three homeland participants come from Laguna province (Luzon island), two come from Mindanao province (Mindanao island) and one from Cebu province (Cebu island).

#### 4.2. The Sri Lankan group

The sociolinguistic interviews and language portraits of 48 Sri Lankan participants were collected between January and September 2024 by Margherita Di Salvo, Maria Paola Noschese, Violetta Cataldo and Marta Maffia. A summary of the Sri Lankan participants is given in Table 8.

**Table 8:** Dataset of the Sri Lankan participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	15	15	33.3	10.3	30
<b>Second generation</b>	5	13	19.2	-	18

The first generation speaker group includes 15 females and 15 males (tot = 30), with an average age of 33.3 years. As for the Filipino migrants, first generation includes all those migrants who arrived in Italy after the onset of puberty: three speakers arrived between the ages of 12 and 16, the others arrived in Italy as adults. The mean length of residence (LOR) in Italy is 10.3 years.

The educational level of the first generation speakers is quite high, with 27 having attended or completed school and university; four who came to Italy as adolescents did

not finish school in their home country, but began their studies in Italian high schools or universities. They are mostly employed as domestic workers (15), or in restaurants or shops (cook or waitress) (4), one is a hair washer, two are employed in a Sri Lankan market, one is a teacher.

With regard to the motivation for migration, female first-generation speakers arrived in Italy for family reasons (12), to seek employment (2) or to pursue studies in Italy (1). In contrast, male participants arrived in Italy for employment (8), for family reunification (2) or for the purpose of studying or residing abroad (3). Two speakers did not report any predominant reason for their migration.

The second generation comprises 5 females and 13 males (tot = 18). For the Sri Lankan group, it is important to note that these speakers may not always fit into the prototypical definition of the second generation, as many of them spent their secondary socialization years (generally from age 3 to 8-9) in Sri Lanka at the behest of their parents. In fact, members of the first generation often want their children to be educated in Sri Lanka. As a result, many young people born in Naples spent their childhood in Sri Lanka, which involved linguistic immersion in a predominantly Sinhalese environment. This is likely to have had an impact on their level of proficiency in both Sinhalese and Italian.

Of the second-generation participants, ten have either completed or are currently pursuing studies at the high school or university level. Eight are employed or engaged in job searches, with a range of occupations including dance teacher (one), trade unionist (one), receptionist (one), cultural mediator (one), and housekeeper (one).

Familial connections are observed between a limited number of participants. These include parent-child relationships between first and second generation participants (two cases) and between a wife and her husband (one case). In numerous instances, participants are friends or even cohabitants devoid of familial ties (three cases).

### 4.3. The Ukrainian group

The Ukrainian sociolinguistic interviews and language portraits were collected from November to December 2024 by Margherita Di Salvo, Maria Paola Noschese and Violetta Cataldo. A summary of the Ukrainian participants is given in Table 9.

Table 9: Dataset of the Ukrainian participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	19	9	43.7	14.7	28
<b>Second generation</b>	0	1	21	-	1

The first generation speaker group includes 28 participants, 9 males and 19 females, with a mean age of 43.7 years. The mean age of the Ukrainians participants is similar to that of the Filipinos, but it is higher than that of the Sri Lankans: this is due to the fact that Ukrainian migration is predominantly made up of women who emigrate alone at an older age and exclusively for working reasons (see also Dedalus 2023: 62). Transnational families are frequent in this group.

As for the age of arrival of first generation participants: five speakers arrived between the ages of 12 and 17, and the others as adults. The mean LOR in Italy is 14.7 years. It is necessary to distinguish between the migration wave that preceded 2022 and the subsequent wave. The former wave was composed of women and men driven by

economic motivations, whereas the latter, which emerged after February 2022, was predominantly made of women and children seeking to escape war.

The educational level of first-generation speakers is notably high, with six individuals having attended or completed university, while 11 participants attended high school. The majority of women are employed in domestic work, with one individual engaged in the role of cultural mediator, another pursuing a career as a musician, and a third working as an educator. The male participants are primarily engaged in occupations related to construction and ethnic commerce.

A noteworthy feature of the Ukrainian group is the prevalence of mixed marriages, observed in four cases within our sample. These unions are exclusively formed between a Ukrainian woman and a Neapolitan partner.

The second generation comprises a single male participant, aged 17 years. He was born in Italy and is currently employed at the Ukrainian shop owned by his parents. He completed three years of secondary education but did not proceed to the next stage. The demographic profile of the community is characterized by a significant imbalance in favor of the female component, coupled with a relatively limited number of family units. This gives rise to an uneven composition of the community which is heavily skewed in favor of the female component.

#### 4.4. The Senegalese group

Data collection took place between February and June 2024. Interviews were collected by a BA student Antonella Alborino and Margherita Di Salvo (Alborino 2024). The Senegalese sample is made up of 16 first-generation participants, seven women and nine men. A summary of the Senegalese participants is presented in Table 10.

**Table 10:** Dataset of the Senegalese participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	7	9	36.8	7.6 years	16

The Senegalese participants in our study are primarily concentrated in the city center of Naples, particularly in the area near the central station, which is characterized by a high density of foreign residents, especially from African countries, Bangladesh, and Pakistan. Their average length of residence (LOR) in Italy is 7.6 years, although this figure masks considerable variation: four participants have been in Italy for more than 10 years, five have arrived within the past year, while the other participants have resided in Italy for between 2 and 10 years. This variation in residence length also affects language choice during interviews, with recent arrivals often using French.

Seven participants reported having lived for over a year in locations other than Naples prior to settling there; of these, four had lived abroad—two in France and two in Spain. Only half of the Senegalese participants are married, and notably, in two cases, the spouse is Italian.

In terms of education, the Senegalese participants generally possess medium-to-high qualifications: ten completed high school, two obtained university degrees, and two attended Koranic school. Regarding employment, five participants are currently unemployed; among those working, six are employed in domestic or care-related occupations—such as housekeeper, colf, or seamstress, two have manual or maintenance

jobs (hotel maintenance and gardener), one individual is engaged in informal street vending combined with cultural mediation, one is employed in the security sector, and one is a company legal representative. The other participants are unemployed or have discontinuous employment.

All participants arrived in Italy as adults, with an average age of arrival of 29 years.

#### 4.5. The Ghanaian group

The Ghanaian group is made up of twelve participants, four of the first generation and eight of the second generation. Data collection took place from May to November 2024. Data collection was mainly carried out by a MA student Hannora Marlene Asienda (under the supervision of Marta Maffia). Other interviews were conducted by Antonella Alborino, Marta Maffia, and Maria Paola Noschese. A summary of the Ghanaian participants is presented in Table 11.

**Table 11:** Dataset of the Ghanaian participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	1	3	45	24.3	4
<b>Second generation</b>	3	5	22.5	-	8

First generation migrants came to Naples for work reasons: men were the first to arrive and were followed by women. Systematic studies on this group are lacking: however, our interviews show that there are marriages, mainly endogamic, and that there are second-generation speakers, including adolescents/young adults. Their average length of residence (LOR) in Italy is 24.3 years, similar to the Filipinos. The absence of demographic studies on this group does not allow us to reconstruct its settlement pattern. However, based on our collected data, we can report that the migrants interviewed claim to reside either in central neighborhoods of Naples or in the Pompeii area. Among the first-generation Ghanaian participants, two have completed high school. One has graduated from university, and another is currently attending university in Naples. Three first-generation participants are employed: one works as a bricklayer, one as a receptionist, and one as a cleaning lady. As for the second-generation participants, most have completed high school. Three are attending university, and one is still in high school. They are also employed in various roles: one is a model, two are waiters, two are tour operators, and one is an athlete.

#### 4.6. Bangladeshi group

The group originating from Bangladesh was surveyed by Margherita Di Salvo and Maria Paola Noschese in April and May 2024. It reflects the main characteristics of Bangladeshi immigration in Naples and the Campania region (for an overview, see Buonomo et al. 2025). The sample consists of 11 young men, only some of whom (four) are married; however, all migrated to Italy alone. In these cases, the migrants left their wives in Bangladesh, with the intention of reuniting after obtaining a residence permit.

In the remaining cases, migration occurred individually and was not connected to established family networks. Given that Bangladeshi migration is predominantly labor-driven and that household formation remains limited (Buonomo et al. 2025), it was not

possible to include individuals related by family ties in the sample. Nonetheless, some participants (four) are connected through bonds of friendship and cohabitation.

**Table 12:** Dataset of the Bangladeshi participants with some basic biographical information

	Female	Male	Age (Mean)	LOR (Mean)	Tot
<b>First Generation</b>	0	11	27.1	1.6 years	4

The settlement pattern reflects a strong trend toward residential segregation in that all respondents reside in the area close to the station and in the poorer neighborhoods of the city center. All participants reported living in houses shared with compatriots. In addition, all have medium-to-high educational qualifications, with all having completed high school. Their length of residence (LOR) in Italy is relatively short, and their proficiency in Italian remains limited. This language barrier contributed to the frequent use of English during interviews (5 out of 11 cases).

Limited Italian language skills further contribute to the marginalization of Bangladeshis in the labor market, confining them to low-status professions with minimal interaction with the native population. Among the participants, one works as a waiter, another as a cook in a Neapolitan restaurant, while the rest divide their time between precarious jobs (mainly street vending, such as selling umbrellas) and other informal or illegal work.

## 5. HELLO Campania Collection at Eurac Research Clarin Centre

In November 2024, with the collaboration of Egon Stemle and Alexander König, the process of archiving the materials collected as part of the HELLO Campania project was initiated. The research team chose to deposit the collected data—including sociolinguistic interviews, language portraits, and linguistic tasks—at the Eurac Research CLARIN Centre (ERCC), a private non-profit research center based in Bolzano and hosted by the Institute for Applied Linguistics (IAL). The HELLO Campania Collection can be accessed at <https://clarin.eurac.edu/repository/xmlui/handle/20.500.12084/94>.

The ERCC is committed to the long-term preservation of the materials archived in its repository. By archiving the HELLO Campania data with the ERCC, the project aims not only to ensure its preservation but also to adhere to best practices in digital preservation and to promote open-access sharing. The ERCC is part of the Europe-wide CLARIN infrastructure, which ensures that (meta)data adhere to well-defined international standards and the FAIR Data Principles. This framework supports knowledge discovery, innovation, and the integration of data and knowledge, while promoting data sharing and open access. The archived materials are intended to benefit a range of stakeholders—including scholars, educators, and researchers working on migrant integration and education, as well as members of the communities and ethnic groups interviewed in the project. The files are licensed under: [CLARIN ACADEMIC END-USER LICENCE \(ACA-BY-NC-NORED 1.0\)](#).

To build a user-friendly archive, we organized the HELLO Campania Collection into seven distinct Communities or subcollection (see Figure 3): one for each ethnic group included in the project (Philippines, Sri Lanka, Ukraine, Senegal, Ghana, Bangladesh), and one containing the transcripts of the collected materials.

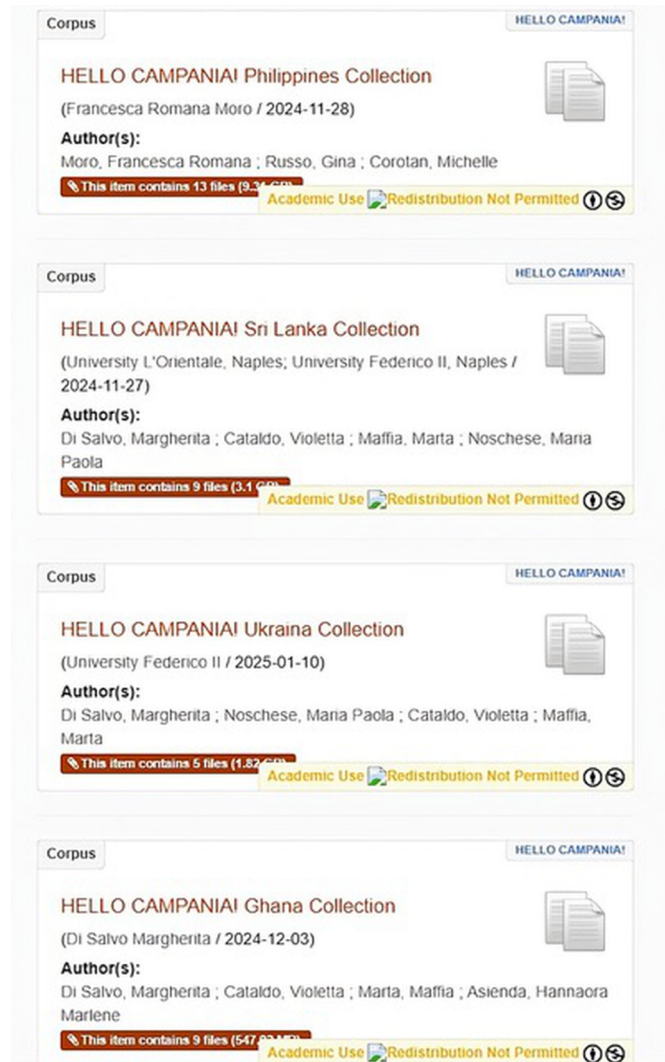
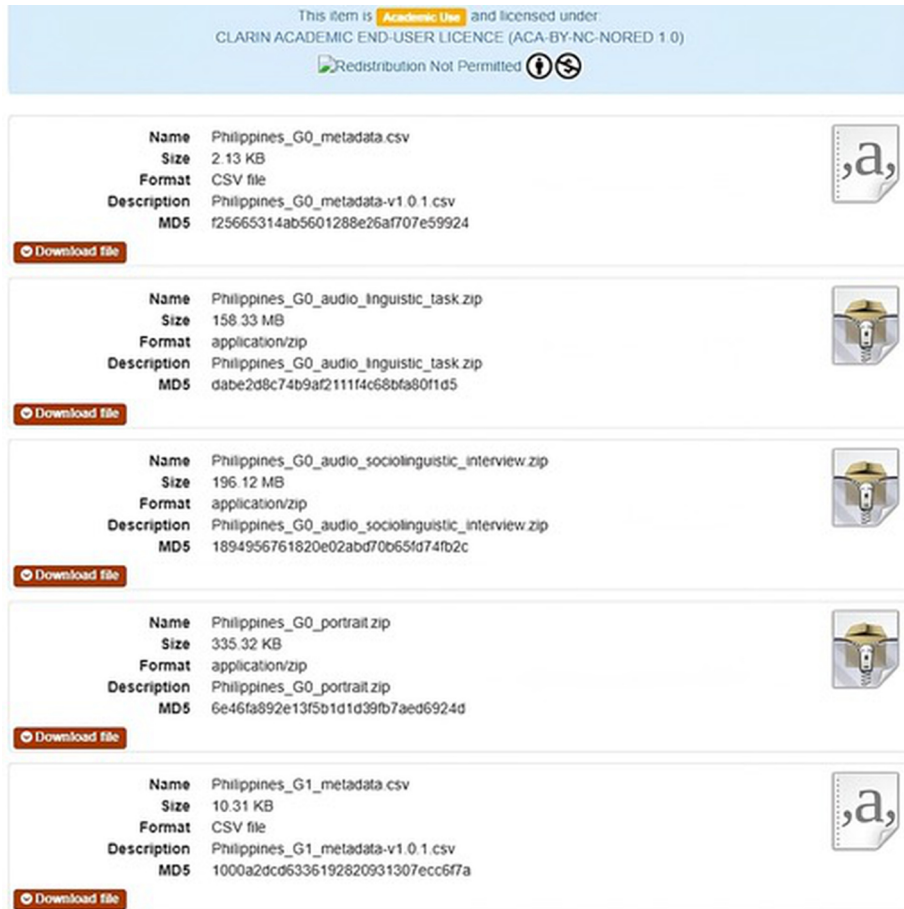


Figure 3: Communities of the HELLO Campania Collection archived at the EURAC Research CLARIN Centre

Each Community is further subdivided internally according to generation—G0 (homeland speakers), G1 (first-generation speakers), and G2 (second-generation speakers)—and by the type of recording: linguistic task, sociolinguistic interview, and language portrait, as shown in Figure 4 below. Furthermore, each folder contains a metadata csv file for each generation (e.g., Philippines\_G0\_metadata, Philippines\_G1\_metadata, and Philippines\_G2\_metadata) providing information about:

- generation, age, gender, educational level, occupation of the participant
- linguistic background of the participant (L1(s) and other languages known)
- type of recording (sociolinguistic interview, Frog story, description of videoclips)
- main language used in the recording
- date of the recording
- length of the recording
- who collected data and who transcribed them.



**Figure 4:** Example of the file organization in the “Philippines sub-collection” of the HELLO Campania Collection, archived at the EURAC Research CLARIN Centre.

All audio files and portraits are open access and fully anonymized. Transcripts are provided in .eaf format, fully anonymized, available upon request.

Each collection has its own item identifier (handle) and a corresponding citation format. For example:

“Di Salvo, Margherita; Cataldo, Violetta; Maffia, Marta; et al. 2024, *HELLO CAMPANIA! Sri Lanka Collection*, Eurac Research CLARIN Centre, <http://hdl.handle.net/20.500.12124/87>.”

The archived material from the HELLO Campania project provides a valuable resource for linguists interested in language use and linguistic choices, for those interested in the study of Italian as a second language and as a first language for heritage speakers from diverse linguistic backgrounds. The dataset includes not only extensive samples of Italian and Neapolitan dialect spoken by migrants and their descendants, but also a rich collection of Tagalog data, with recordings from both homeland speakers in the Philippines and heritage speakers in Italy. In addition, the sociolinguistic interviews provide detailed information on participants’ language acquisition trajectories, language use, and multilingual practices, offering insights into how heritage languages are maintained, transformed, or shift in a migration context. This wealth of material enables cross-linguistic and cross-generational comparisons, making it a significant resource for researchers working on second language acquisition, heritage language studies, sociolinguistics, and language contact phenomena.

## 6. Conclusions

The HELLO Campania project represents a significant contribution to the study of migration-induced multilingualism in Italy, providing the first systematic documentation of the linguistic practices of six immigrant communities in the Campania region. Through a carefully designed and multidimensional approach – combining sociolinguistic interviews, language portraits and linguistic tasks – the project provides a rich, open-access dataset that captures the complex interplay between heritage languages, Italian and Neapolitan. Focusing specifically on the Filipino community, the HELLO Campania project also offers new insights into contact-induced language change in heritage Tagalog in Italy. By using an intergenerational family design, the project allows for a direct comparison between first- and second-generation speakers as well as home baselines, providing a more nuanced understanding of language change and maintenance. Finally, by archiving the HELLO Campania dataset at the Eurac Research CLARIN Centre and by establishing the CLARIN K-Centre Heritage Languages in Europe, the project ensures the long-term preservation and accessibility of valuable linguistic resources for future research. We hope that the HELLO Collection and the K-Centre will serve not only linguists, but also educators, policymakers, and community members committed to promoting linguistic diversity, social inclusion, and cultural heritage in contemporary Italy.

## Acknowledgements

The research presented in this project—together with its outcomes and the writing of this article—was financially supported by the Italian National Project PRIN PNRR 2022 “*Heritage Languages and Languages of the Others (HELLO Campania)*” (P2022WJ8YF). We would like to thank all the people who helped us with the data collection. Among others, we would like to mention Charito Basa (Filipino Women’s Council), Fra’ Primi (Church Santa Maria della Verità, Naples), Corrado Maffia (Scuola di Pace), Giancamillo Trani (Caritas Migrantes), Pierre Preira (Senaso), Mishel Gayaan Fernando Christeguge. We also thank the Sri Lankan Consul in Naples, Avv. Capasso. For the Ukrainian community, the anthropologist Tamara Mykhaylak (University Federico II) gave us great support during the fieldwork. We express our sincere gratitude to an anonymous reviewer for their insightful feedback, and to Egon Stemle, Alexander König and the CLARIN network for their invaluable support in the archiving process. Finally, we would acknowledge the support of the CLARIN K-Centre *Heritage Languages in Europe* (<https://k-centreherlan.unior.it/>) in the writing of this article. All remaining errors, of course, are our own.

## References

- ALBORINO, A. (2024) *La comunità senegalese a Napoli: aspetti sociolinguistici* (Unpublished bachelor's thesis). Napoli: Università di Napoli Federico II.
- AUSTIN, P. K. (2013) "Language documentation and meta-documentation". In M. C. JONES & S. OGILVIE (Eds.), *Keeping languages alive: Documentation, pedagogy and revitalization*. (Pp. 3-15. Cambridge: Cambridge University Press.
- BENASSI, F., NACCARATO, A., IGLESIAS-PASCUAL, R., SALVATI, L., & STROZZA, S. (2023) "Measuring residential segregation in multi-ethnic and unequal European cities". *International Migration*, 61: 341-361.
- BOHNEMEYER, J., BOWERMAN, M., & BROWN, P. (2001) "Cut and break clips". In *Manual for the field season 2001*. Pp. 90-96. Nijmegen: Max Planck Institute for Psycholinguistics.
- BONDOC, I. P., W. O'GRADY, K. DEEN & N. TANAKA. (2018). "Agrammatism in Tagalog: voice and relativisation". *Aphasiology* 32(5), 598-617.
- BUONOMO, A., BENASSI, F., DE FILIPPO, E., & STROZZA, S. (2025) *Gli immigrati di Napoli e le Napoli degli immigrati*. Milano: FrancoAngeli.
- BUONOMO, A., DI SALVO, M., & MAFFIA, M. (2025) "Lingue in migrazione: Repertori, competenze e usi". In A. BUONOMO, F. BENASSI, E. DE FILIPPO, & S. STROZZA (Eds.), *Gli immigrati di Napoli e le Napoli degli immigrati*. Pp. 161-182. Milano: FrancoAngeli.
- CATALDO, V. (2025) "Phonetic features of Neapolitan and Campania Regional Italian in migrants' speech". *International Journal of Linguistics*, 17(5): 201-220.
- CHINI, M., & ANDORNO, C. (2018) *Repertori e usi linguistici nell'immigrazione: Un'indagine su minori alloggiati dieci anni dopo*. Milano: FrancoAngeli.
- COHAL, A. L. (2014) *Mutamenti nel romeno di immigrati in Italia*. Milano: FrancoAngeli.
- D'ANNA, L. (2017). *Italiano, siciliano e arabo in contatto: Profilo sociolinguistico della comunità tunisina di Mazara del Vallo*. Palermo: Centro di studi filologici e linguistici siciliani.
- DE BLASI, N. (2013) "Persistenze e variazione a Napoli (con una indagine sul campo)". *Italienisch*, 69(1): 75-92.
- Dedalus (2023) *Modelli insediativi e livelli di integrazione dei cittadini immigrati nella città di Napoli*. Rapporto di ricerca, Napoli.
- DE FILIPPO, E., & STROZZA, S. (2015) *Immigrazione e integrazione in Campania*. Milano: FrancoAngeli.
- DELLA PUTTA, P. (2022) "Neapolitan, Regional and Standard Italian in the linguistic repertoire of Ukrainian private carers in Naples: Sociolinguistic competence and attitudes towards a complex linguistic context". In F. GOGLIA & M. WOLNY (Eds.), *Italo-Romance dialects in the linguistic repertoires of immigrants in Italy*. Pp. 47-75. Palgrave Macmillan.
- DI SALVO, M. (2025) "Italiano regionale campano come L2 di srilankesi di prima e seconda generazione: Un quadro sociolinguistico". *Italiano LinguaDue*, 17(1): 614-631.

- DI SALVO, M., & DI MASSA, V. P. (2025) "Language variation and urban space: A sociolinguistic study of immigrants in Naples (Italy)". *International Journal of Linguistics*, 17(5), 136-152.
- DI SALVO, M. (in press) "Sri-Lankans in Naples: Language maintenance and family language policies across generations". *International Journal of Multilingualism*.
- GIULIANO, P., ANASTASIO, S., & RUSSO, R. (2014) "Passato remoto, passato prossimo e imperfetto: Uso biografico e fittizio delle forme al passato nelle interlingue di immigrati di area partenopea". In A. DE MEO, M. D'AGOSTINO, G. IANNACCARO, & L. SPREAFICO (Eds.), *Varietà dei contesti di apprendimento linguistico (Studi Aitla)*. Pp. 299-315. Milano: Officinaventuno.
- GUADAGNO, E. (2022) "Territori in movimento: La comunità srilankese nella spazialità napoletana". *Bollettino della Società Geografica Italiana*, 5(2): 95-108.
- MAFFIA, M., & DE MEO, A. (2015) "Literacy and prosody: The case of low-literate Senegalese learners of L2 Italian". In *Adult literacy, second language, and cognition: LESLLA Proceedings 2014*. Pp. 129–147. Centre for Language Studies.
- MAFFIA, M. (2020) "Apprendenti srilankesi di italiano L2: Appunti per un profilo sociolinguistico". *Italiano LinguaDue*, 12(2): 123-134.
- MAFFIA, M. (2023) *Analfabetismo, abilità orali e lingue seconde: Uno studio su senegalesi apprendenti di italiano L2 (Studi Aitla, 16)*. Milano: Officinaventuno.
- MATURI, P. (2016) "L'immersione in una realtà linguistica complessa: Gli immigrati tra i dialetti e l'italiano". In A. DE MEO (Ed.), *L'italiano per i nuovi italiani: Una lingua per la cittadinanza*. Pp. 123-128. Napoli: Università degli Studi di Napoli L'Orientale.
- MILANO, E. (2020) "Verso la fossilizzazione: I processi di acquisizione dell'italiano L2 di giovani migrati filippini a Napoli". In M. BORREGUERO ZULOAGA (Ed.), *Acquisizione e didattica dell'italiano: Riflessioni teoriche, nuovi apprendenti e uno sguardo al passato*. Pp. 269-286. Peter Lang.
- MORO, F. R., & IRIZARRI VAN SUCHTELEN, P. (2017) "Dominant language transfer in heritage languages". In H. PEUKERT & I. GOGOLIN (Eds.), *Dynamics of linguistic diversity*. Pp. 143-162. Amsterdam: John Benjamins.
- MORO, F. R., & FRICKE, H. (2020) "Contact-induced change in Alorese give-constructions". *Oceanic Linguistics*, 59(1–2): 116-147.
- MORO, F. R., & KLAMER, M. (2021) "What is "natural" speech? Comparing free narratives and Frog stories in Indonesia". *Wacana: Journal of the Humanities of Indonesia*, 14: 238-313.
- MORO, F. R., & RUSSO, G. (2024) "Family language policy in multilingual Filipino families in Italy". *Journal of Multilingual and Multicultural Development*, 46(9): 2965-2979.
- MORO, F. R. (under review) "Syntactic and lexical changes in trilingual heritage Tagalog speakers". *Critical Multilingualism Studies (Special issue: Language outcomes in multilingual heritage speakers: The role of the contact scenario)*.
- NAGY, N. (2009) *Heritage Language Variation and Change*. <http://projects.chass.utoronto.ca/ngn/HLVC>
- NAGY, N. (2015) "A sociolinguistic view of null subjects and VOT in Toronto heritage languages". *Lingua*, 164: 309-327.

- OR, E. M. (2025) “An ethnographic study of translingual practices of Filipino migrants in Italy”. Paper presented at the conference Heritage Languages in Italy, Naples, Italy, October 2–3, 2025. <https://sites.google.com/view/prinpnrrhellocampania/events?authuser=0>
- PEROTTO, M. (2009) *Lingua e identità dell’immigrazione russofona in Italia*. Napoli: Liguori Editore.
- POZZI, S. (2014) “Trasmissione delle lingue, integrazione e identità nelle famiglie immigrate”. In M. V. CALVI, I. BAJINI, & M. BONOMI (Eds.), *Lingue migranti e nuovi paesaggi*. Pp. 37-50. LED Edizioni.
- PRASAD, G. (2014) “Portraits of plurilingualism in a French international school in Toronto: Exploring the role of visual methods to access students’ representations of their linguistically diverse identities”. *Canadian Journal of Applied Linguistics*, 17(1): 51-77.
- SAVY, R. (2006) “Specifiche per la trascrizione ortografica annotata dei testi raccolti”. In F. ALBANO LEONI & R. GIORDANO (Eds.), *Italiano parlato: Analisi di un dialogo*. Pp. 1-37. Napoli: Liguori.
- TANAKA, N., BONDOC, I. P., & DEEN, K. (2022) “Examining main clause similarity and frequency effects in the production of Tagalog relative clauses”. *Journal of the Southeast Asian Linguistics Society*, 15(2), 70–86. <https://hdl.handle.net/10524/52506>
- TANAKA, N., O’GRADY, W., DEEN, K., KIM, C.-E., HATTORI, R., BONDOC, I. P. M., & SORIANO, J. U. (2016) “Relative clause elicited production task”. Nozomi Tanaka Collection (NT1-002), Kaipuleohone: <http://hdl.handle.net/10125/42193>.
- VIETTI, A. (2010) “Italiano e spagnolo a contatto: Immigrazione e varietà etnica”. In M. V. CALVI, G. MAPELLI, & M. BONOMI (Eds.), *Lingua, identità e migrazione: Prospettive interdisciplinari*. Pp. 221-235. Milano: FrancoAngeli.
- VILLERIÚS, L., MORO, F. R., & KLAMER, M. (2019) “Encoding transfer events in Surinamese Javanese”. *Journal of Language Contact*, 12(3): 784-822.
- YOSHIDA, R., & NICHOLS, S. (2024) “The multiple resources of refugee students: A language portrait inquiry”. *International Journal of Multilingualism*, 21(1): 379-399.

## Appendix 1: Sociolinguistic Questionnaire

### Part 1: Personal Information

1. Age (completed years): .....
2. Gender:
  - Male
  - Female
  - Non-binary
3. Place of birth: What is the name of the city where you were born?
4. Where do you live (name of the city)? [Researcher should also ask about the neighborhood]  

---
5. Are there many X (your nationality) in the area where you live? Do you socialize with them?  

---
6. Have you ever had problems renting or buying a house because you are X?  

---
7. Have you ever lived elsewhere for a continuous period of at least one year?
  - Yes
  - No
8. (Only if the answer to question 7 is Yes)  
Where?  

---
- (Possible comparison between the areas where they have lived,  
also regarding interaction with fellow nationals)*
9. How long did you live there (in other places)?  

---
10. What was your job there?  

---
11. Marital status [Also ask for how many years]:
  - Married to an Italian person
  - Married to an X person
  - Married to a person from another country
  - Single
  - Engaged
  - Widowed
12. Where was your spouse/partner born?  

---

13. How long have you been together? How many years were you together?

---

14. Have you ever been married before?

---

15. If yes, for how long?

---

16. Do you have children? How many? How old are they?

---

17. Who do you currently live with?

---

## Part 2: Biographical Information about the Past

18. Who lived with you in the house where you grew up in your home country?

---

19. In your home country, which language(s) did you use with your family?

Father: \_\_\_\_\_

Mother: \_\_\_\_\_

Siblings: \_\_\_\_\_

Elder relatives: \_\_\_\_\_

Other relatives: \_\_\_\_\_

*(N.B.: Important question—try to obtain a detailed answer)*

20. Which language(s) did they use with you?

---

21. Did you attend school in X (country)?

- Yes
- No

22. If yes, where and for how many years?

---

*(Only if the answer to question 21 is Yes. If they studied elsewhere, specify where below)*

---

23. What is the highest level of education you completed in X (country)?

- Pre-School Education (4 years, ages 2–6)
- Primary School (6 years, ages 6–12)
- Junior Secondary School (3 years, ages 12–15)
- General Senior Secondary School (3 years, ages 15–18)
- Technical/Vocational Secondary School (3 years)
- Non-University Studies (3 years)
- Bachelor's Degree (3–4 years)

- Graduate Diploma/Master's Degree (1–2 years)
- Doctorate

If studied abroad: \_\_\_\_\_

24. Which languages were used by teachers during lessons?

\_\_\_\_\_

25. Which languages were taught as subjects?

\_\_\_\_\_

26. Did you have a job in X (country)?

- Yes
- No

If yes, what kind of job?

\_\_\_\_\_

27. How long have you been living in Italy?

\_\_\_\_\_

28. How old were you when you arrived in Italy?

\_\_\_\_\_

29. Who did you arrive with?

\_\_\_\_\_

30. Why did you choose to come to Italy?

- To find a job and settle
- To find a job and save enough money to return to X
- To reunite with family already in Italy
- For other reasons (specify): \_\_\_\_\_

31. Have you attended any school (language courses or others) in Italy?

- Yes
- No

If yes, what type of school/course?

\_\_\_\_\_

For how long? \_\_\_\_\_

32. Are you currently employed?

- Yes
- No (*If no, since when?*)

If yes, what is your job?

\_\_\_\_\_

33. Have you had other jobs in Italy?

\_\_\_\_\_

34. What are the nationalities of your co-workers? More Italians, X, or others?

---

35. Have you ever faced difficulties finding a job because you are X?

---

36. Which language do you mostly use at work?

---

37. Are you currently attending school in Italy?

- Yes (*If yes, which school?*)
- No

38. Since when?

---

39. Have you ever been treated differently at school because of your origin?

---

### **Part 3: Relationship with the Country of Origin**

40. Do you have contacts with your family and friends in X (country)?

- Yes
- No

*(If yes, specify type of contact)*

Phone/video calls

- More than once a month
- About once a month
- Three to four times a year
- Once a year
- Never

Letters / email

- More than once a month
- About once a month
- Three to four times a year
- Once a year
- Never

Travels

- More than once a month
- About once a month
- Three to four times a year
- Once a year
- Never

Other types of contacts (specify) (e.g. WhatsApp, Facebook, Viber)

- More than once a month
- About once a month

- Three to four times a year
- Once a year
- Never

41. When was the last time you visited your country of origin? For how long?

---

42. Do you have any relatives or friends from X (your country) in Italy?

---

If yes, how often do you meet them? (Specify how many times per month)

---

43. Do you have any Italian relatives or friends?

---

If yes, how often do you meet them? (Specify how many times per month)

---

44. Do you have relatives or friends of other nationalities?

---

If yes, how often do you meet them? (Specify how many times per month — once a week is considered often)

---

45. Where do you imagine your life to be in 10 years? Are you planning to return to your country of origin?

---

#### **Part 4: Parents' Information**

46. Where was your mother born?

---

47. Where does your mother live now?

---

48. How long has she been living there?

---

49. Where did she live before?

---

50. Where was your father born?

---

51. Where does your father live now?

---

52. How long has he been living there?

---

53. Where did he live before?

---

54. When did your parents arrive in Italy?

---

55. How old were they when they arrived?

---

56. Who did they come with?

---

**Part 5: Language Knowledge and Self-Assessment**

57. Which language did you first learn in childhood, speaking with your parents and siblings?

---

58. What other languages do you speak? (Please self-evaluate your level for each one.)

Language	Very Well	Well	Not So Well	Definitely Not Well
a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

59. Among the languages you mentioned, which ones can you read and write?

<i>Lingue</i>	LM	(a)	(b)	(c)	(d)	(e)	(f)
Capire <i>Understand</i>							
Parlare <i>Speak</i>							
Leggere <i>Read</i>							
Scrivere <i>Write</i>							

60. Which language do you feel you express yourself best in?

---

61. Where did you learn the languages you know?

<i>Lingue</i>	(a)	(b)	(c)	(d)	(e)	(f)
A scuola <i>At school</i>						
In famiglia <i>Family</i>						
A lavoro <i>At work</i>						
Altro <i>Other</i> <i>(specificare)</i>						

### Part 6: Language Use

62. Which language do you use with the following interlocutors?  
(Indicate emotional importance, practical importance, and time spent with speakers of language X)

Language	Situations	Topics	Emotional Importance (0–10)	Practical Importance (0–10)	Time Spent (less than 1d/week, at least 1d/week, more than 1d/week, every day)
Heritage Language (L1)					
English					
Italian					
L4					
L5					
Neapolitan dialect					
Other					

63. Which language do you speak most frequently?

---

64. Usually, today, which language do you speak with...?

<b>Interlocutor</b>	<b>At Home</b>	<b>Outside Home (Interethnic)</b>	<b>Community Events (Intraethnic)</b>	<b>Specific Topics? (If any)</b>
Mother				
Father				
Brother(s)				
Sister(s)				
Partner				
Husband/Wife				
Children				
Nephews/Nieces				
Elder relatives in Italy				
Friend of same nationality				
Italian friend				
Priest or religious figures				
Employer				
Colleague 1				
Teacher				
Other relatives in Italy				

65. Are there differences in the languages you use today compared to when you were a child?

---

66. What language do your family members (parents, children, etc.) use when speaking to you?

---

67. What language do your parents use when speaking to each other?

---

68. What language did your parents speak when they first arrived in Italy?

---

69. What language do you use when watching TV, using media, or on social networks?

---

70. In which language do you usually read?

---

71. What kind of music do you listen to? (Specify language if possible.)

---

72. In which language do you think?

---

73. In which language do you dream?

---

74. Have you ever attended a Language X course (in Italy or abroad)?

---

### **Part 7: Future of Language X in Italy**

75. In your opinion, how much is Language X currently used in Italy?

---

76. Do you think that speaking Italian could make you forget your heritage languages?

---

77. Do you think that living in Italy could make you forget your cultural roots or traditions?

---

78. Do you think that living in Italy could make your children forget their heritage languages and traditions? Would that make you feel sorry?

---

79. How is Language X spoken today in Italy?

---

80. Are there differences between how the first and second generation speak Language X? Can you give examples?

---

81. Which culture do you feel you belong to? (Do you perceive your cultural identity as more Italian, more X, or a mixture of both?)

---

### **Part 8: Ethnic Orientation**

82. How do you consider yourself? Italian, X, or both?

---

83. How do you feel you were welcomed by Italian society? (Well, badly, so-so?)

---

84. Do you feel integrated into Italian society? (On a scale from 1 to 10)

---

85. In your opinion, is there discrimination against people from X?

---

