ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

TRADUZIONE, INTERPRETAZIONE E INTERCULTURALITA'

Ciclo 38

Settore Concorsuale: 10/L1 - LINGUE, LETTERATURE E CULTURE INGLESE E ANGLO - AMERICANA

Settore Scientifico Disciplinare: L-LIN/12 - LINGUA E TRADUZIONE - LINGUA INGLESE

STARTING THE FLO: AN EMPIRICAL EXPLORATION OF TRANSLANGUAGING IN EUROPEAN DIGITAL NOMAD COMMUNITIES

Presentata da: Novella Tedesco

Coordinatore Dottorato

Chiara Elefante

Supervisore

Silvia Bernardini

Co-supervisore

Cristiana Cervini

Esame finale anno 2026

Abstract

Contemporary mobility is reshaping how individuals mobilise their linguistic resources to negotiate meanings within increasingly superdiverse social settings. Digital nomadism, an emerging form of voluntary mobility grounded in remote work, digital infrastructures and fluid community practices, provides an exemplary context for observing these transformations. Digital nomads can be broadly defined as professionals who travel frequently while working remotely thanks to the affordances of digital infrastructures. They are characterised by a high capital of mobility, within which languages, and especially English, play a central role in enabling participation, networking and access to transnational communities. Rather than merely treating digital nomads as a lifestyle group, the thesis conceptualises them as participants in transcultural exchanges, where linguistic resources and communicative practices are negotiated and emergent through social interaction.

Through an innovative methodological design – where a participatory ethnographic approach provides the from-within perspective that complements and expands more traditional sociolinguistic instruments such as corpora and surveys – the research project aimed to develop open and FAIR language resources (i.e., datasets and methodological protocols) to provide an empirical account of linguistic fluidity characterising transcultural communication within digital nomad communities based in Europe. The study focuses on the European context because digital nomadism here is comparatively less contested, because the EU promotes plurilingualism and transcultural engagement, and because my position as a digital nomad and researcher moving across the European territory has facilitated in-depth access to the European communities.

Within this framework, FLO – the European Fluid Languages Observatory – represents the central research output initiated by this thesis. This is envisaged as a long-term initiative, that builds on the datasets and methodologies developed here and continues beyond the thesis to document, analyse and openly disseminate knowledge about communicative practices shaped by contemporary mobility. Project personal branding through the development of *thenomadlinguist.eu* website provides a public-facing platform for the continued development and expansion of FLO.

Following a short introduction summarising research aims and approaches (i.e., grounded, ethnographic, participatory), Chapter 2 frames digital nomadism within broader debates on contemporary mobilities and digitalisation. A review of literature on the topic of digital nomadism shows how this brench of research has mainly expanded across sociology and tourism studies, while linguistic perspectives remain almost absent.

Chapter 3 then provides the thesis theoretical foundations. Building on mobility studies, translanguaging theories, literature on English as a lingua franca, studies on plurilingual practices like intercomprehension, the thesis advances a theory of linguistic fluidity. In this sense, fluidity is defined as an emergent characteristic of language in use, shaped by mobility trajectories, interactional needs, socially bounded language representations (i.e., attitudes and ideologies). This conceptualisation is operationalised to inform the methodological decisions guiding the empirical work of the project.

Methodological innovation is a major contribution of the study. Chapter 4 presents a triangular mixed-methods design combining:

(i) community-based participatory netnographic and ethnographic observations in Madeira and the Canary Islands (2022–2025);

(ii) a survey documenting repertoires, mobility trajectories and attitudes toward English and translanguaging among Europe-based digital nomads;

(iii) data-intensive linguistic methods leading to the creation of NomadLingo, a corpus of approximately 12 hours of transcribed informal translingual interactions involving 80 participants.

Chapter 5 presents the NomadLingo corpus and addresses methodological issues linked to the processing and management of naturally occurring translingual data. The dataset includes contextual metadata and linguistic annotations enabling several levels of analysis of translanguaging practices, repair strategies, misunderstanding management and meaning negotiation.

The case-studies presented in Chapter 6 illustrates how the resources collected through the project can be explored to answer questions in different linguistic fields: I) macro-sociolinguistics, considering emergent discourse and attitudes challenging monolingual ideologies; II) plurilingualism studies, analysing resources and strategies for meaning negotiation; III) variation studies, problematising a homogeneous view on English as a lingua franca; IV) usage-based cognitive linguistics, leveraging the effects of translanguaging on speech fluency. Beyond the significance of the initial findings presented in Chapter 6, these analyses illustrate how linguistic fluidity can be empirically observed and measured, highlighting the centrality of ethnographic perspectives and data triangularity.

The concluding chapter synthesises the project contributions. Empirically, the thesis provides unprecedented documentation of communication within highly mobile European communities, offering the first integrated set of linguistic resources dedicated to digital nomadism. Theoretically, it bridges different perspectives to advance a comprehensive view of linguistic fluidity. Methodologically, it demonstrates the value of integrating linguistic ethnography with corpus approaches within a FAIR framework, offering a replicable model for future research. Socially and educationally, it contributes to public understanding of translingual practices and to the development of open resources that support language education, transcultural awareness and policy innovation.

Overall, the thesis seeks to offer a renewed understanding of how language operates in a world shaped by continuous mobility.

Preface & Acknowledgments

This thesis is about movement and travel: about how mobility reshapes who we are, and about how, through language, identities are negotiated, performed and transformed. The research project here presented developed in a moment of life when my identity was still in the making, so that research, personal experience and maturation moved together. My Ph.D. has been itself a journey, made of encounters, detours, uncertainty, and, eventually, growth.

I am deeply grateful to my Professors at the Department of Interpreting and Translation of the University of Bologna, who believed in my ideas and allowed me the freedom to explore unconventional paths. I thank my supervisors, Professor Silvia Bernardini, who has been a fundamental guide since my Master's years, nurturing my curiosity through trust and challange, and Professor Cristiana Cervini, for her constant presence and support. I am also grateful to the members of the CoLiTec research group, to Professor Adriano Ferraresi and Professor Maja Milicevic Petrović, to Emanuela and Eleonora, for their generosity and insightful feedback. Thanks to Bianca Gualandi and the Data Steward team of the University of Bologna for their thorough assistance. To my fellow PhD travellers, thank you for sharing doubts, ideas, and moments of mutual support. Thanks to Martina who made Forlì and the DIT feel like home.

My thanks extend to the Professors who hosted me in Madeira and Gran Canaria, supporting my work and facilitating crucial phases of data collection. I am also grateful to the CIUTI Association for financially supporting my visiting period at KU Leuven, where I had the luck of working with Professor Benedikt Szmrecsányi and the QLVL research group. Their engagement, feedback, and scholarly environment allowed me to look at the data from new perspectives and sharpen my analytical tools. Thank you to Benedikt**,** Ruiming, and Chiara for your time and expertise.

Some of the most enduring anchors throughout this journey have been the friends who have known me long before it began. Thank you, Domi, for your constant presence, for reminding me who I am even when I struggled to find my centre. Thank you, Fede, for being such a good travel companion, and to my cousin and lifelong bestfriend Chiara, for being the strongest woman I know. Some friendships do not simply accompany a journey, they give it direction.

This work rests on the emotional, psychological, and material support of my family; without them, it would not exist. My deepest gratitude goes to them. To my mother Andreina, for her unconditional support, resilience, and unwavering belief in me. To zia Picca, for her brilliant mind and steady care. To my brothers Giulia and Luli, who have always been there, near or far, offering support in ways both visible and invisible. To my cousin Mario, who taught me patience and love, and whose hands and eyes have showed me, since we were children, that language is only a small part of communication. To Nonna Wanda, who has been a teacher and a fellow anglicist: it is probably from her that I got my passion for languages and innovation. I

am deeply proud to be your granddaughter. I dedicate my achievements to my dog Sette, who found me when I was seventeen and kept on saving my life from that moment on.

This research was born in movement and shaped by travel across spaces of community. Fuerteventura is where everything began. What started as everyday life, slowly turned into deep conversations, questions, shared reflections, and eventually research. To everyone I met there, thank you for your openness, trust, and generosity. And especially to Gian, my first and only surf mate and one of the most enthusiastic supporters of this project: questa ricerca è nata durante le nostre conversazioni di fronte ai mulini della North Shore e al tromanto del Cotillo.

The journey continued in Madeira, and I wish to thank the Madeiran community as well as the coworking managers in Ponta do Sol and the creators of NomadX for being open, supportive and participative. I am also grateful to the people of the Kite & Surf Coliving in Gran Canaria, and to my Las Palmas flatmates: extraordinary individuals, all with different stories, to whom I have tried to do some justice through these pages.

Canary Islands, Madeira, but also Bali, Thailand, Cabo Verde, Brussels, San Sebastián: writing across places taught me that knowledge is always situated, and that thinking changes when the ground beneath you changes. In my travel to Cabo Verde, I came in contact with the Creole language, and with the amazing people from there: people whose identity is built through translanguaging, and who have made the *trans space* their home. Then, last June, to participate in the International Symposium on Bilingualism, I lended in Donostia. In the sound of its Atlantic waves I found the best soundtrack for my writing sessions, and in the Euskal Herria a renovated curiosity for languages and a further source of inspiration.

Throughout these years, yoga and surfing accompanied my research. They taught me patience, perseverance, balance, attention to my inner self and presence in my body, reminding me that understanding is not only intellectual but also embodied, and that learning often happens through moments of difficulty and quiet listening.

I conclude this journey with a strong desire to continue exploring, applying what I have learned, and contributing to studies on languages and communication. In a time marked by uncertainty and conflict, understanding how we create spaces of interaction and new forms of community may help us rethink the way we live together, and imagine a future where diversity is seen as a richness rather than a threat.

Contents

Chapter 1. Introduction 1

1.1 Research scope and objectives 1

1.2 Doing research from the inside: Researcher's position 3

1.3 General approach to research 4

1.3.1 Doing exploratory and empirical research 4

1.3.2 The ethnographic perspective: Observation, participation, community and knowledge sharing 5

1.3.3 Transdisciplinary dimensions 6

1.4 Structure of the thesis 8

Chapter 2. The context of investigation: Digital nomadism and digital nomad communities in Europe 9

2.1 Approaching the topic of digital nomadism: Introduction 9

2.2 Digital nomadism and digital nomad communities 10

2.2.1 Systematic literature review on digital nomadism 10

2.2.2 Digital nomadism on the web 21

2.2.3 Nomadism, globalisation, voluntary mobility and digitalisation 27

2.2.4 Socio-cultural features of digital nomad communities 34

2.3 Zooming in on the European Context 39

2.3.1 Delimiting geographical research focus to the EU territory 39

2.3.2 EU: Towards an ecology of cultures and languages 40

2.3.3 Digital nomadism in Europe 45

Chapter 3. A theoretical framework to understand linguistic fluidity 50

3.1 Theoretical position and structure of the chapter …….…….. 50

3.2 Mobility and language: A sociolinguistic perspective on language fluidity ……...51

3.3 Languaging, transcultural communication, translanguaging and plurilingualism …....54

3.4 Transcultural communication in action ……....57

3.4.1 English as a lingua franca ……..……....58

3.4.2 Intercomprehension …….....…...61

3.4.3 Translanguaging as a communicative practice ……..……..…....63

3.4.4 Informal translation ……..…….…….....65

3.4.5 Technology-augmented communication …….……..…...67

3.5 Informal language learning and language socialisation ……..……..69

3.6 A theory of fluidity: Towards a practical definitional of language fluidity ….……..71

Chapter 4. Methodology: Triangularity, FAIR data and resource creation …73

4.1 Introduction to project methodology ….……73

4.2 Starting the FLO: The European Fluid Languages Observatory ….……...75

4.2.1 FLO project overview ….……...…..75

4.2.2 Creating FAIR & OPEN language resources ….……..……..…..77

4.2.3 Resource access and project promotion through personal branding ..…….……...86

4.3 Foregrounding Research ….……..…..90

4.3.1 Digital nomadism and language ….……..……...…...90

4.3.2 Developing an operational definition of digital nomad and digital nomad communities …93

4.3.3 Identifying sites for field research: Madeira and Canary Islands ….…….……..95

4.3.4 Developing a methodological framework ….…….……..…...95

4.3.5 Initial findings ….……...…….…..101

4.3.6 Foregrounding research: General remarks ……105

4.4 Mixed methods for linguistic ethnography: From grounded theory to data-driven research ……106

4.4.1 General methodological stance ……106

4.4.2 Community-based participatory ethnography and netnography ……108

4.4.3 Survey research ……148

4.4.4 Data-intensive linguistics: Corpus methods and spoken corpora ……166

Chapter 5. Building the NomadLingo corpus ……169

5.1 Introduction to Chapter 5 ……169

5.2 Approach and methods for natural spoken data collection ……169

5.2.1 Naturally occurring data and consequences of ethnographic participation ……169

5.2.2 General principles and tools for data collection ……170

5.2.3 Data collection in Madeira Islands ……173

5.2.4 Data collection in Canary Islands ……177

5.2.5 The raw data ……179

5.3 Metadata and contextual annotation ……180

5.3.1 Keeping contextual notes and LaMeta for metadata storage ……181

5.3.2 In-text metadata, contextual and structural annotation ……182

5.3.3 Issues with existing TEI guidelines ……183

5.3.4 Corpus accompanying documentation ……185

5.4 Processing translingual spoken data ……186

5.4.1 Transcription pipeline and tools evaluation ……186

5.4.2 Applying pseudoanonymisation to the dataset ……191

5.5 Linguistic annotation ……193

5.5.1 Translanguaging ……194

5.5.2 Misunderstanding ……195

5.5.3 Repair ……197

5.6 Using the corpus: Available formats, download and access ……198

5.6.1 NomadLingo1.1 open ……198

5.6.2 NomadLingo on NoSketchEngine ……199

Chapter 6. Analysing FLO data: Towards a better understanding of linguistic fluidity ……203

6.1 Orientation to Chapter 6 ……203

6.2 Case study 1 – Plural identities and cultures ……204

6.2.1 Standpoint and aims ……204

6.2.2 Case study 1: Data and analytical approach ……204

6.2.3 Identity as / through mobility ……205

6.2.4 Plurality in language resources as identity trait ……207

6.2.5 Performing community membership ……209

6.2.6 Summing up: Digital nomad identity as relational and co-constructed ……209

6.3 Case study 2 – Nomad English: Beyond English as a lingua franca ……210

6.3.1. Reconsidering ELF through mobility ……210

6.3.2 Analytical framework: Integrating variation, repertoires and interaction ……210

6.3.3 Linguistic variation as index of mobility, transcultural repertoires and community building ……211

6.3.4 Translinguistic variation and lexical richness ……214

6.3.5 Speakers' agency, linguistic awareness and sociolinguistic ideologies ……216

6.3.6 English in translanguaging ……222

6.3.7 Linguistic capital and social affordances: A culture of acceptability, a model of accessibility ……223

6.4 Case study 3 – Negotiation and misunderstanding in transcultural communication ….223

6.4.1 Misunderstanding: A space for meaning co-construction in transcultural interactions ….223

6.4.2 Analytical framework for the analysis of misunderstanding sequences…….……..…....224

6.4.3 Signalling trouble and initiating negotiation…….…….…….…..224

6.4.4 Cultural assumptions, peer translation, and narrative re-keying …….……....225

6.4.5 Irony and humour as alignment and misalignment resources …….…….....226

6.4.6 Misunderstanding as an interactional resource in transcultural communication …....227

6.5 Case study 4 – Exploring the cognitive load of translanguaging through statistical modelling of disfluencies …….…….……..228

6.5.1 Introduction: A usage-based perspective on translanguaging and speech (dis)fluency 228

6.5.2 Operationalisation of disfluency 229

6.5.3 Case study 4: Dataset 231

6.5.4 The method: A mixed-effect linear regression model to predict disfluencies based on translanguaging phenomena 234

6.5.5 Model 1 results: Translanguaging, plurilingualism and disfluency ……..236

6.5.6 Model 2 results: The relationship between translangauging and different disfluency types ……...……238

6.5.7 Model 3 results: Impediments as indexes of cognitive load and/or precursors of translanguaging …….…239

6.5.8 A closer look: Qualitative analysis of the relationship between speech impediments and translanguaging ……..241

6.5.9 Case study 4: Limitations ……..246

6.5.10 Case study 4: Implications for views of plurilingual competence in mobility contexts 246

Chapter 7. Final Remarks ……247

7.1 Project summary, main outcomes and results …….247

7.1.1 Project deliverables ……..248

7.1.2 Main findings…… 250

7.1.3 Empirically-grounded theoretical advancements ……251

7.1.4 Social impact and educational initiatives ……252

7.2 Limitations ……253

7.2.1 *Research design* ……253

7.2.2 *Theoretical issues* ……255

7.2.3 *Technical issues* ……256

7.3 Foreseeable Developments, Applications and Future Research Directions ……257

7.4 Conclusion ……258

References ……261

Appendices ……293

Appendix 1. ***Literature review entries*** ……293

Appendix 2. ***FLO Data Management Plan*** ……321

Appendix 3. ***Where's your language at? Questionnaire*** ……347

Appendix 4. ***GC2_Circumstantial Annotation*** ……357

Appendix 5. ***Description of Madeira Islands data collection sessions*** ……359

Appendix 6. ***Description of Canary Islands data collection sessions*** ……366

Appendix 7. ***NomadLingo Raw Data_Metadata on data collection settings and participants*** ……373

Appendix 8. ***NomadLingo1.1_Data_Metadata on data collection settings and participants*** ……375

Appendix 9. ***Annotation scheme for the NomadLingo Corpus*** ……379

Appendix 10. ***NomadLingo1.1 open readme.txt*** ……381

List of figures

Figure 1. Corpus composition: Disciplinary Clusters. …….13

Figure 2. Screenshot of an Instagram post by Nomadhire, a marketing platform for information sharing and consulting on remote work and digital nomadism. ……...…….17

Figure 3. Screenshot of the 'How it works' webpage of Wifi Tribe (as of September 2025). …....24

Figure 4. Screenshot of Nomads.com homepage (as of September 2025). …….….25

Figure 5. Digital nomadism as product of contemporary globalisation processes. …….…….29

Figure 6. Comparison between digital nomads and expats in an Instagram post by Freaking Nomads. …….……...36

Figure 7. Conceptual representation of digital nomad communities ……38

Figure 8. 'Hola no-turista': sign on the front door of a coliving in Fuerteventura. ……46

Figure 9. FLO's data overview and accessibility. ……83

Figure 10a. An Instagram post sharing news about research outputs and professional goals. …89

Figure 10b. An Instagram post sharing a personal picture of life as a digital nomad in Madeira. 89

Figure 10c. An Instagram post to thank participants after a data collection event and promote participation in the online survey. …….…90

Figure 11. Response location in Where's your language at? PILOT (Tedesco 2024b). ……98

Figure 12. Countries of residence of respondents to *Where's your language at? Extended Pilot*.100

Figure 13a. Respondents' level of education in *Where's your language at? Extended pilot*. …..100

Figure 13b. Respondents' field of education in *Where's your language at? Extended pilot*. …..101

Figure 14. Frequency of language usage in digital nomad everyday life. ……..102

Figure 15. Transcultural communicative strategies employed by digital nomads. ……103

Figure 16. Researcher's mobility trajectory from March 2022 to November 2025. ……..114

Figure 17. Sangha coworking, and its welcome page with information after subscription. ……116

Figure 18. A hiking activity organised by Madeira Digital Nomad Community. ……..116

Figure 19. The researcher leading a yoga class on the rooftop terrace of a coliving in Funchal, and the flyer posted on Instagram by one community member to promote the activity. 116

Figure 20a. A board where people could write their impressions about living within

the Madeiran digital nomad community. ……..……117

Figure 20b. A couple of nomads sharing biscuits and Dutch cultural tradition with the other members of the community. ……..117

Figure 21. A group of nomads during the nomad retreat organised by NomadX in Porto Santo.118

Figure 22a. The board at the entrance of Kite&Surf Nomad House in Las Palmas de Gran Canaria. ……...……..118

Figure 22b. Messages on the Whatsapp group of Kite&Surf Nomad House. ……119

Figure 23. A thanking message on the Kite&Surf Nomad House Whatsapp Group with evidence of translanguaging. …….……119

Figure 24a. Screenshot of a message on the Slack Channel for the promotion of a social event at Las Palmas de Gran Canaria. ……...……..120

Figure 24b. A workshop promoted through the Gran Canaria community Slack channel. …120

Figure 25. Talleres Palermo, a coworking in Las Palmas de Gran Canaria. ……121

Figure 26. A coworking café in Las Palmas de Gran Canaria. ……122

Figure 27. Welcome dinner party in Corralejo, Fuerteventura. ……122

Figure 28. Screenshots from the Whatsapp group of the coliving and coworking Surfing Colors in Corralejo, Fuerteventura. …….……..123

Figure 29. A dinner on the beach after a hike with the digital nomad community in Fuerteventura. ……124

Figure 30. Field ethnography locations. ……124

Figure 31. Interview with one of the community managers of Digital Nomads Madeira at the Coworking space in Ponta Do Sol. ……..……128

Figure 32. A picture taken at the Estalagem hotel in Ponta do Sol. ……130

Figure 33. Organisation of the Slack channel used by the digital nomad community in Madeira. …….134

Figure 34. Promotion of a professional workshop and networking event on the Las Palmas de Gran Canaria Slack channel. ……...…136

Figure 35. A post on the Madeira Digitla Nomad Slack Channel. ……..…137

Figure 36. A social event promoted on the Las Palmas de Gran Canaria Slack Channel. …138

Figure 37. A screenshot of messages posted on the Madeira Friends Whatsapp chat in the sub-group related to the community based in Ponta do Sol. ……..…138

Figure 38. A screenshot of a conversation on the Whatsapp group of the Surf&Kite Nomad House. ……...139

Figure 39. A screenshot of an Instagram post published on Digital Nomads Madeira. ……..141

Figure 40. Screenshots of digital nomads Instagram bios. ……141

Figure 41. Screenshot of an Instagram post by one of Corralejo's community managers. ….143

Figure 42. A screenshot of a post published on Remote Work Community Instagram profile. ...145

Figure 43. Conceptual mapping for survey design. ……..151

Figure 44. A question on the practice of tanslanguaging in interaction in *Where's your language at?*. ……151

Figure 45. Questions about linguistic repertoire in *Where's your language at?*. ……153

Figure 46. An open question about translanguaging in *Where's your language at?*. …...…..153

Figure 47. Behavioural questions to identify the target population in *Where's your language at?*. ……154

Figure 48. A comparison between a question in *Where's your language at? Extended Pilot* and *Where's your language at?*. ……155

Figure 49. Flyers advertising *Where's your language at?* at a coworking space in Los Lajares and a meet-up event in Corralejo (Fuerteventura). ……...……158

Figure 50. IP location of responses in *Where's your language at?*. ……...…161

Figure 51. Languages spoken daily by respondents in *Where's your language at?*. …….163

Figure 52. Language learning modalities in *Where's your language at?*. ……..163

Figure 53. Frequency of English use across domains as represented in *Where's your language at?*. ……165

Figure 54. Selected linguistic behaviours and attitudes of participants in *Where's your language at?*. ……166

Figure 55. A picture from data collection in Madeira: participants completing the booklet with consent and privacy forms, and background questionnaire. ……...174

Figure 56. A picture from a data collection session at the Coworking in Ponta do Sol, Madeira.175

Figure 57. Flyer containing link to privacy and consent forms and prompts for NomadLingo metadata collection through audio-recording (English Version). ……178

Figure 58. NomadLingo contextual metadata workflow. ……180

Figure 59. LaMeta *people* metadata input form. ……181

Figure 60. NomadLingo1.1 XML basic structure. ……182

Figure 61. Speaker nationality distribution in NomadLingo1.1 produced through *Text type analysis* on NoSketchEngine. ……200

Figure 62. Concordance lines for the word *what* within misunderstanding sequences. …….201

Figure 63. Translanguaging highlighted through keyword extraction using NoSketchEngine. ..202

Figure 64. Translanguaging occurrences according to context type. ……220

Figure 65. Distribution of translanguaging tags per session. ……221

List of tables

Table 1a. Corpus composition: Type of Publication. ……..……12

Table 1b. Corpus composition: Chronological Distribution. ……...12

Table 2. Digital nomadism terminology. ……16

Table 3. Corpus composition: Methodological approaches. ……18

Table 4. Demographic information about respondents in *Where's your language at? Extended pilot*. ……99

Table 5. Digital nomad attitudes towards English. ……102

Table 6. Examples of translanguaging given by the extended pilot respondents with explanations. ……..104

Table 7. Demographic information on respondents to the survey *Where's your language at?*. 158

Table 8. Recording procedures in data collection sessions for the NomadLingo corpus. …170

Table 9. General information about data collected in Madeira between October and December 2023. ……175

Table 10. Madeiran data collection sessions: An overview. ……175

Table 11. General information about data collected in Canary Islands between February and April 2024. ……..177

Table 12. Canary Islands data collection sessions: An overview. ……..177

Table 13. NomadLingo: Raw Data. ……..178

Table 14. A comparison between TEI Spoken Guidelines and NomadLingo contextual and structural annotation schema. ……..184

Table 15. Speaker codes in NomadLingo. ……191

Table 16. The two-layer pseudoanonymisation system in NomadLingo. ……..192

Table 17. Frequency of *apartment* and *flat* in reference corpora. …….……210

Table 18. ARRIVE vs. COME across speaker groups. ……..211

Table 19. Examples of comparative forms extracted from NomadLingo. ……211

Table 20. Lexical items and semantic fields in translanguaging. ……...214

Table 21. Overview of analytical models for measuring effects of translanguaging on disfluencies. ……..233

Table 22. Fixed effects estimates for Model 1 (Translanguaging on general disfluency). ….235

Table 23. Model fit statistics for Model 1. ……..…...236

Table 24. Effects of translanguaging on disfluency types (Model 2). ……237

Table 25. Fixed effects of Model 3: Translanguaging strategies predicting impediments. …238

Table 26. Samples from NomadLingo1.1 for qualitative analysis of translanguaging and speech impediments. ……...241

Chapter 1

Introduction

The present thesis investigates linguistic fluidity in contexts of mobility by examining how speakers negotiate meaning, manage resources, and co-construct communicative efficacy in transcultural interaction (Baker 2022). The work focuses on informal, everyday encounters among highly mobile individuals, particularly on the newly rising communities of digital nomads in Europe (Woldoff & Lichtfield 2021).

The study addresses how language practices emerge at the intersection of mobility, plurilingual diverse repertoires, and interactional demands. Rather than approaching language as a bounded system or as a set of discrete codes, a usage / practice-based perspective is adopted, by analysing *languaging* as something speakers *do* in interaction (Love 2017). From such a standpoint, linguistic practices are understood as dynamic, contingent, and socially situated, emerging from speakers' engagement with superdiverse communicative ecologies and shaped by their biographies and mobility trajectories.

The research is grounded in an original empirical infrastructure and combines qualitative and quantitative approaches to explore translanguaging from a usage-based perspective, which is here defined as *linguistic*, or *language*, *fluidity*. At the core of the research hence lies the concept of *linguistic fluidity*, used to capture contingent forms of variability and change affecting multiple levels of language use – phonological, morphosyntactic, pragmatic, and interactional – under conditions of transcultural contact (Otsuji & Pennycook 2010; Jasper & Madsen 2019; Prinsloo 2023; Baker 2024). This view aligns with approaches that treat languages as social constructs and *languaging* as an idiosyncratic yet fundamentally social process (Love 2017).

Digital nomad communities are taken as a particularly relevant site of investigation because they constitute superdiverse and digitally-augmented social formations exemplifying mobility and digitalisation patterns typical of the globalised societies (Blommaert 2010; Dreher & Triandafyllidou 2023).

The chapter proceeds by outlining the scope and objectives of the research (Section 1.1), clarifying the researcher's positionality and the situated nature of the study (Section 1.2), and presenting the general methodological and epistemological approach adopted (Section 1.3). It then concludes with an overview of the structure of the thesis (Section 1.4).

1.1 Research scope and objectives

The primary scope of this research is to investigate linguistic fluidity as a characteristic of transcultural communication in contexts shaped by mobility, superdiversity (Blommaert & Varis 2015), and culturally heterogeneous interactional settings. The study focuses on informal spoken communication among speakers with diverse linguistic repertoires, examining how they mobilise linguistic, semiotic, and pragmatic resources to achieve understanding, manage interactional trouble, and negotiate social alignment, and their perception and conciouseness around their own communicative behaviours and attitudes.

Within this framework, the research is guided by the following overarching questions:

1. How does transcultural communication unfold within digital nomad communities, and how is it perceived by speakers?
2. Which languaging processes and communicative strategies do speakers use in order to communicate successfully in contexts characterised by high linguistic and cultural diversity?
3. What are the salient characteristics of language use in these communities, and how do they reflect broader patterns of linguistic fluidity under conditions of mobility?

In addressing these questions, the study takes into account:

- the sociolinguistic context of the communities under investigation (including dominant languages, community norms, degrees of institutionalisation, the cultural relation between nomad and local communities);
- the communicative practices enacted by community members in everyday interaction;
- speakers' attitudes and reflections on their own linguistic practices, in line with the participatory orientation of the research;
- selected sociopragmatic functions, such as misunderstanding management, mediation and negotiation, irony, and stance-taking;
- formal and structural features of language use, examined through a corpus-informed approach and, where relevant, in dialogue with construction-based perspectives (Fried & Van Trijp 2004-2026).

More specifically, the objectives of the thesis are:

- to develop an empirically grounded framework for analysing linguistic fluidity that integrates insights from translanguaging, English as a lingua franca and intercomprehension research;
- to create resources for the exploration of transcultural communication
    - and, within this general aim, particularly to build a spoken corpus designed to capture translingual naturally occurring interaction in informal social contexts within digital nomad communities;
- to identify and analyse transcultural communicative practices as they unfold in the contexts under investigation from several sociolinguistic perspectives.

By articulating these objectives and research questions, the thesis positions itself at the intersection of sociolinguistics, applied linguistics, and mobility studies, contributing both conceptual clarification and empirical evidence to current debates on language in highly mobile, transcultural settings.

1.2 Doing research from the inside: Researcher's position

This research is conducted from an insider perspective. My long-term personal and professional experience within digital nomad communities and other mobile social environments has positioned me not only as an observer, but also as a participant in many of the interactional contexts from which the data were collected. This prolonged engagement has shaped both my access to the field and my sensitivity to the social meanings, norms, and expectations embedded in everyday communicative practices.

The notion of positionality, originally developed within feminist epistemologies and critical social research, has progressively entered sociolinguistics and linguistic ethnography as a way to account for the researcher's social location, trajectories, and involvement in the field (Bucholtz 2001; Heller 2008; Copland & Creese 2015). Rather than being treated as a source of bias to be neutralised, positionality is here conceptualised as an epistemic resource. This stance aligns with ethnographically oriented approaches that recognise knowledge production as situated, relational, and historically embedded (e.g., Haraway 1988; Bourdieu 1999).

This perspective – here emrbaced – allows a view on the researcher as constitutively part of the social world under investigation. Access, trust, and the interpretation of meaning are shaped by shared experiences, linguistic practices, and forms of social alignment that emerge through participation. In the context of this study, being embedded in digital nomad communities enabled access to interactional settings, facilitated trust-based data collection, and supported a fine-grained understanding of interactional behaviours and language practices that would have been difficult to capture through short-term or purely observational methods.

At the same time, conducting research from the inside requires systematic reflexivity. Throughout the project, I explicitly documented methodological choices, analytical decisions, and interpretative boundaries, acknowledging how my own trajectories, linguistic repertoires, and social positioning may shape data collection as well as analytical processes. This reflexive stance is operationalised in the research design described in Chapter 4, where issues of access, participant recruitment, data elicitation, and ethical considerations are addressed in detail.

By adopting this positioned and reflexive approach, the study aligns with research traditions that view knowledge as co-constructed with participants, rather than extracted from them. The researcher's involvement is thus not framed as a methodological weakness, but as a condition that, when critically examined and transparently accounted for, enhances analytical depth and ecological validity. This positioning is particularly relevant in the study of mobile, informal, and transitory communities, where meaning-making practices are ephemeral and deeply

embedded in shared lived experience and cannot be fully understood without sustained participation.

1.3 General approach to research

The thesis adopts an empirically grounded, mixed-methods research design, in which qualitative and quantitative approaches are systematically combined and mutually inform one another. While qualitative, interaction-oriented analysis plays a central role in interpreting situated communicative practices, the study also relies extensively on quantitative data and analyses to identify patterns, tendencies, and correlations. The research design is therefore qualitatively informed and quantitatively supported, reflecting the complex nature of the empirical material and the research questions addressed.

Rather than testing predefined hypotheses derived from abstract or idealised models of language, the study proceeds inductively, allowing regularities and analytically relevant phenomena to emerge from the data, while remaining theoretically informed. This approach is consistent with the exploratory nature of the project and with the choice to investigate a relatively underexplored sociolinguistic context, where rigid hypothesis-testing would risk imposing external categories on practices that are still poorly documented.

This general approach also reflects the conviction that linguistic fluidity cannot be adequately captured through a single method or theoretical lens, but requires methodological pluralism and conceptual flexibility. Linguistic fluidity is treated as a complex phenomenon that unfolds at the intersection of interactional, cognitive, and sociobiographical dimensions, and its investigation therefore calls for analytical tools capable of addressing both micro-level interactional dynamics, broader distributional tendencies within the corpus and macro-level social tendencies.

Accordingly, the research design combines (i) exploratory and empirically driven analysis, (ii) an ethnographic perspective grounded in observation, participation, and community engagement, and (iii) a transdisciplinary orientation that brings together insights from sociolinguistics, applied linguistics, mobility studies, and communication research. These three dimensions structure the general approach adopted in the thesis and are developed in detail in the following sub-sections (1.3.1 – 1.3.3).

1.3.1 Doing exploratory and empirical research

Given the scarcity of corpora explicitly designed to capture translingual spoken interaction and research addressing the topic of language and communitcation within the communities under study, the present work is exploratory in nature. The construction of the NomadLingo corpus is therefore not merely a preparatory step, but a core component of the research contribution which documents underrepresented communication practices.

Exploration operates at multiple, interrelated levels. First, it concerns the identification of interactionally salient phenomena emerging from various kinds of ethnographic data, without imposing rigid a priori categories. Second, it informs the development of the annotation schemes, which required the operationalisation and partial reconceptualisation of theoretical notions drawn from translanguaging, English as a lingua franca, and intercomprehension research within an integrated, practice-oriented framework. Rather than mapping pre-existing concepts directly onto the data, annotation categories were progressively refined through iterative engagement with empirical material, allowing theory and data to mutually inform one another.

Third, exploration characterises the selection and combination of analytical strategies. Quantitative analyses are employed not with the aim of simplifying complexity, but to identify recurrent tendencies, distributions, and associations across the datasets. These patterns are then interpreted through fine-grained qualitative and, in the case of the corpus, interactional analysis, ensuring that statistical tendencies remain anchored in the situated reality of communicative practices.

The iterative movement between empirical observation and theoretical refinement aligns with grounded and inductive approaches to sociolinguistic research, in which analytical categories are treated as provisional and revisable, and theory is progressively shaped through sustained engagement with data rather than imposed in advance (Glaser & Strauss 1967; Charmaz 2006). At the same time, the study adopts a theoretically informed form of exploration, acknowledging that data-driven analysis is always guided by prior conceptual frameworks, disciplinary traditions, and research questions (Hammersley & Atkinson 2019).

Within this perspective, exploratory research is not understood as preliminary or methodologically weak, but as a necessary and productive strategy for investigating underexplored social phenomena and communicative ecologies. By opening analytical space for unexpected patterns and emergent practices, the exploratory design adopted here enables the development of new hypotheses, methodological tools, and theoretical insights concerning linguistic fluidity in contexts of mobility.

1.3.2 The ethnographic perspective: Observation, participation, community and knowledge sharing

An ethnographic orientation underpins the entire research process. Data collection was informed by long-term multisited observation, sustained participation in everyday activities, and continuous interaction within the communities under study. Language is therefore analysed not as an abstract system, but as embedded in social relationships, shared practices, and locally meaningful routines. This perspective allows communicative practices to be interpreted in relation to speakers' trajectories, interactional histories, and the specific socio-material conditions in which encounters unfold.

The study draws on the tradition of linguistic ethnography, which has been central to sociolinguistics since the discipline's emergence. From the foundational work of Hymes (1964, 1972) and Gumperz (1968), ethnography has provided a framework for understanding language as social action, requiring the integration of linguistic analysis with ethnographic sensitivity to context. More recent developments in linguistic ethnography have further emphasised the need to link micro-level interactional analysis with macro-level social processes, foregrounding meaning-making as situated, negotiated, and temporally layered (Rampton et al. 2004; Copland & Creese 2015; Rampton et al. 2015).

Within this tradition, ethnography is not conceived merely as a set of methods, but as a research philosophy that assumes the co-constitutive relationship between linguistic form, communicative practice, and social life (Creese & Blackledge 2015). Importantly, as Rampton et al. (2015) note, linguistic ethnography does not prescribe a fixed toolkit, but rather encourages methodological adaptability in response to the community, the research questions, and the forms of data involved. This flexibility is particularly relevant for the study of digital nomad communities, characterised by transient co-presence, diverse semiotic repertoires, and changing interactional ecologies.

In this project, the ethnographic perspective is closely intertwined with a participatory orientation. Rather than treating participants as passive sources of data, the research adopts a view of knowledge as co-constructed with community members. This approach aligns with participatory and community-based research traditions, which emphasise reciprocity, accountability, and the social relevance of academic inquiry (Balazs & Morello-Frosch 2013). While participatory action research (PAR) is often associated with explicitly emancipatory or justice-oriented contexts, in this study it is primarily adopted to foreground the processes through which knowledge is produced, shared, and negotiated, and to reflect critically on my own role within those processes (as introduced in Section 1.2).

The participatory dimension of the project was motivated by three main factors. First, digital nomad communities themselves are embedded in a participatory culture (Jenkins 2009), where peer-to-peer knowledge exchange, collaboration, and informal mentoring are everyday practices. Second, my long-term involvement in mobile and nomadic lifestyles enabled authentic insider access. Third, this orientation allowed the project to contribute not only to academic knowledge, but also to the development of community competences, reflexive awareness, and shared linguistic sensitivity among participants, through feedback moments, informal discussions, and public-facing activities.

Ethnography also informs the ethical stance of the project. Transparency, reflexivity, and responsibility are treated as integral to the research process rather than as external constraints.

Finally, the ethnographic perspective is combined with a grounded-theory sensitivity (Glaser & Strauss 1967). While the study does not apply grounded theory in a strict or orthodox manner, it adopts its core principle of theory development emerging from sustained engagement with empirical material. This data-driven orientation supports the project's

broader aim of developing an empirically grounded account of linguistic fluidity that remains sensitive to the lived realities, communicative practices, and social dynamics of the communities under investigation.

1.3.3 Transdisciplinary dimensions

The thesis adopts an explicitly transdisciplinary orientation, grounded in sociolinguistics and applied linguistics but systematically engaging with insights from mobility studies and human geography, linguistic anthropology, education, discourse studies, and research on technology-mediated interaction. This choice reflects the assumption that linguistic fluidity in contexts of high mobility and digital mediation cannot be adequately understood within the boundaries of a single disciplinary tradition.

Sociolinguistics itself has long been characterised by an interdisciplinary impulse, insofar as it requires social theory, ethnographic sensitivity, and contextual interpretation to complement linguistic analysis. This is particularly evident within linguistic ethnography, which has been described as an *interdisciplinary region* rather than a closed methodological framework, encouraging researchers to combine tools and perspectives in ways that are responsive to the research context and questions at stake (Rampton et al. 2004; Tusting & Maybin 2007) In this sense, the present study aligns with traditions that view interdisciplinarity not as the juxtaposition of perspectives, but as their integration around a shared empirical problem.

However, the term *transdisciplinary* rather than *interdisciplinary* is deliberately adopted in order to signal a further epistemological step. While interdisciplinarity typically involves collaboration and exchange across established disciplinary boundaries, *transdisciplinarity* emphasises the partial suspension and reconfiguration of those boundaries in response to complex social phenomena that exceed any single disciplinary logic (Klein 2010; Nicolescu 2014). In this sense, the prefix *trans-* foregrounds movement *across*, *between*, and *beyond* disciplines, allowing concepts, methods, and analytical categories to be reassembled around the object of inquiry rather than inherited from disciplinary traditions.

The thesis enters into dialogue with research on digital nomadism and lifestyle mobility, with studies of technology-mediated and platform-based communication, and with work on informal and non-institutional learning environments. These perspectives make it possible to situate interactional language practices within broader socio-technical and cultural transformations, rather than treating them as purely linguistic and/or pragmatic phenomena.

This transdisciplinary stance is also consistent with recent work on transcultural communication, which emphasises that identity negotiation, communicative alignment, and participation emerge at the intersection of linguistic, social, psychological, and technological factors, and therefore require integrated analytical models rather than discipline-specific explanations (Byram 2008; Deardorff 2009; Baker 2015).

Hence, transdisciplinarity in this thesis is not pursued for its own sake, but as a methodological necessity dictated by the complexity of the phenomenon under investigation. The integration

of sociolinguistic analysis, ethnographic observation, corpus-informed methods, and insights from mobility and technology studies enables a more comprehensive account of how linguistic fluidity is enacted, constrained, and made meaningful in everyday interaction. This integrative approach is fully aligned with the mixed-methods design and ethnographic orientation discussed in this section, and underpins the research framework developed throughout the thesis.

1.4 Structure of the thesis

The thesis is structured as follows.

Chapter 2 contextualises the study by outlining digital nomadism as a sociolinguistic setting and discussing mobility as a key condition shaping contemporary communication. This contextualisation is further developed and empirically expanded through the preliminary findings presented in Chapter 4, where initial data are used not only to illustrate the methodological design but also to refine and substantiate the theoretical assumptions introduced in Chapters 2 and 3. Chapter 3 presents the theoretical framework particularly inspired by translanguaging theories, introducing an operational definition of linguistic fluidity based on literature about transcultural communication, translanguaging, English as a lingua franca, intercomprehension, and language socialisation. Chapter 4 details the methodological design of the study, focusing on survey and ethnographic participation. Importantly, this chapter also reports preliminary empirical results that contribute to validating and extending the theoretical framework, thereby informing the final methodological choices adopted in the study. Given its central role within the thesis, the construction of the NomadLingo corpus is treated in a dedicated chapter. The corpus is innovative both in the type of linguistic phenomena it represents – still underrepresented in existing corpora – and in the methodology developed to capture and structure such data. Chapter 5 therefore introduces the corpus, describing in detail its design, construction procedures, structure, and access modalities, while also highlighiting its metascientific value for contemporary corpus linguistics research. Chapter 6 presents the empirical analyses, combining quantitative and qualitative methods to explore different aspects of translanguaging from a usage-based perspective. Finally, Chapter 7 synthesises the main contributions of the thesis, reflects on its limitations, discusses social impact and envisaged result applications, and outlines directions for further research.

Chapter 2

The context of investigation: Digital nomadism and digital nomad communities in Europe

2.1 Approaching the topic of digital nomadism: Introduction

In this chapter, digital nomadism and digital nomad communities in Europe are presented as a prime example of transcultural communicative contexts. Digital nomadism refers to a form of location-independent work and lifestyle in which individuals leverage digital technologies to sustain their professional activities while engaging in continuous and/or long-term travel. The phenomenon can be situated at the intersection of remote work, travel, and lifestyle migration, combining the flexibility of digital infrastructures with aspirations to freedom and exploration (Makimoto & Manners 1997; Reichenberger 2018; Hermann & Paris 2020). Thereby, digital nomads are individuals who work remotely through digital means while traveling or relocating frequently across different geographic contexts. They are not a homogeneous group but encompass diverse profiles, from freelancers and entrepreneurs to salaried employees engaged in telework. What unites them is the valorisation of mobility and autonomy, often articulated through narratives of location independence, work–life balance, and self-realisation (Mancinelli 2020; Hannonen 2020; Cook 2023).

The present chapter addresses both digital nomadism, conceptualised as a sociocultural phenomenon of contemporary mobility and digitalisation, and digital nomad communities, which are introduced as *light* and *fluid* communities *of practice* (Section 2.2). Such discussion is based on a systematic literature review updated to July 2025 (an overview of which is presented in Chapter 2, Sub-section 2.2.2). Given the relative novelty of the phenomenon under study, publications are continuously coming out and much useful information can also be retrieved from non-academic sources. For this reason, a review of non-academic materials findable on the web related to the phenomenon is provided in Sub-section 2.2.2.

The geographical borders delimitating the area of interest for this research project are those of the European Union. Therefore, Section 2.2 offers a discussion of the main socio-cultural and linguistic characteristics of the EU, while introducing concepts like plurilingualism and pluriculturalism, also relating them to the Common European Framework of Reference (CEFR). Included in this section is an overview of the main digital nomad communities present in the European territory, exploring the links between European contemporary culture and

digital nomadism. This also provides information on the distinguishing features of the European communities compared to those located elsewhere, like the ones based in Asia (e.g., in Bali, Indonesia, or in Chang Mai, Thailand [see, for example, Green 2020 or Qotim et al. 2025]).

2.2 Digital nomadism and digital nomad communities

Digital nomadism has expanded significantly worldwide since the Covid-19 pandemic, and the last years have witnessed a proliferation of academic contributions as well as online texts, guides, marketing materials engaging with the topic and providing valuable information and perspectives. Hence, the following two sub-sections aim to provide a systematic and as comprehensive an account as possible of the relevant literature and other sources like websites and social media platforms. The first aims to identify current research trends and gaps regarding the topic, while the second appears essential for a deeper understanding of the phenomenon, especially given the participatory approach adopted in the present research (see Section 1.3.2) and the 'participatory' horizontal nature of the communities under study (Jenkins 2009).
Then, drawing on review findings as well as on the broader frameworks of mobility and globalisation studies, in Sub-section 2.2.3 a general definition of digital nomadism is provided, and the phenomenon is presented from a socio-cultural perspective. By describing their features, an operational definition of what is here meant by the term *digital nomad communities* is given in Sub-section 2.2.4, and the main characteristics of these communities are described in the light of contemporary sociological research.

2.2.1 Systematic literature review on digital nomadism

This literature review presents an updated, systematic overview of open access scholarly research on digital nomadism, based on a corpus of 224 academic sources. The review follows a structured approach aimed at identifying key trends and thematic clusters, disciplinary orientations and methodological patterns in the existing literature. A complete bibliography of the reviewed literature, including information like type of publication, URLs or DOIs for access, is available in Appendix 1. Only works explicitly cited are present in the list of References. Based on the systematic literature review, then, a focus on few publications that

proved to be particularly relevant for the characterisation of digital nomadism within the present research is provided.

The literature was selected according to a set of inclusion and exclusion criteria. Only documents written in English were considered, and particular attention was paid to accessibility: materials had to be openly accessible and/or retrievable via standard academic databases. In few cases access was restricted through academic account sign in, for this reason, only a sub-set of the literature corpus containing abstracts (or introductions when abstracts are missing) is available for download on the GitHub FLO repository created for this thesis (more information concerning data management can be read in Section 5.1).[1]

Although the term digital nomadism first appeared as the title of a book by Makimoto & Manners in 1997. The topic has attracted growing attention across disciplines since 2021, with a marked rise in studies that seek to understand this type of emergent mobility as well as the forms of identities and communities it generates. To allow for the inclusion of both pioneering contributions and up-to-date research outputs, no constraints were applied with regard to the year of publication. More specifically, inclusion and exclusion criteria were applied as follows.

Inclusion criteria:

- The document must be written in English;
- The source must be available open access, or through academic or institutional access;
- The study must focus primarily on the phenomenon of digital nomadism, rather than mentioning it tangentially.

Exclusion criteria:

- Non-academic content such as blog posts, journalistic articles, or promotional websites;
- Incomplete bibliographic records (i.e., lacking author, year, or title);
- Content which could not be fully accessed by the researcher.

[1] https://github.com/users/novella-tedesco/projects/1.

It should be noted that many texts encountered during the searches were not accessible and were therefore excluded. Such a choice – necessary due to project budget – also represents a general commitment to Open Science which characterises the whole work (see Section 4.2).

Moreover, as already mentioned, community and commercial websites represent an essential source of information and have been reportedly used in studies on digital nomadism; an in-depth discussion on such non-academic sources is offered in Section 2.3.2.

Texts were retrieved via multiple rounds of keyword searches ('digital nomad*', 'remote work and mobility', 'digital nomad communities') on Google Scholar, Semantic Scholar, and DOAJ. The corpus was last updated in July 2025.

General information about corpus composition, including all materials available for download (after the selection and exclusion criteria were applied), can be found in Tables 1a and 1b and Figure 1.

Table 1a. Corpus composition: Type of Publication.

Type of Publication	**Number of Sources**
Journal/magazine articles	161
Conference papers	24
PhD/Master's/Bachelor's theses	19
Book chapters	10
Books	4
Reports	1
Working paper/ preprints	5

After journal articles, which are the most common type of academic outputs in social sciences, the number of conference papers indicates a dynamic, still-forming field: as shown in Table 3 below, many contributions are exploratory or conceptual, while some recent ones report empirical studies.

Table 1b. Corpus composition: Chronological Distribution.

Year	**Number of studies**
2006–2016	8

2017	2
2018	7
2019	11
2020	19
2021	21
2022	24
2023	31
2024	68
2025	33
2026	1

From 2019 onwards the literature has grown steadily, with a notable surge during/after the COVID-19 pandemic (see Section 2.4). The 367 distinct authors suggest broad interest for the topic. An effort towards transdisciplinary research is unveiled by the many works written by more scholars together, consistent with the phenomenon's complexity as well as the diversity of disciplinary fields represented in the corpus (see Figure 1).

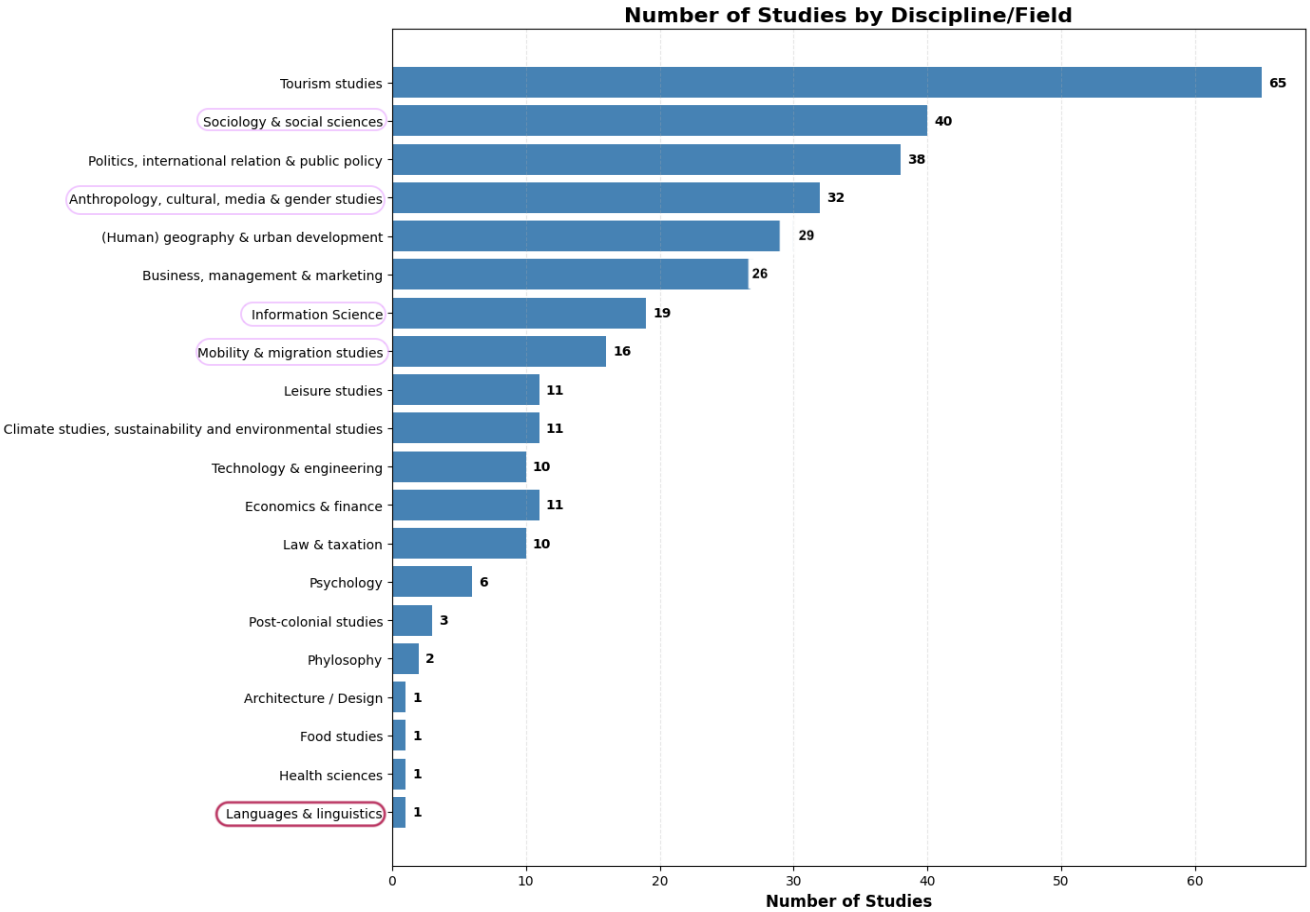


Figure 1. Corpus composition: Disciplinary Clusters.

To produce the categorisation shown in Figure 1, articles were manually classified based on the main disciplinary perspective, author(s)' field(s) of study, or journal scope, with overlaps noted for interdisciplinary work, meaning that the same article may be counted as pertaining to several disciplines. In Figure 1, disciplines most relevant to the present research are circled in pink, while linguistics – the main field of this thesis – is circled in red.

Sociology and tourism studies appear as the dominant perspective of investigation, with a steady increase in publications from IT, political science, and law in the latest years. Studies exploring communication generally focus on marketing, self-representation or, in a few cases, community building. While some recent publications focus on social media communication, digital nomad self-representation and discourse (see, for example, Isaac & Dorrestein 2025, or Chen & Pun 2025), linguistics perspective and approaches are almost unrepresented.

A corpus-driven analysis of frequent words, expressions and keywords in the open-access sub-corpus of abstracts highlights the following recurring themes:

- Digital infrastructure & ICT (e.g., Mohlin & Chatzipanagiotou 2023);
- Remote work, location independence and mobility patterns (e.g., Hermann & Paris 2020; Ingvarsson 2023)
- Coworking/coliving spaces (e.g., von Zumbusch & Lalicic 2020; Lee et al. 2019)
- Lifestyle & identity (e.g., Reichenberger 2018; Cook 2023)
- Digital nomadism and local contexts (e.g., Wibowo 2024)
- Marketing and tourism development (e.g., Getman 2021)
- Social inequality, privilege & policy (e.g., Mancinelli 2020; Dreher & Triandafyllidou 2023)
- The Covid-19 pandemic impacts on mobilities and work (e.g., Shawkat et al. 2021; Ehn et al. 2023).

These themes often overlap, confirming digital nomadism as a socio-cultural, economic, spatial and geopolitical phenomenon that requires integrative, cross-disciplinary perspectives.

Before the pandemic, researchers typically framed digital nomadism as an emergent lifestyle anchored in mobility ideals, identity work, and enabling infrastructures. Core strands in the earliest considered publications (up to 2020) include: (i) definitional/typological debates and lifestyle–identity construction (e.g., Reichenberger 2018]); (ii) coworking/coliving as social

and spatial infrastructures that facilitate belonging and work/leisure balance (e.g., Orel 2019]); (iii) information infrastructures and digital platforms enabling location-independent work (e.g., Wang et al. 2019) ; (iv) early place-based ethnographies and mobilities questioning how 'home' and 'community' are reworked on the move (e.g., Green 2020]).

The pandemic is an inflection point. With the normalisation of remote work, mobility patterns were disrupted and then reconfigured, and policy frameworks rapidly evolved (see Section 2.2.3). Consequently, post-2020 studies broaden in at least five directions:

- Normalisation of remote/hybrid work and new practices: *workcation/work-cation* (a neologism that identifies leisure trips in which people keep on working remotely), work-leisure boundary management, and well-being/engagement enter the agenda, often via surveys or experimental designs (e.g., Ji et al. 2024; Kozak et al. 2024);
- Policy and governance: comparative analyses of digital-nomad visas, immigration/taxation, and labour rights proliferate, tying nomadism to migration and public-policy debates (e.g., Koskela & Beckers 2024; Dreher & Triandafyllidou 2023; Rasnača 2023; Narotzki & Kuperberg 2023; Arbutina et al. 2023);
- Urban/real-estate and destination effects: attention shifts to housing markets, gentrification risks, and destination strategy/marketing (e.g., Sequera 2024; Taltavull 2024; Bednorz 2024; Zhou et al. 2024);
- Inequalities and social differentiation: growing focus on gender, race, and privilege, plus loneliness/mental health in globally mobile lives (e.g., Linao et al. 2024; Miguel & Lutz 2024);
- Field consolidation: a wave of bibliometric/systematic reviews mapping the domain and signaling maturation (e.g., Šímová 2023; Özyürek et al. 2024; Başaran 2025; Sousa et al. 2025).

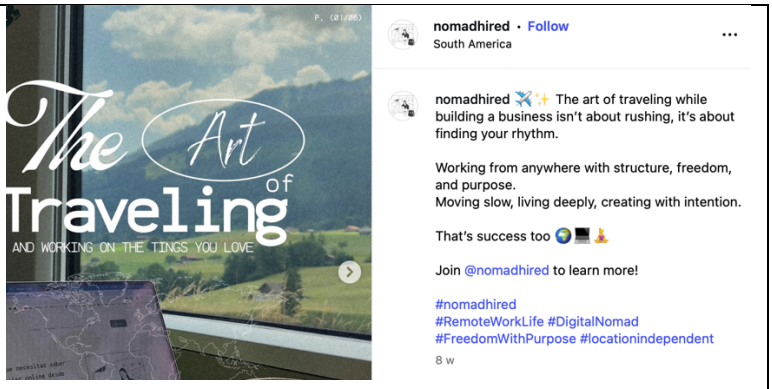
Linguistic and communication studies remain scarce, underscoring the gap this thesis seeks to address (i.e., exploration of language fluidity, translanguaging practices, and technology-augmented interaction for community building). In short, before 2020 the main topics under discussion were identity/lifestyle, coworking as community infrastructure, platform affordances, and limited policy/market impact analysis. These studies were characterised by qualitative analysis, and the recurrent adoption of ethnographic perspectives. Whereas after 2020, the focus shifted towards remote/hybrid normalisation, workcation practices,

destination/real-estate effects, visa/tax governance, inequalities and well-being. While the ethnographic perspective is still prominent, these newer publications show more mixed and quantitative methodologies, including some field-mapping reviews. These shifts, both in themes and in methods, contextualise the present thesis: they justify a move from lifestyle-centric accounts to an integrated perspective combining mobility and transculturality by analysing language and its actual use.

The analysis also highlights digital nomadism as an emergent terminological domain. Interestingly, the terms, often neologisms, extracted from the sub-corpus are also repeated in hashtags used either by the nomads or by companies to attract their attention in social media posts, marketing articles and blog posts (see Sections 2.2.2 and 4.2.3). Some examples are provided in Table 2.

Table 2. Digital nomadism terminology.

Term/ word	**Frequency in the reviewed academic literature**	**Example of use within community networks**
Remote work	50	'Remote work community': the name of one of the most popular networking website for digital nomads (https://www.remoteworkcommunity.com)
Workation / workcation / work-cation	17	'What places would you suggest for a workation during the winter months?': question asked on Reddit 10 months ago (https://www.reddit.com/r/SoloTravel_India/comments/1ghqkw7/what_places_would_you_suggest_for_a_workation/)
Location independence / location independent	24	#locationindependence 38.7k posts on Instagram, #locationindepent 558k posts on Instagram (example shown in Figure 2)

		 Figure 2. Screenshot of an Instagram post by Nomadhire, a marketing platform for information sharing and consulting on remote work and digital nomadism.

These patterns suggest that a shared lexicon circulates across academic, industry and community spaces. Thus, digital nomad culture appears to be co-produced through polycentric practices of knowledge-making and sharing (as will be further explored in Sections 2.2.2 and 2.2.4).

As briefly mentioned above, this literature review also led to relevant insights into methodological practices. To extract information about methodology, a script was developed which automatically classifies papers into the methodological categories listed in the left column of Table 3. Each abstract was lowercased, tokenised, and then matched against a set of manually curated keyword dictionaries through a Python code available, like the corpus of abstracts, on Github[2]. These dictionaries reflect typical methodological markers, which were found through a corpus-driven word-list extraction performed on the BAWE corpus[3], as follows:

- Qualitative research keywords: 'interview', 'semi-structured', 'ethnography', 'focus group', 'thematic analysis', 'discourse analysis', 'fieldwork'.
- Quantitative research keywords: 'survey', 'questionnaire', 'regression', 'ANOVA', 'experiment', 'statistical analysis', 'sample size'.

[2] https://github.com/novella-tedesco/FLO/blob/Digital-Nomadism-LitRev/LitRevMethodology.ipynb

[3] The British Academic Written English (BAWE) corpus, was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics, Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, Reading) and Paul Wickens (School of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

- Mixed methods: detection of both qualitative and quantitative markers in the same abstract, or explicit mentions of ‘mixed methods’ or ‘triangulation’.
- Bibliometric/systematic reviews keywords: ‘systematic literature review’, ‘bibliometric’, ‘scientometric’, ‘meta-analysis’, ‘state of the art’.
- Technical studies/reports keywords: ‘policy report’, ‘white paper’, ‘prototype’, ‘simulation’, ‘system design’, or by publication type metadata (e.g., reports).
- Theoretical or opinion-based: default category applied when no methodological markers were detected in the abstract.

The classification logic was implemented as a sequence of conditional rules (*if/elif* statements) applied to each cleaned abstract. For instance, if both qualitative and quantitative markers appeared in the same text, the script returned ‘Mixed methods’; if only review-related markers were found, it returned ‘Bibliometric/systematic reviews.’

The results of the automatic classification are shown in Table 3.

Table 3. Corpus composition: Methodological approaches.

Methodology	**Number of studies**
Theoretical or opinion-based	96
Qualitative research	70
Mixed methods	27
Quantitative research	15
Bibliometric/systematic reviews	13
Technical studies / reports	3

Four features stand out:

I) ‘Theoretical or opinion-based’ is the largest category, partly reflecting the state of the field – rapidly evolving, concept-heavy and often positioned at the intersection of mobility, work and lifestyle studies. However, this is also due to the methodology applied: our automatic classifier relies only on abstract-level cues; moreover, where explicit markers are absent, items default to ‘opinion-based’. This conservative bias minimises false positives but may undercount empirical designs when abstracts are non-transparent.

Opinion-based contributions play an important role because the phenomenon is constantly evolving and still regarded by many as controversial. Some early papers, often published in interdisciplinary journals, were written as conceptual or reflective essays rather than empirical studies (e.g., Schlagwein 2017; Cook 2020; Brusseau 2021). These works were essential in opening up the field, raising questions on definitions, scope and implications of digital nomadism, even though they did not always rely on systematic data collection.

II) Qualitative designs remain prominent, i.e. ethnography and netnography. Interviews and case studies are the backbone of both early and contemporary work (e.g. Aroles et al. 2022; Prester 2022), often complemented by interpretive analyses of netnographic elements like Twitter or Instagram contents (e.g. Polson 2024; Lutz & Perez-Vega 2024) (see Sections 2.2.2 and 4.4.2). For instance, an influential study on digital nomadism by Reichenberger explores identity construction among the nomads using data gathered through semi-structured interviews linking them to 'on-line […] existing non-academic information' (2018).

III) Mixed and quantitative approaches are expanding, especially in tourism and management: survey-based modelling and experimental designs examine motivations, well-being, intention to stay/revisit, place attachment and policy impacts (e.g. Ji et al. 2024; Kozak et al. 2024; Lagidze & Pirtskhalava 2024; Miocevic et al. 2025). For instance, Hannonen (2024) combines survey data with qualitative insights to conceptualise 'emergent geographies' of digital nomadism, while Miocevic et al. (2025) apply both statistical modelling and qualitative interviews to analyse adaptation and food practices. Prengler et al. (2025) integrate ethnographic and organisational perspectives to understand mobility paradoxes in remote work. These examples show that researchers are increasingly interested in triangulating perspectives, combining quantitative representativeness with qualitative depth.

IV) Bibliometric and systematic reviews signal the need to consolidate a rapidly proliferating domain (e.g., Šímová 2023; de Sousa et al. 2025; Başaran 2025; Özyürek et al. 2024).

Summing up, digital nomadism research seems to have moved from definitional/theoretical groundwork towards diversified empirical inquiry, with growing attention to policy (e.g., visa regimes, Koskela & Beckers 2024), urban/real-estate effects (Taltavull 2024; Sequera 2024), taxation and law (e.g., Narotzki & Kuperberg 2023; Arbutina et al. 2023), destination strategies (e.g., Zhou et al. 2024; Bednorz 2024), and gendered/inequality dynamics (Linao et al. 2024; Mancinelli 2020). As previously explained, post-pandemic literature has connected individual

practices to infrastructures and policy regimes more explicitly than pre-2020 work. Identity and lifestyle remain central, but are reframed through regulation (visas, taxation), local impacts (housing, coliving supply), and marketised destination strategies. This thematic shift is accompanied by methodological diversification: alongside qualitative staples, there is more survey-based modelling (motivation, well-being, place attachment, revisit intention), mixed methods, and comparative policy analysis (Isaac & Dorrestein 2025).

To conclude this overview, it should be noted that some contributions more than others have been fundamental for the present research, in two main directions.

First, to define the phenomenon and the identity of digital nomads as well as the nature of their communities, studies addressing conceptual definitions, historical origins, and the broader context of mobility have been especially valuable. Works in the fields of tourism, mobility and sociology studies by Chevtaeva (2024), Reichenberger (2018), Richter & Richter (2020), Prengler et al. (2025), Miocevic et al. (2025), and Zikic et al. (2025) analyse digital nomadism in relation to globalisation, digitalisation, the gig economy and precarious work, also paying particular attention to the COVID-19 pandemic, seen as an accelerator of flexible mobility regimes (Mancinelli & Germann Molz 2023). Research by Woldoff & Litchfield (2021), Hannonen (2020, 2024) and Schlagwein (2017, 2018) provides insights into how digital nomads form, sustain and negotiate their communities of practice. These works focus on everyday life organisation, informal support networks, and the role of online/offline platforms in producing a sense of belonging. They illustrate how digital nomad communities are fluid, open and continuously renewing, but still crucial for knowledge exchange and identity validation. These studies highlight the tension between a will for freedom and autonomy, and precarity typical of the globalised world; they represent the basis for the contextualisation of digital nomadism within 'liquid modernity' (Bauman 2000) provided in Sections 2.2.3 and 2.2.4.

Second, to grasp the specific characteristics of the phenomenon in Europe, several empirical studies have proved fundamental, which analyse the European context, with attention to policy, urban development and local impacts. For example, Rasnača (2023) examines labour rights for mobile workers; Bozzi (2024) discusses infrastructures and city governance; Mohlin & Chatzipanagiotou (2023) analyse digital technology and isolation in European settings; Demaj et al. (2021) focus on Tirana as a potential hub; and Taltavull (2024) explores real estate dynamics. These contributions shed light on the specificities of European digital nomadism,

where mobility intersects with EU cultural, mobility, and linguistic policies, labour markets, urban regeneration, and digitalisation, as will be discussed in Section 2.2.

Because of the phenomenon's nature and novelty, websites are, as previously stated, a relevant source of information – however partial and often marked by a biased positive view of the trend due to their marketing orientation. Reports such as MBO Partners' 'State of Independence' (2025) or popular guides on digital nomad visas (like Pumble[4]) contribute to shaping the narrative on digital nomadism and constitute a direct source of information, even if they are not scholarly literature. For this reason, aligning with investigative practices of many previous works on digital nomadism (e.g., the next section provides an overview of such non-academic information (e.g., Reichenberger 2018).

2.2.2 Digital nomadism on the web

Digital nomadism is a phenomenon that has been extensively shaped and participated in through online contents. Alongside academic research, a large portion of the available knowledge on digital nomads is produced, circulated, and updated on non-academic websites of various kinds. This material includes reports by marketing companies, corporate platforms, community websites, blogs, and networking environments, each contributing to the visibility and understanding of digital nomadism as a lifestyle and a mobility practice.

Commercial and marketing reports play a crucial role in quantifying the phenomenon and making it visible to a wider audience. They offer statistical data, demographic profiles, and trend analyses that are frequently referenced in online articles, consultancy documents, and community websites. One of the most cited reports in academic literature is the MBO Partners State of Independence in America Report (latest version is dated 2024). The relevance of the MBO Partners *State of Independence* report within digital-nomad scholarship is reflected in how frequently it is used as an empirical baseline for interpreting the expansion of remote and independent work. Its figures are often cited to compensate for the absence of official global statistics and to provide a minimum demographic grounding for discussions of nomadic mobility. For example, Hannonen (2024) draws on the report to contextualise the rapid growth of location-independent workers in North America, treating it as indicative of broader global dynamics. According to this survey-based study, the number of self-identified digital nomads

[4] https://pumble.com/learn/digital-nomad-visa/statistics/.

in the United States reached 18.1 million in 2024, a sharp increase from 7.3 million in 2019. The report disaggregates the nomad population by professional categories (independent consultants, freelancers, employees working remotely), age cohorts, gender, and ethnic groups. It highlights motivations such as the pursuit of flexibility, lifestyle improvement, and international mobility, but also barriers like visa restrictions and financial uncertainty. As the authors note, 'digital nomads have moved from the fringe to the mainstream, reshaping expectations of work-life integration' (MBO Partners 2024: 3).

At the international level, the Flatio Digital Nomad Report 2023 provides complementary insights. Based on more than 1,200 responses from nomads in 96 countries, it offers a global overview of mobility patterns and housing practices. The survey identifies Portugal as the most popular destination (27.1%), followed by Spain (12.4%), Thailand (8.7%), and other locations in Southeast Asia and Latin America. Within Europe, Lisbon, Madeira, Barcelona, and the Canary Islands emerge as recurrent hubs, whereas outside Europe Bali, Chiang Mai, and Mexico City are frequently mentioned. The report also emphasises accommodation challenges: while 35% of respondents rely on platforms such as Airbnb, almost 40% report difficulties in finding adequate housing, particularly for mid-term stays. The study further highlights that digital nomads tend to value 'climate, affordability, and internet quality' above other factors (Flatio 2023: 12).

These reports serve multiple functions. On the one hand, they provide quantitative confirmation of the phenomenon's growth; on the other, they contribute to defining its boundaries by identifying key destinations, demographic segments, and infrastructural needs. They also reveal the centrality of Europe – especially Southern Europe – as a hub for digital nomads, situating the continent alongside long-standing expat Asian destinations. In this sense, the Flatio 2023 Report devotes particular attention to Portugal and Spain, confirming their centrality in the European geography of digital nomadism. Within Portugal, Lisbon and Madeira stand out as leading destinations: Lisbon for its urban vibrancy, international connectivity, and cultural offer, and Madeira for its attractive climate, relative affordability compared to mainland cities, and local government initiatives explicitly aimed at attracting remote workers. Similarly, in Spain, Barcelona and the Canary Islands (with Fuerteventura and Gran Canaria in particular) emerge as strategic hubs. Fuerteventura is highlighted for its combination of surfing culture, year-round warm weather, and growing coliving infrastructures, making it especially appealing to younger nomads and those seeking an outdoor

lifestyle. The report suggests that these Southern European destinations function as 'anchor points' in the nomad mobility circuit: they offer a balance of infrastructure (high-speed internet, coworking spaces), lifestyle advantages (climate, leisure), and community presence. Moreover, both Portugal and Spain benefit from relatively accessible residency and visa policies, further consolidating their position as European destinations for international nomads (Flatio 2023: 12–15).

In parallel, the companies providing such reports position themselves as mediators between nomads and services, offering practical tools, legal advice, or housing solutions. They often combine service provision with the communication of values and lifestyle narratives. Indeed, Flatio (https://www.Flatio.com), beyond publishing its annual report, functions as an accommodation marketplace specifically adapted to digital nomads. It offers furnished apartments, simplified rental agreements, and utilities included in the price, responding directly to the needs of mid-term tenants. Indeed, a key component of the online ecosystem is represented by accommodation platforms, which increasingly target digital nomads as a distinct market segment. Flatio, for instance, explicitly brands itself as a 'housing marketplace for digital nomads and remote workers'. Beyond standard rental listings, the platform integrates filters for coliving spaces, mid-term rentals, and verified landlords, signalling an attempt to professionalise the often-precarious housing arrangements associated with mobile workers. The 2023 Flatio Report highlights European hotspots such as Lisbon, Madeira, and Fuerteventura, as well as global hubs like Bali and Mexico City, confirming how accommodation provision plays a central role in consolidating these destinations within the nomad circuit. Importantly, the report frames housing not only as logistics but as a core part of lifestyle construction, linking accommodation choices to values such as community, affordability, and well-being.

Similarly, platforms like WiFi Tribe and NomadX blend accommodation with community values. On WiFi Tribe's website, sections such as 'How it works' emphasise belonging, shared adventures, and co-created experiences, where the house is not just a place to stay but a *collective hub*. As shown in Figure 3, the page features a diagram of community values.

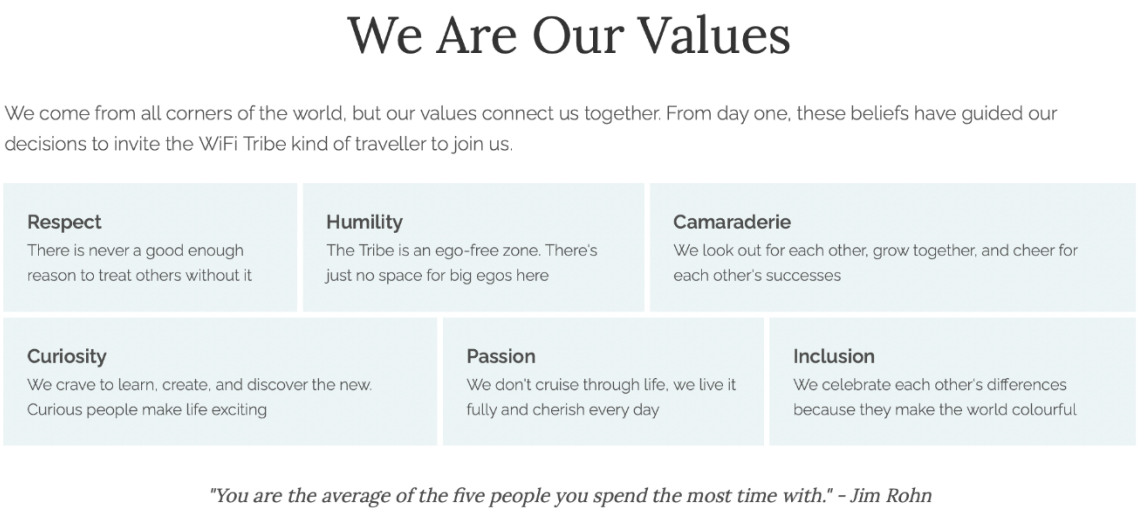


Figure 3. Screenshot of the 'How it works' webpage of Wifi Tribe (as of September 2025).

Nomadx[5], which presents itself as a relocation and housing platform assisting professionals who wish to move and work abroad, integrates city guides, visa information, and accommodation offers. The emphasis is placed on transparency of costs, community integration, and practical support for mobility. Legal Nomads (https://www.legalnomads.com), originally a personal blog by Jodi Ettenberg, demonstrates how individual storytelling can evolve into a widely recognised resource. The platform contains travel narratives, reflections on food and culture, and advice on long-term mobility, including issues of law and health. It illustrates how personal experiences become sources of authority and reference for a broader audience of aspiring nomads. These websites articulate digital nomadism as a lifestyle underpinned by values of freedom, flexibility, and community. Company websites therefore act as both service providers and cultural intermediaries, reinforcing key imaginaries while responding to practical demands.

Another significant portion of the web landscape around digital nomadism is composed of community-driven platforms and blogs, which mix practical resources with cultural narratives. Like the sets of websites explored above, these platforms are not neutral: they present digital nomadism as a desirable lifestyle, often framing it in aspirational and consumer-oriented terms. The homepage of Digital Nomad World[6] is a case in point. The visual layout

[5] https://www.nomadx.com.

[6] https://digitalnomads.world.

combines a background map of the world with a group of diverse young adults walking together, carrying objects that symbolise leisure and creativity (a guitar, a bicycle). This imagery communicates both global belonging and sociability, framing nomadism not just as work-related mobility but as a lifestyle infused with community and freedom. The textual layer reinforces this idea: it highlights 'detailed and up-to-date city guides', 'exclusive discounts' on coliving, coworking, and insurance, and promises opportunities to 'meet like-minded digital nomads.' At the bottom of the page, numbers such as 45,007 members, 170 city guides, and 200+ discounts & benefits act as metrics of legitimacy, giving authority to the platform through quantification. What emerges is a hybrid logic of commerce, community, and expertise, where nomadism is packaged as both a lifestyle product and a social network.

Other platforms adopt similar strategies. Nomads.com (formerly known as The Nomad List, https://nomads.com) ranks cities on variables like cost of living, internet speed, safety, and weather (as shown in Figure 4).

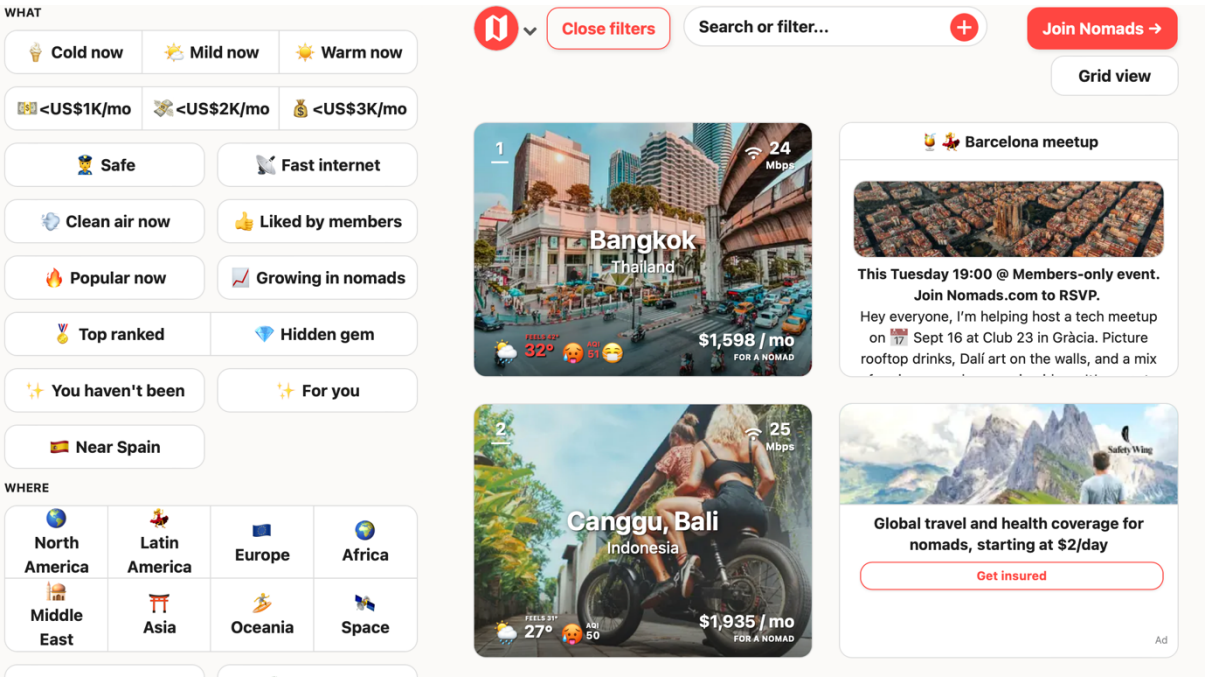


Figure 4. Screenshot of Nomads.com homepage (as of September 2025).

By quantifying cities, it turns mobility choices into a matter of data-driven decision-making, while also offering Slack (a digital communication platform designed for teamwork and group coordination) and Discord (another well-known community app) channels that transform those metrics into opportunities for networked belonging (the use of such networking platforms for

community building and organisation is explored in greater detail in Section 4.4.2). Similarly, Digital Nomads World (https://digitalnomads.world) and Working Nomads (https://www.workingnomads.com) qualify themselves as comprehensive portals: the former by curating guides, visa information, and community events, the latter by aggregating remote jobs. In both cases, nomadism is mediated as an organised lifestyle, supported by structured databases and continuous updates. The visual emphasises principles of respect and contribution to community life, framing digital nomadism not merely as a work arrangement but as a lifestyle embedded in community values. Unlike statistical infographics found in industry reports, this type of visualisation functions normatively: it communicates what it means to 'be a member' of the community and invites users to align their identity with the collective ethos of the Tribe. In these websites, accommodation is not presented as a neutral service but as an entry point into a structured narrative of belonging.

A further example is provided by Remote Work Community (https://www.remoteworkcommunity.com), a networking website explicitly oriented towards remote workers and digital nomads. The platform functions as both a resource hub and a community-building tool, offering articles, curated guides, and opportunities for interaction (an interview with its creator is included as part of the ethnographic data for the present Thesis, and can be accessed on the project webpage https://www.thenomadlinguist.eu/blog/)[7]. Its emphasis is not only on practical aspects but also on community visibility, positioning itself as one of the largest online communities dedicated to remote work. The homepage explicitly describes the network as a 'community for digital nomads, remote workers and entrepreneurs,' highlighting how the categories overlap in practice and reinforcing the idea of nomadism as a hybrid space between professional identity and lifestyle.

Another noteworthy case is the Madeira Digital Nomads project (Startup Madeira 2021), which differs from global platforms by anchoring nomadism to a specific territory. The initiative created the so-called 'Digital Nomad Village' in Ponta do Sol, combining coworking spaces, housing solutions, and cultural events. The website communicates practical resources (visa support, accommodation options, event schedules), but also emphasises values of community and integration with locals, presenting Madeira as a 'home' rather than just a stopover. Unlike global ranking platforms, Madeira Digital Nomads is explicitly tied to regional development policies, using nomadism as a strategy to diversify the local economy.

[7] Further information about the webpage and project's personal branding can be found in Section 4.4.2.

This local model has been replicated in other contexts, such as the Canary Islands (Nomad City Las Palmas, Remote Workers Fuerteventura), where community websites function as destination branding tools as much as actual community hubs. They show how governments and local actors co-opt nomadism into place-based economic development strategies, a dynamic that contrasts with the more global, transnational logic of platforms like Nomads.com.

These sources have been particularly useful to understand the nature and characteristics of digital nomad communities, discussed in Section 2.2.4.

Finally, networking infrastructures extend beyond websites into informal communication channels such as Slack groups, WhatsApp chats, Telegram networks, and Discord communities. While harder to map systematically, these digital spaces act as everyday infrastructures of mobility, where knowledge is exchanged on destinations, visas, or job opportunities. Taken together, these platforms – ranging from formal housing providers to informal group chats – show how the web does not simply describe digital nomadism but actively produces and organises it, transforming a loosely defined mobility practice into a structured, marketable lifestyle (see Section 4.4.2).

2.2.3 Nomadism, globalisation, voluntary mobility and digitalisation

Digital nomadism has emerged as one of the most emblematic phenomena of contemporary globalisation, where the forces of mobility and digitalisation converge to produce new forms of work and lifestyle. Drawing upon the sources above discussed as well as on research on mobility and globalisation, the present sub-section provides a discussion of digital nomadism as a 'phenomenon of contemporary globalisation, where the forces of mobility and digitalisation converge to produce new forms of work and lifestyle' (Chevteava & Denizci-Guillet 2021: 1).

In the 21st century, mobility has emerged as a defining feature of social and economic life, driven by advancements in transportation, communication technologies, and the demands of a globalised economy (Urry 2007). Scholars have described this as part of the 'new mobilities paradigm', which foregrounds not only migration but also lifestyle migration, youth mobilities, and short-term professional relocations (Sheller 2018; Benson & O'Reilly 2009; King 2018). Europe, in particular, has witnessed increasing movements of individuals for work, education, and leisure (i.e., voluntary mobility). According to Eurostat (2022), intra-European migration for professional purposes has surged, alongside a growing segment of individuals engaging in

flexible, location-independent lifestyles facilitated by digitalisation. These mobility patterns are characterised by an elevated 'capital of mobility', defined by Moret as the economic and social freedom to move – or stay – according to one's own will (2020). Importantly, this capital is not equally distributed but stratified by nationality, class, and professional profile, privileging certain passports and socio-economic backgrounds while excluding others (Glick Schiller & Salazar 2013).

This resonates with Bauman's (2000) concept of liquid modernity, where social bonds, careers and memberships are increasingly light, flexible and reversible, encouraging people to keep options open and commitments temporary. Yet, as Favell (2008) notes, the ability to embrace this liquidity is not universal; European mobility regimes create new opportunities for highly skilled professionals while simultaneously restricting access for others through legal and bureaucratic barriers. Liquid mobility thus intersects with policy and rights in non-trivial ways: as remote work stretches across jurisdictions, long-standing couplings between place, obligations and entitlements begin to loosen, raising questions of social citizenship and fairness for both movers and stayers (Webb 2024). The proliferation of digital nomad visas worldwide illustrates these shifts: while governments market themselves as 'talent destinations', critics argue that such measures institutionalise new hierarchies of belonging where rights are increasingly tied to economic capital rather than long-term residence (Dreher & Triandafyllidou 2025).

In Blommaert's (2010) seminal work, 'The Sociolinguistics of Globalisation', the traditional concept of migration, i.e., the movement of people from one territory to another to settle in a new sociocultural environment, is expanded to include cases typical of contemporary times. Migration is thus conceptualised as a never-ending process giving rise to transnational deterritorialised subjectivities. This idea seems to be well exemplified by digital nomads, whose social identity is characterised by online social networks (Ahuja et al. 2020), location-independent work (Bødker 2016), and a substantial mobility capital.

As defined in the literature presented in Sub-section 2.2.1, digital nomads are professionals who work digitally over the Internet to enable a lifestyle of constant traveling and expat living (Reichenberger 2018; von Zumbusch & Lalicic, 2020). Initially coined by Makimoto & Manners in 1997, the concept, like the phenomenon itself, has gained widespread popularity after the Covid-19 pandemic (Ehn et al. 2022). As shown in Figure 5, at the core of digital nomadism are voluntary mobility and digitalisation, as phenomena grounded in contemporary globalisation. Recent organisational and migration research suggests that digital nomadism is

not only a lifestyle label but also a career form and a policy category co-produced by platforms and states (Dreher & Triandafyllidou 2025).

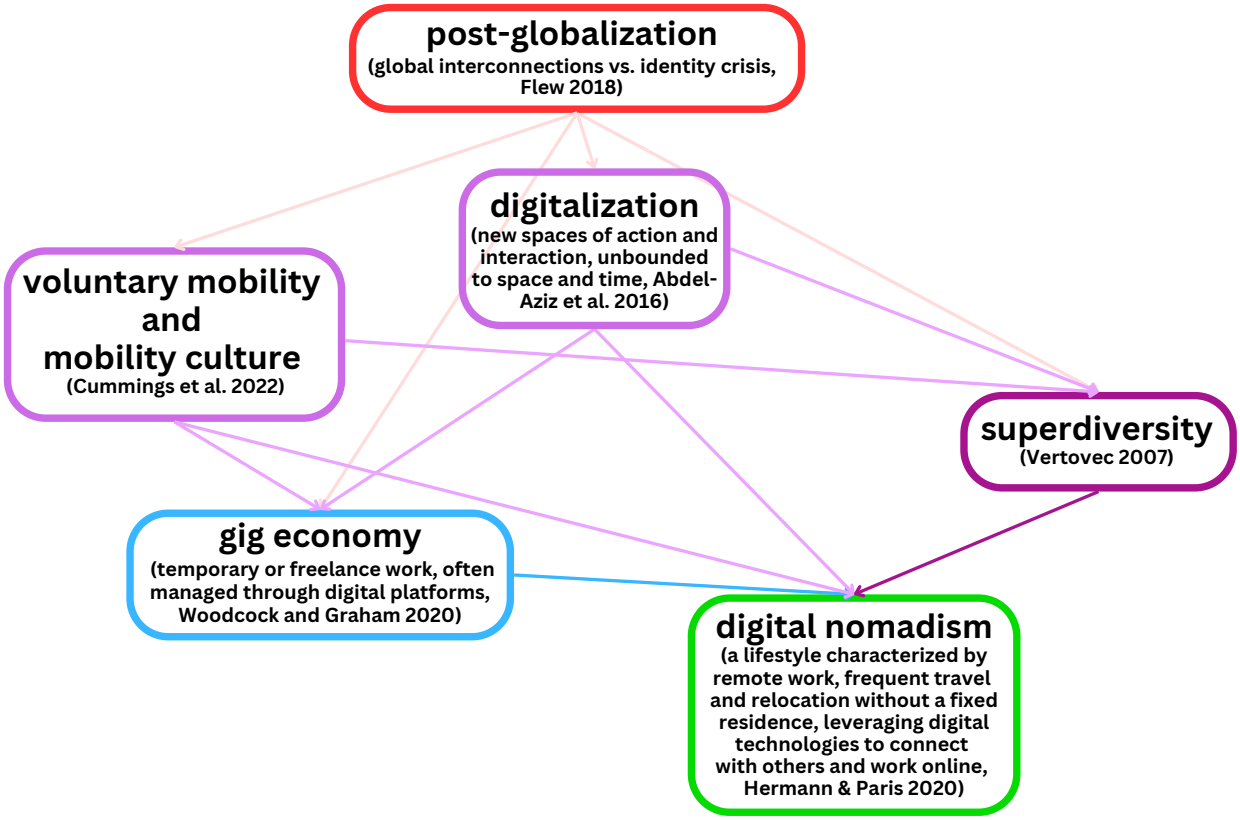


Figure 5. Digital nomadism as product of contemporary globalisation processes.

The concept of *digital nomadism* itself suggests mobility and digitalisation as core elements of the phenomenon; hence, the role of digitalisation also deserves to be discussed. Beyond enabling location-independent work through broadband infrastructures, collaborative platforms and cloud services, digitalisation now increasingly involves automation and artificial intelligence (AI). Recent research demonstrates that AI is reshaping nomad workflows by compressing routine tasks, facilitating asynchronous collaboration, and expanding multi-client management capacities. At the same time, these technologies intensify the 'always-on' expectation, potentially eroding work–life boundaries (Dalisaymo & Patalay 2024). Recent critical research further suggests that digital infrastructures do not merely enable mobility but actively reshape labour geographies. Couldry and Mejias (2019) describe this as 'data colonialism,' whereby platforms extract value from workers' constant connectivity, transforming autonomy into a managed dependency. Similarly, Woodcock and Graham (2019)

argue that digital platforms redistribute labour opportunities across borders while simultaneously deepening competition and precarity. For digital nomads, these dynamics mean that enhanced efficiency and global client access come at the cost of online algorithmic visibility, rating systems, and a constant pressure to remain available online. In this sense, digitalisation does not only provide the material conditions for nomadism but also actively transforms its temporalities and rhythms: mobility becomes tied to digital infrastructures of scheduling, monitoring, and algorithmic mediation.

This duality – enhanced freedom versus deeper platform dependency – mirrors the broader ambivalence of liquid modernity, in which flexibility and precarity evolve side by side (Bauman 2000). Moreover, it should be noted that access to a digital nomad lifestyle remains uneven, with individual circumstances such as nationality, socioeconomic background, and digital literacy playing a significant role in determining who can participate in this mode of hypermobile work.

Summing up, from the perspective of mobility studies, digital nomads appear as privileged subjects who employ new means for resilient success in in a global socioeconomic landscape marked by delocalisation and insecurity (Richter and Richter 2020), represented in Figure 5 by references to the 'gig economy'. Yet this privilege is ambivalent: the physical freedom enabled by mobility coexists with a growing dependence on digital infrastructures, platforms and labour regimes that shape when, how and where mobility can occur. This tension – between autonomy and infrastructural reliance – echoes broader global dynamics, although its manifestations vary significantly across geographical and social contexts.

Seen from the perspective of tourism studies, digital nomadism is increasingly framed not only as a mobility lifestyle but also as a form of tourism-driven migration that reshapes local economies and urban fabrics. Research has shown that destinations such as Lisbon, Barcelona, Bali, and Mexico City actively brand themselves to attract nomads, positioning them alongside long-term tourists and lifestyle migrants (Orel 2019; Hannonen 2020; Richards 2021). This influx contributes to the visibility and internationalisation of destinations but also raises concerns about gentrification of urban centres and housing pressure. Studies conducted in Southern Europe demonstrate that short-term, flexible residents such as digital nomads overlap with tourism flows , thereby intensifying competition in local rental markets (Sequera & Nofre 2019; Taltavull 2024). The result is often a dual process: on the one hand, local authorities and businesses highlight the economic benefits of attracting skilled, mobile professionals; on the

other, residents experience rising housing costs, neighbourhood transformation, and the displacement effects familiar from urban tourism research. This ambivalence has led scholars to conceptualise digital nomads as both 'desirable tourists' and 'unsettling migrants' (Cohen et al. 2015; Mancinelli 2020). As Germann Molz (2021) argues, nomadism should be understood as part of a continuum with tourism mobilities, where the boundaries between leisure and work blur, and where the socio-economic impacts on host communities cannot be detached from broader debates on sustainable tourism and urban justice. These processes are especially visible in metropolitan hubs such as Lisbon and Barcelona, where digital nomads intersect with lifestyle migrants and tourists, reinforcing dynamics already studied in urban tourism and gentrification research (Sequera & Nofre 2019; Cocola-Gant et al. 2020).

Policy frameworks such as digital nomad visas are often justified as economic diversification strategies in fragile island economies, where dependence on tourism makes local societies particularly sensitive to mobility-driven transformations (Webb 2024; Dreher & Triandafyllidou 2025). The representation of digital nomads in hosting community narratives is thus controversial, especially in destinations actively branded as nomad hubs such as Madeira and the Canary Islands. Qualitative research highlights an ambivalent reception. Local actors see nomads as a source of economic diversification, particularly valuable in post-pandemic recovery contexts where tourism economies were severely disrupted (Orel et al. 2021).

In Madeira, initiatives like the Ponta do Sol 'Digital Nomad Village' were presented by policymakers and entrepreneurs as strategies to attract international talent, boost co-working infrastructures, and reposition the island within global mobility circuits (Startup Madeira 2021). Local residents often acknowledge the economic benefits but also point to cultural distance, a sense of temporariness, and worries about housing affordability (Demaj et al. 2021).

In the Canary Islands, ethnographic accounts show that while digital nomads are welcomed as modern, cosmopolitan visitors, they are also associated with rising rents, spatial segregation in co-living enclaves, and the reproduction of inequalities between highly mobile foreigners and precarious local workers (Sequera & Nofre 2019). This reflects a broader tension: while destination marketing discourses emphasise community integration and *win-win* models, everyday interactions in host societies often reveal asymmetries in power, privilege, and belonging (Nowicka & Ryan 2015; Mancinelli & Germann Molz 2023). The Canary Islands offer a telling example of how local communities are negotiating the rise of digital nomadism

in tandem with long-standing concerns about mass tourism. In 2025, protests erupted across the archipelago under the slogan 'Canarias tiene un límite' (transl. 'the Canary Islands have a limit'), gathering thousands of residents demanding stricter regulations on tourism, vacation rentals, and real estate speculation (RTVE 2025). While local authorities, such as the Cabildo of Tenerife, have simultaneously promoted the arrival of digital nomads as part of a diversification strategy, segments of the population perceive them as contributing to rising rents and the shortage of affordable housing (eldiario.es 2024). Journalistic accounts stress the ambivalence of this reception: nomads are framed by policymakers as skilled, cosmopolitan visitors who can support the local economy year-round, yet for many residents they epitomise the same structural pressures brought by mass tourism – housing displacement, neighbourhood change, and the precarisation of everyday life. This contradiction is sharpened by state-led initiatives: while regional authorities frame digital nomads as 'desirable migrants' to be courted through targeted visas and incentives, residents often experience their arrival as part of the same structural pressures associated with mass tourism, namely housing scarcity and rising costs (Bozzi 2024; Dreher & Triandafyllidou 2025).

This tension illustrates how, in fragile island economies, digital nomadism is entangled with broader debates on sustainability, the right to housing, and urban justice (for a broader discussion on digital nomad communities in Europe see Section 2.2.2).

The experiences of both global cities and insular economies highlight how digital nomadism is embedded in uneven geographies of globalisation, where cosmopolitan imaginaries confront local contestations and structural inequalities. In this sense, digital nomadism exemplifies the contradictions of post-globalisation: celebrated as an engine of innovation and cosmopolitanism, yet contested as a driver of urban pressure, gentrification, and social differentiation in fragile island economies. While a comprehensive discussion on these issues would take us too far, the social mobility perspective allows for the identification of some essential characteristics of digital nomadism, including cultural – and hence linguistic – fluidity.

The *mobile-worker paradox* discussed in mobility-focused studies is further highlighted by organisational studies: mobility promises freedom, yet sustainable work often requires *placefulness* (routines, anchors). Nomads navigate this paradox differently depending on their 'wanderlust' dispositions and infrastructural supports (Prengler et al. 2025). Quantitative and qualitative work on place identity and socio-cultural adaptation indicates that repeated moves

do not erase place bonds; nomads layer multiple attachments and manage belonging and well-being across locales (Lagidze & Pirtskhalava 2024; Miočević et al. 2025). This complex socio-emotional layering is mirrored in individual voices and communicative practices (Tedesco 2025a). Moreover, as an evident example of post-globalisation, denoted by hyperconnectivity and boundless communication, the phenomenon raises important interrogatives about linguae francae and plurilingualism which inspired the present research work.

Signals of normalisation come from market and policy data: in the United States, 18.1 million workers self-identified as digital nomads in 2024 (≈11% of the workforce), while more than 60 countries have introduced visa pathways targeting remote workers, suggesting that platforms and governments are co-evolving the field (MBO Partners 2024; Global Policy Report 2024).

Thus, digital nomadism nowadays appears to have spread globally. To understand the societal changes which created a fertile ground for the phenomenon to grow exponentially, it can be useful to recall the concept of globalisation. Globalisation can be holistically described as 'the spread of transplanetary connections' (Scholte 2007: 3). More recently, the concept of post-globalisation has been proposed to describe the tension between global interconnectedness and emergent crises of identity and localisation (Flew 2018). Post-globalisation highlights a world that is increasingly hyperconnected while simultaneously grappling with fragmentation, uncertainty, and cultural diversity. Placed in this frame, digital nomadism is a post-globalisation formation: hyperconnected, platform-mediated and deterritorialised, yet constantly negotiating local frictions and identities across uneven geographies (Dreher & Triandafyllidou 2025).

From a sociolinguistic perspective, Blommaert and Rampton (2011) and Arnaut et al. (2016) have noted that globalisation processes lead to greater sociolinguistic diversity, applying the concept of *superdiversity*, i.e. 'the dynamic interplay of variables among an increased number of […] scattered, multiple-origin, transnationally connected, socio-economically differentiated and legally stratified immigrants' (Vertovec 2007: 1024), to the study of languages and sociolinguistic systems. In Figure 5, digital nomadism is thus located at the intersection between digitalisation, globalisation, social and economic insecurity (and mobility), and digital nomad communities are presented as examples of superdiverse social ecosystems (see Section 2.2.4). A recent migration-studies synthesis explicitly ties digital nomadism to neoliberalism and postcoloniality, proposing a three-level framework (micro: practices, meso:

platforms/visas, macro: regimes) to analyse how privileges and constraints are organised (Dreher & Triandafyllidou 2025). In parallel, a career-netnography of Reddit discourse shows that online communities circulate practical *knowing-how/whom* scripts that codify mobility as shareable know-how. These findings underscore how digital nomadism materializes superdiversity in practice, and validate the investigation of digital nomad communities as empirical lenses on mobility, work and identity, and, crucially, on language in superdiverse, liquid settings.

2.2.4 Socio-cultural features of digital nomad communities

Digital nomads thus seem to combine traits of traveling professionals, freelancers, telecommuters and neo-nomad backpackers (Hannonen 2020). Thus, while remote-working, the deliberate adoption of a nomadic lifestyle and an intense use of communication technologies represent some of the key defining features of digital nomads, self-representation (Bonneau et al. 2023) and community building (Woldoff and Lichtfield 2021) are dominant behavioural aspects. Indeed, few digital nomads fully embrace an individualist lifestyle associated with continuous travel; most seek to share their nomadic experiences with 'like-minded' people, leading to an increasing number of structured communities around the world – and the web (Woldoff and Lichtfield 2021). In doing so, they form communities that – though often temporary, dispersed, and heterogeneous – become crucial sites of belonging and support, settings for transcultural encounters, knowledge and identity co-creation (Tedesco 2024a).

Providing a clear definition of digital nomad communities is a rather difficult task; the idea of nomad communities may even sound paradoxical. At first sight, digital nomads do not appear to share a common traditional cultural or social background, if these notions are understood in relation to their place of origin, familiar histories, nationally-rooted socialisation practices. Despite this heterogeneity, they nonetheless develop strong forms of belonging and identification within the communities they create while on the move (Woldoff and Lichtfield 2021). The definition here proposed of digital nomad communities is based on Blommaert and Varis' concept of light communities, 'focused but diverse occasioned coagulations of people' arising 'in the offline-online contemporary world' (2015, 54, 70). The presented conceptualisation of digital nomad communities is also enriched by the concepts of digital community (Kindsmüller et al. 2009), and community of practice (Iverson 2010). While traditionally applied to professional communities (Wenger 1998), recent studies like Li et al. (2009) and De Paiva Duarte (2013) have broadened the use of the term 'community of

practice', focusing on the co-construction of a horizontal knowledge sharing infrastructure, which seems to be a term that well represents the digital networks of information sharing, the use of coworkings and the organisation of cultural, educational and professional-development initiatives within the local communities.

Before further discussing the definition and characteristics of digital nomad communities, a terminological note is in place that can also help clarify the boundaries within which the term 'digital nomad community' can be applied. The terms digital nomad and remote worker are often used interchangeably in public discourse and even in some academic sources. This conflation derives from the shared reliance on digital infrastructures and the capacity to perform work at a distance (Cook 2023). Yet there are important distinctions. *Remote work* refers broadly to employment arrangements in which tasks are performed outside of a traditional office, often from a fixed base such as one's home (Messenger & Gschwind 2016). Many remote workers may never move abroad or participate in nomadic hubs. Digital nomadism, by contrast, embeds mobility and travel into the very definition of work-life organisation: mobility is not incidental but constitutive of the lifestyle (Chevtaeva & Denizci-Guillet 2021; Peña Pérez 2025). Research demonstrates that digital nomads frame mobility as a value in itself, intertwining work and travel as an identity practice, whereas remote workers can remain sedentary while still benefiting from digitalisation (Dreher & Triandafyllidou 2025).

The peculiarities of digital nomad communities become even clearer when contrasting them with expatriate (or *expats*) communities. Expat communities are not new: anthropological and sociological research since the 1970s has described expatriates as privileged migrants, often moving under corporate, diplomatic, or development contracts (Fechter & Walsh 2013; Kunz 2016). Such communities are typically more stable and locally embedded, often organised around national identity, international schools, clubs, or diplomatic infrastructures (Nowicka & Ryan 2015). They provide strong support networks, but are relatively closed, relying on shared socio-economic and cultural backgrounds.

Digital nomad communities differ in several respects: (i) temporality: while expat communities tend to form long-term enclaves, nomad communities are short-lived, rotational, and constantly renewing (Hannonen 2020); (ii) identity: expats often define themselves in relation to nationality, while digital nomads foreground lifestyle, mobility, and professional identity (Cook 2023); (iii) infrastructure: expat communities rely on formal institutions (corporations,

embassies), while nomad communities rely on informal, hybrid infrastructures such as coworking spaces, coliving hubs, and online forums (Orel et al. 2021; Woldoff & Litchfield 2021); (iv) communication, culture and language: expats often reproduce linguistic and cultural homogeneity, whereas digital nomad communities exemplify superdiversity and plurilingual interaction (Tedesco 2025a). To make this difference sharper, thorugh netnography (see Section 4.4) digital nomads have been observed to consciously distance their selves from the typical *expat* profile. In Figure 6, a screenshot of a post on the Instagram social media channel dedicated to digital nomads named 'Freaking Nomads' explain in simple terms the difference between these two lifestyles.

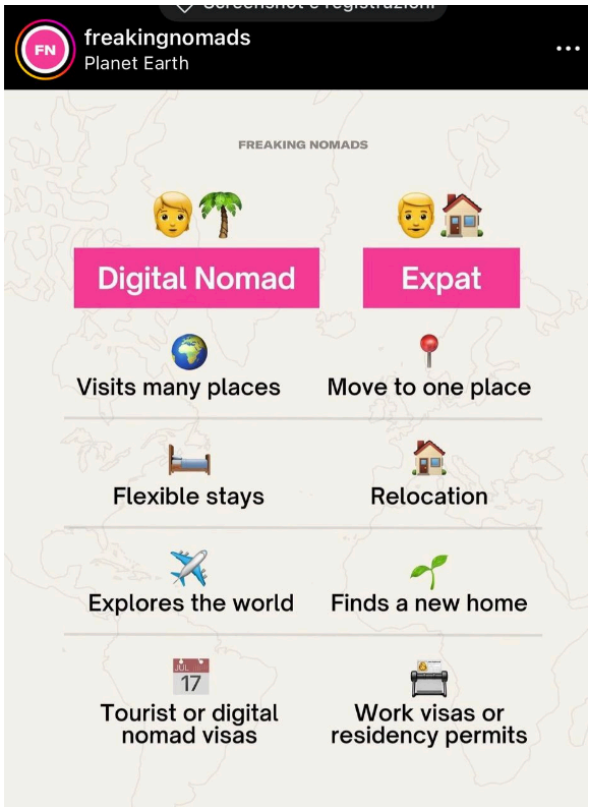


Figure 6. Comparison between digital nomads and expats in an Instagram post by Freaking Nomads.

Nevertheless, the comparison with expat communities situates digital nomadism within a longer genealogy of mobile professional communities. Expat enclaves, corporate 'colonies', and lifestyle migrants have long demonstrated how mobility, privilege, and identity intertwine (Fechter & Walsh 2013; Benson & O'Reilly 2009). Lifestyle migration studies, in particular, show how middle-class professionals seek meaning and identity through relocation to attractive destinations (Benson & Osbaldiston 2014). Digital nomads inherit some of these traits but

reframe them around digital infrastructures and the ideology of flexibility. Their communities are less tied to nation and more to global imaginaries of freedom, self-actualisation, and networked belonging.

Summing up, while some overlaps are observed between the terms *remote work* and *digital nomadism*, and while both share continuities with expat communities, digital nomad communities represent a novel hybrid formation. They combine the fluidity of mobility with the intentionality of community building, positioning themselves as superdiverse, transcultural, and digitally mediated collectives that diverge from the bounded, nationality-oriented models of expatriate enclaves.

Indeed, recent studies stress that belonging in nomad communities is hybrid, mediated by both material infrastructures (coworkings, colivings, cafés) and intangible ones (digital platforms, affective ties). Bonneau et al. (2023) highlight self-representation as a dominant behavioural practice, while Woldoff and Litchfield (2021) emphasise community building. The sense of belonging to places, however, is fragile: nomads constantly negotiate attachment to temporary homes while sustaining transnational ties (Pentina et al. 2023). This reflects Blommaert's (2010) notion of deterritorialised subjectivities and Vertovec's (2007) superdiversity applied to mobile elites.

Hence, a final definition digital nomad communities describe them groups of remote-workers who interact with each other using online communication infrastructures, thanks to which they also organise themselves into local communities who live and work in the same places, actively participating in some or all aspects of community life. Besides their activity within the digital labour market, digital nomad communities are *digital* because they are based on social media communication. They are *intentional* because belonging to the community is agentively chosen by all members, who gather around nomad hubs (Thomposon 2019), where they put into *practice* a certain lifestyle based on the blending of leisure and work-related activities (Hannonen 2020), and on the development of non-formal and informal infrastructure for knowledge sharing. The latter can be online forums as well as workshops and meetings of different kinds at coworkings and colivings, where information about hot digital nomad topics, such as VISA, taxation, personal branding, business digitalisation and AI are shared (Tedesco 2024a). In Figure 7, a representation of digital nomad communities at the intersection of agency, local practices and online interaction is provided.

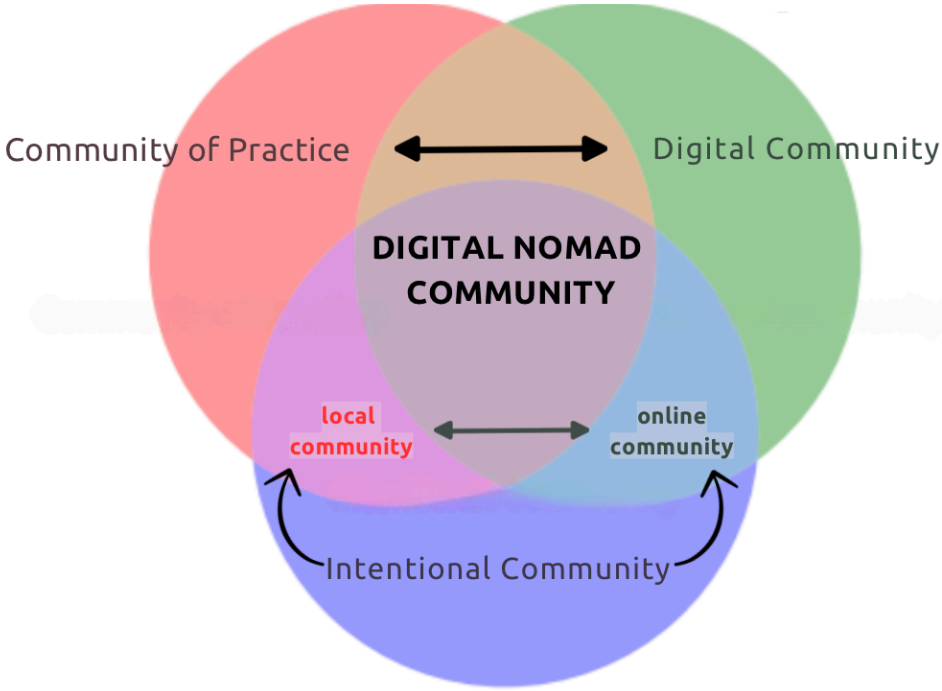


Figure 7. Conceptual representation of digital nomad communities.

Digital nomad communities are essentially digital: they live in what Orel et al. (2021) call 'digital trans-spaces'. Platforms such as Slack, WhatsApp, Reddit, Instagram pages, and Facebook groups (e.g. Digital Nomads Europe) are central infrastructures. Dedicated websites like Nomad List (nomadlist.com), Remote Work Community (remoteworkcommunity.com), WiFi Tribe (wifitribe.co), or Nomads.com provide both knowledge resources (visas, taxation, branding, AI) and a sense of imagined togetherness (see Section 2.2.2). For instance, WiFi Tribe (wifitribe.co) frames itself explicitly as a tribe of 'like-minded explorers,' offering not only accommodation and coworking but also a lifestyle of curated sociality, emphasising belonging and shared values. Similarly, Coworking Europe (coworkingeurope.net) stresses how coworking hubs double as sites of both productivity and community, echoing Orel's (2019) idea of coworking as 'civic justification' for digital nomadism.

These vast and constantly growing, digital communities double up as territorial communities, in which the intense use of social networks supports local practices, events, co-living and co-working experiences (Orel et al. 2021). Blurring the boundaries between global and local citizenship, these communities embody a new living concept, based on self-organisation, plurality and inclusivity. Thus, an essential feature of digital nomad communities, which leads to such a great diversity in cultural and linguistic repertoires, is their openness, guaranteed by a system which allows anybody who seeks social identification, advice and support from the

community to become a member – provided they retain the needed capital of mobility[8]. And it is precisely this openness that foregrounds the possibility for such a great diversity of cultural and linguistic repertoires to co-exist and blend within the same community network.

2.3 Zooming in on the European Context

2.3.1 Delimiting geographical research focus to the EU territory

The choice of Europe as the primary context of investigation in this study is motivated by a number of structural, cultural, and policy-related considerations. As discussed in Section 2.2, digital nomadism is a global phenomenon emerging at the intersection of mobility, digitalisation, and new forms of work. Yet its manifestations differ considerably across regions. Europe represents a particularly instructive case because the socio-economic inequalities and housing crises associated with digital nomadism are generally less acute than in many parts of the Global South. For example, while gentrification and housing pressure in Southern European destinations such as Lisbon, Barcelona, and the Canary Islands have sparked debate and even protest (Sequera & Nofre 2019), these challenges occur in contexts that still provide comparatively strong social protections and regulatory frameworks[9]. By contrast, in Asian destinations such as Bali, Chiang Mai, or parts of the Philippines, scholars have noted forms of digital-nomad-driven neo-colonialism, where local economies are profoundly restructured around transient, often privileged foreign residents, with limited redistributive policies in place (Qosim et al. 2025).

[8] Although this study does not aim to offer a critical analysis of privileged mobility per se, but rather to investigate emerging communicative practices within specific pluricultural contexts, it cannot remain detached from the broader inequalities that structure contemporary global mobility. In a world marked by war, economic disparity, and inhumane living conditions, mobility is not a neutral or equally accessible experience: what may be framed as travel, choice, or opportunity for some is, for others, a forced and often life-threatening journey. As several scholars have noted, access to mobility is deeply stratified, and even seemingly basic factors—such as nationality or passport ownership—can radically transform the meaning and consequences of movement across borders (Bauman 1998; Sheller & Urry 2006; Sheller 2018).
With this awareness in mind, the present research is conceived not as a celebration of mobility, but as a contribution to more reflexive and socially responsive approaches to language education. By foregrounding communicative practices in transcultural settings, the study aims to inform the design of educational projects that foster transcultural awareness, critical linguistic reflexivity, and sensitivity to inequality among younger generations. In doing so, it seeks to respond to the concrete communicative needs of contemporary global citizens, while supporting broader efforts toward cultural and linguistic inclusivity.

[9] For example, the Spanish Ley de Ordenación Sostenible del Uso Turístico de Viviendas (Law on Sustainable Planning of Tourist Use of Housing), approved on November 12, 2025, which freezes new VV licenses in Canary Islands for up to five years, requiring strict compliance with local zoning and urban plans.

Europe also constitutes a privileged lens because of its explicit political and legal commitments to cultural and linguistic diversity. Since the 1990s, the European Union and the Council of Europe have developed an elaborate policy infrastructure aimed at promoting plurilingualism, pluriculturalism, and inclusive citizenship (Council of Europe 2001; European Commission 2017). These commitments are not simply rhetorical: they are inscribed in binding treaties, language education frameworks such as the Common European Framework of Reference for Languages (CEFR), and migration policies that frame mobility as both an economic resource and a social challenge to be managed inclusively (Barni & Extra 2008; Krzyżanowski & Wodak 2011). This normative environment is crucial for understanding digital nomadism in Europe, since the phenomenon is not only shaped by market logics and platform affordances but also by policy regimes that foreground transcultural communication, plurilingual repertoires (see Sections 2.2.2 and 3.2), and the rights of mobile citizens.

The European context therefore offers both an empirical and a normative vantage point**.** Empirically, Europe is one of the global hotspots of digital nomadism, with destinations ranging from urban hubs such as Berlin, Barcelona, and Lisbon to island economies like Madeira and the Canary Islands (Demaj et al. 2021; Parreño-Castellano et al. 2022). Normatively, Europe provides a political and discursive environment that foregrounds diversity, integration, and social rights, allowing for the investigation into how digital nomads interact with, and sometimes challenge, existing ecologies of language and culture.

The discussion in the remainder of this chapter is thus organised into two main parts. Section 2.2.2 examines in depth the European policy orientation towards pluriculturalism and plurilingualism, with particular emphasis on the CEFR and related Council of Europe and EU documents. Section 2.2.3 then addresses the phenomenon of digital nomadism in Europe, by situating it within the broader context of European mobility regimes, labour markets, and tourism geographies.

2.3.2 EU: Towards an ecology of cultures and languages

The European Union has long positioned itself as a space where linguistic and cultural diversity is not merely tolerated but actively promoted as a cornerstone of integration. The EU's motto, 'United in Diversity', captures a normative orientation towards what can be described as an ecology of cultures and languages, in which multiple linguistic repertoires and cultural affiliations coexist within shared institutional frameworks (European Union 2007). This orientation is not an accidental product of enlargement or migration but a deliberate outcome

of decades of language and education policy, crystallised in working and official documents, the most important of which is probably the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001, 2020; Krzyżanowski & Wodak 2011).

From the early 1990s onwards, European institutions have recognised that the sustainability of the Union requires tools to mediate cultural diversity while fostering common citizenship. This was particularly urgent after the Maastricht Treaty (1992), which strengthened the principle of European citizenship. Language was identified as both a barrier to mobility and a resource for cultural bridging, leading to a surge of initiatives under the EU's multilingualism policy (Krzyżanowski & Wodak 2011; Barni & Extra 2008). These initiatives stressed that linguistic diversity is not a problem to be solved but an asset to be cultivated. As a result, the EU progressively shifted from notions of 'multilingualism' (the coexistence of many languages) to plurilingualism, defined as 'the ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent, has proficiency of varying degrees in several languages and experience of several cultures' (Council of Europe 2001: 168).

The CEFR is central to this agenda. Conceived as both a pedagogical and political instrument, it frames language proficiency not in terms of static competence but as a dynamic repertoire. Its descriptors (A1–C2) are widely used in education systems and professional settings, but the framework also embeds a philosophy of transcultural communication. In the CEFR's companion volume (Council of Europe 2020), the concept of *pluricultural competence* is explicitly articulated, referring to the ability of learners to draw on multiple cultural frames of reference in interaction. This orientation aligns with the broader EU ambition to enhance mobility and transcultural spaces (Byram 2021).

The EU's promotion of plurilingualism is not limited to education. Policy documents consistently frame it as a driver of labour mobility and economic competitiveness. The 2007 European Commission Communication 'Multilingualism: An Asset for Europe and a Shared Commitment' (COM(2007) 566 final) argued that 'linguistic diversity is an integral part of European identity' and that citizens should be enabled to learn at least two foreign languages in addition to their mother tongue. This *mother tongue plus two* principle remains a guiding objective, reiterated in the 2017 Communication 'Strengthening European Identity through Education and Culture' (European Commission 2017). In this vision, plurilingual and

pluricultural competencies are prerequisites for full participation in the European labour market and in democratic life.

At the same time, EU institutions are aware that mobility regimes generate new tensions around migration, identity, and inclusion. Krzyżanowski and Wodak (2011) note that multilingualism policies often operate in tension with securitisation discourses, especially in relation to third-country nationals. While the CEFR and related initiatives foreground inclusivity, migration policies frequently demand rapid integration and linguistic assimilation. For example, several Member States have introduced mandatory language tests for residency and citizenship, justified as mechanisms of integration but criticised for functioning as gatekeeping tools (Extra & Spotti 2009).

Nevertheless, the EU's overarching discourse remains one of pluricultural dialogue. The 2008 European Year of Intercultural Dialogue and subsequent Council Conclusions explicitly stressed the role of languages as mediators of social cohesion. The notion of transcultural communication**,** which will be addressed from a theoretical perspective in Section 3.2, is central to the CEFR, and also echoed in the Council of Europe's *White Paper on Intercultural Dialogue* (2008), where intercultural competence is framed as essential to democracy in culturally plural societies.

These policy frameworks matter for the study of digital nomadism in Europe for at least two reasons. First, they establish an institutional environment where mobility and linguistic diversity are normalised**,** providing fertile ground for the reception and regulation of hypermobile groups such as digital nomads. Second, they create expectations that mobile populations should engage with local cultures and languages not merely as consumers but as participants in a pluralistic ecology. This is significant when comparing Europe with other contexts: whereas Asian destinations often frame digital nomads primarily as economic assets, with limited attention to cultural integration, the European context explicitly foregrounds language and culture co-creation as principles for both traveling subjects and hosting communities (Qosim et al. 2025).

Summing up, the European Union project can be described from a sociocultural perspective as an attempt to construct an ecology of cultural and linguistic diversity: an environment where mobility is framed as both a right and a responsibility, and where plurilingualism is cultivated as a core competence for citizens. The CEFR and related policies operationalise this vision,

embedding pluriculturalism into education systems, migration frameworks, and labour policies. As the next section shows, these orientations are directly relevant to understanding how digital nomadism unfolds in Europe, where hypermobile professionals encounter a policy environment that celebrates diversity while grappling with the practical tensions of integration and social cohesion.

Building on the policy and conceptual background outlined above, it is important to examine how the EU and Council of Europe concretely articulate pluriculturalism and plurilingualism, especially in relation to migration, mobility, and intercultural communication.

The CEFR (2001) and its Companion Volume (2020) go beyond linguistic competence by explicitly embedding pluricultural and plurilingual competence into their descriptors. This reflects a major conceptual shift in language education: learners are not expected to become 'native-like' speakers of a second language, but to develop the ability to mobilise different linguistic and cultural resources in interaction. The CEFR defines plurilingual competence as 'the ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent, has proficiency of varying degrees in several languages and experience of several cultures' (Council of Europe 2001: 168).

The Companion Volume (2020) further expands on this by introducing detailed descriptors for mediation and interculturality. Mediation descriptors cover activities such as paraphrasing, summarising, and facilitating understanding between speakers of different languages or from different cultural backgrounds. Intercultural descriptors foreground the ability to reflect on cultural assumptions, to manage misunderstandings, and to engage in dialogue across differences (Byram 2021). Together, these innovations position plurilingualism not simply as a skill but as a social practice: the capacity to navigate superdiverse environments where languages, identities, and values overlap and conflict.

This orientation has been very influential. National curricula in several EU member states (e.g. France, Germany, Spain, and Italy) have adopted CEFR descriptors, shifting classroom practice from purely grammatical instruction to communicative and intercultural goals (North & Piccardo 2016). Moreover, the CEFR has informed professional mobility instruments, such as the Europass Language Passport, which allows workers to document language skills in a standardised, recognisable format. For mobile professionals, including digital nomads, such

frameworks are crucial: they legitimise plurilingual repertoires as professional capital and provide symbolic recognition of intercultural skills.

While education policies celebrate plurilingualism, migration and integration policies often introduce more complex dynamics. As Extra and Spotti (2009) note, many EU member states have implemented mandatory language and civic integration tests for long-term residence or citizenship. These tests frequently draw on CEFR levels (typically A2 or B1) as benchmarks. For example, the Netherlands requires A2 proficiency for permanent residence and B1 for citizenship; Germany demands similar levels. These measures illustrate the dual role of the CEFR: it is both an enabler of mobility and a gatekeeping instrument.

This ambivalence reflects broader tensions in EU mobility regimes. On the one hand, plurilingualism is framed as a right and resource, empowering citizens and migrants to participate in plural societies. On the other, it becomes a responsibility, used to filter 'desirable' from 'undesirable' migrants (Krzyżanowski & Wodak 2011; Van Avermaet 2009). This duality highlights how Europe's ecology of languages is not neutral but shaped by power, hierarchy, and political priorities.

Importantly, the EU has also attempted to reconcile these tensions by foregrounding intercultural dialogue. The Council of Europe's *White Paper on Intercultural Dialogue* (2008) frames linguistic diversity as a condition for democratic participation, while EU directives on anti-discrimination (e.g. Directive 2000/43/EC) stress equal treatment irrespective of cultural or linguistic background. Nevertheless, the practical application of these principles often depends on national implementation, which varies considerably.

Plurilingual and pluricultural policies in Europe have direct implications for digital nomads. First, they shape the institutional environment: nomads arriving in Europe encounter destinations that valorise linguistic diversity and offer infrastructures (language courses, multilingual services, international schools) designed to support mobility. Second, they shape identity narratives: digital nomads can frame themselves as not just economic actors but as participants in a European ecology of languages and cultures, aligning with policy ideals of pluricultural citizenship.

However, as was discussed in Section 2.2, reality is often more complex. As research on host communities in Lisbon, Barcelona, and Madeira shows, digital nomads are frequently perceived as outsiders who remain socially and linguistically distant (Sequera & Nofre 2019;

Sardinha et al. 2023). While policy discourses celebrate diversity, everyday encounters reveal asymmetries in who is expected to adapt, learn local languages, or participate in community life. The European emphasis on plurilingualism thus sets high normative expectations, but digital nomads often rely on English as a lingua franca, reproducing inequalities between globally mobile elites and local residents (Holliday 2018).

The EU policy frameworks construct a distinctive ecology of cultures and languages, rooted in the CEFR and in broader commitments to pluriculturalism and plurilingualism. While these frameworks open space for intercultural dialogue and provide symbolic resources for mobile professionals, they also operate as instruments of governance and stratification. The tension between empowerment and gatekeeping is particularly salient in the case of digital nomads, who embody the ideals of mobility and linguistic diversity but also expose the limits of Europe's integration model. The next sub-section explores these dynamics further by analysing how digital nomadism unfolds specifically within the European context.

2.3.3 Digital nomadism in Europe

Digital nomadism in Europe represents both a continuation of broader global trends and a distinctive regional manifestation shaped by the continent's historical trajectories, policy frameworks, and socio-economic environments. In Europe, digital nomadism intersects with long-standing traditions of mobility – such as Erasmus student exchanges, intra-EU labour migration, and lifestyle migration.

Empirical studies consistently identify Europe as one of the most important global hubs for digital nomads (Flatio 2023; MBO Partners 2024). This attractiveness is reinforced by a series of structural, economic, and cultural factors that make Europe a particularly fertile ground for remote work–driven mobilities. Within Europe, three geographical clusters stand out.

First, urban centres such as Berlin, Lisbon, Barcelona, and Tallinn attract digital nomads because of their connectivity, cultural life, and reputation as creative hubs (Richards 2021; Bozzi 2024). These cities have positioned themselves as international nodes in the knowledge economy, offering coworking spaces, start-up ecosystems, and digital infrastructures that appeal to globally mobile professionals.

Second, Southern Europe and island destinations – notably Portugal (Lisbon, Porto, Madeira), Spain (Barcelona, Canary Islands), Croatia (Split, Dubrovnik), and Greece – have emerged as

highly visible hotspots (Demaj et al. 2021; Sequera & Nofre 2019). Their attractiveness lies in a combination of relatively low cost of living compared to Northern Europe, favourable climate, and deliberate policy initiatives aimed at diversification beyond mass tourism. For example, the Ponta do Sol 'Digital Nomad Village' in Madeira, launched in 2021, was explicitly designed to attract remote workers, combining coliving and coworking facilities with events fostering community engagement (Startup Madeira 2021). Similarly, the Canary Islands government has branded the archipelago as 'the European hub for remote workers,' emphasising year-round sunshine and advanced connectivity. This marketing shift has been embraced also by accommodation providers. For example, a picture taken at the entrance of the coliving Surfing Colors in Fuerteventura, shown in Figure 8, captures the use of the word 'no-turista' / 'no-tourist' to address the guests of the structure.

Figure 8. 'Hola no-turista': sign on the front door of a coliving in Fuerteventura.

Third, peripheral and emerging hubs in Central and Eastern Europe are gaining traction. Tallinn, already known for its e-residency programme, has become a symbolic destination for location-independent entrepreneurs (Vesala 2020). Croatia introduced a digital nomad visa in 2021, while other countries such as Romania and Hungary have considered similar schemes. These initiatives often serve a dual purpose: attracting foreign income and talent, and rebranding national economies as innovation-friendly.

A defining feature of European digital nomadism is the proliferation of visa pathways and policy experiments. By 2025, more than 20 European countries had introduced digital nomad or remote work visas, including Portugal, Spain, Greece, Estonia, Croatia, and Malta (Dreher & Triandafyllidou 2025). These visas typically allow stays from six months to one year, are often renewable, and are targeted at third-country nationals. While EU citizens can already move freely across member states under the principle of free movement, digital nomad visas are primarily designed to attract skilled professionals from outside the EU, particularly from North America, Latin America, and Asia.

The rationale behind such policies is multifaceted. Governments frame them as strategies to diversify tourism, stimulate local economies, and counter demographic decline (Bozzi 2024; Richards 2021). By attracting relatively affluent mobile workers who spend locally but earn globally, policymakers hope to secure a sustainable inflow of capital without the seasonal fluctuations of mass tourism. Yet critical voices argue that digital nomad visas institutionalise new hierarchies of belonging, privileging high-earning foreigners while excluding less affluent migrants (Webb 2024).

Research indicates that digital nomads in Europe contribute to both economic opportunities and socio-spatial tensions. On the positive side, nomads are credited with stimulating local economies through consumption of housing, coworking spaces, cafes, and leisure services. They are often perceived by policymakers as desirable, long-term tourists who enhance a destination's international visibility (Hannonen 2020; Richards 2021). In Madeira, local businesses have reported significant increases in demand since the launch of the Digital Nomad Village, while in Lisbon and Barcelona, nomads contribute to the development of coworking sectors and creative industries (Riesgo Gómez et al. 2025)

On the negative side, scholars highlight the gentrification and housing pressures associated with nomads' presence. In Southern Europe, nomads compete with locals and tourists for rental housing, contributing to rising rents and displacement (Sequera & Nofre 2019; Cocola-Gant et al. 2020; Taltavull 2024). This has sparked protests, particularly in Barcelona and the Canary Islands, where residents link nomadism to the broader touristification of neighbourhoods (RTVE 2025). The result is an ambivalent reception: while governments and businesses welcome nomads as economic assets, local communities often perceive them as outsiders who exacerbate inequalities and disrupt social cohesion (Mancinelli & Germann Molz 2023).

Digital nomadism in Europe unfolds within the continent's distinctive policy environment of plurilingualism and pluriculturalism, discussed in Section 2.3.2. However, while EU policies encourage intercultural dialogue and plurilingual engagement, research shows that digital nomads often remain linguistically distant from local communities, especially during short stays (Riesgo Gómez et al. 2025). This raises questions about the gap between normative policy ideals and everyday practices, which the present Thesis is seeking to contextualise, challenge and address.

Nevertheless, Europe offers infrastructures that encourage engagement. As will be further described in Sections 4.3 and 4.4, language courses, cultural integration programmes, and migrant support services are widely available. In some contexts, nomads have engaged in local initiatives, workshops, and meet-ups that foster exchange with residents. For instance, in Lisbon and Berlin, coworking spaces have become sites of intercultural networking, where local entrepreneurs and international nomads collaborate on projects (Orel 2019; Richards 2021). Such examples suggest that the European ecology of languages and cultures provides resources for more inclusive forms of nomadism, even if they are not always realised.

All in all, Europe seems to have become a fertile ground for the development of digital nomad communities. European nomad hubs such as Lisbon, Berlin, and Las Palmas illustrate how digital nomadism intersects with broader urban and regional development strategies. Local governments and private actors market these communities as evidence of cosmopolitanism and innovation, using them to reposition cities in global mobility circuits (Bozzi 2024). Yet the inclusivity of such communities remains contested: while open in principle, participation requires economic and cultural capital, excluding many locals and lower-income migrants (Mancinelli 2020; Pentina et al. 2023).

Digital nomadism in Europe thus emerges as a complex phenomenon shaped by global trends, regional policies, and local dynamics. Many of its core characteristics – high mobility, remote-work infrastructures, and flexible transnational lifestyles – are aligned with broader global patterns. Yet certain aspects of the European context diverge from, or complicate, this global trend. Europe is both a privileged site of nomadism, thanks to its dense infrastructures, the relative ease of mobility within the Schengen area, and high levels of cultural and linguistic diversity, and a contested one, where tensions around housing shortages, local resistance, and tourism-driven gentrification mirror debates in migration and urban studies (Germann Molz 2011; Mancinelli 2020). Digital nomadism in Europe therefore cannot be understood solely as

an individual lifestyle choice: it is also shaped by layers of governance, policy frameworks, and community negotiation that condition its emergence and sustainability.

The following chapters will build on this analysis by examining how nomad communities communicate, organise, and negotiate belonging within the European context(s) here outlined.

Chapter 3

A theoretical framework to understand linguistic fluidity

3.1 Theoretical position and structure of the chapter

Drawing on established and innovative research across sociolinguistics, applied linguistics, mobility studies, and transcultural communication, this chapter develops an integrated theoretical framework for the investigation of linguistic fluidity in contexts of mobility. Rather than treating linguistic fluidity as a vague descriptive label, the chapter aims to conceptualise it as a theoretically grounded and analytically operational notion, capable of accounting for contemporary communicative practices shaped by mobility, digitalisation, and superdiversity.

The objective of the chapter is twofold. First, it situates linguistic fluidity at the intersection of complementary theoretical traditions that have progressively challenged static, monolithic, monolingual conceptions of language. Second, it provides a coherent conceptual foundation for the empirical analyses developed in the subsequent chapters, ensuring a transparent alignment between theoretical assumptions, methodological choices, and analytical outcomes.

The chapter adopts an explicitly integrative theoretical position. Translanguaging constitutes the core theoretical lens through which linguistic practices are interpreted, but it is not treated in isolation. Instead, it is embedded within a broader constellation of concepts – including languaging, transcultural communication, plurilingualism, English as a lingua franca, intercomprehension, informal translation, and technology-augmented interaction – which together allow for a complex understanding of how communication unfolds in highly mobile, superdiverse cultural settings.

The chapter is organised into six main sections, each building cumulatively on the previous one.

Section 3.2 introduces a sociolinguistic perspective on mobility and language, outlining how intensified mobility challenges sedentary models of language and calls for analytical frameworks capable of accounting for complexity, variability, and translocality. This section establishes the macro-theoretical backdrop against which linguistic fluidity becomes a necessary analytical concept rather than a descriptive metaphor.

Section 3.3 engages with key theoretical constructs – *languaging*, *transcultural communication, translanguaging*, and *plurilingualism* – clarifying their definitions, theoretical genealogies, and relationships. Particular attention is paid to their points of convergence, as well as to the distinct analytical contributions each concept offers to the study of fluid translingual practices.

Building on this conceptual groundwork, Section 3.4 shifts from theoretical constructs to definition of communicative practices by examining how transcultural communication materialises through a set of closely related communicative modes. These include English as a lingua franca, intercomprehension, translanguaging as a communicative practice[10], informal translation, and technology-augmented communication. Rather than treating these phenomena as discrete or mutually exclusive, the section conceptualises them as interconnected manifestations of linguistic fluidity in action.

Section 3.5 focuses on informal language learning and language socialisation, explicitly linking it to the previous discussion. Transcultural communicative settings are presented as fertile sites for language socialisation, where speakers continuously acquire, reshape, and recalibrate linguistic and semiotic resources. This section thus bridges interactional practices and longer-term processes of competence development.

Finally, Section 3.6 synthesises the theoretical threads developed throughout the chapter to propose a working definition of linguistic fluidity. This definition is explicitly oriented towards empirical analysis and serves as the conceptual basis for the methodological design and analytical procedures discussed in Chapters 4, 5, and 6.

3.2 Mobility and language: A sociolinguistic perspective on language fluidity

In general, sociolinguistics is concerned with the relationship between language and society, traditionally articulating this relationship through a distinction between micro- and macro-level perspectives. At the micro-sociolinguistic level, research has focused on how social structures and categories – such as class, gender, age, ethnicity, and social networks – influence linguistic variation and speakers' choices. Within this tradition, linguistic variables are typically treated as dependent variables whose distribution can be statistically related to independent social variables, revealing systematic patterns rather than random deviation (Labov 1972; Fasold 1990). Speech, in this sense, is socially emblematic: it indexes group membership, contributes to the construction of social identities, and participates in the reproduction of social differentiation. At the macro-sociolinguistic level, the analytical focus shifts from individual speakers to societies and institutions, asking what societies *do* with language. This includes the functional distribution of linguistic varieties, processes of standardisation, language policies and ideologies, and the symbolic and material value attributed to different languages and registers (Fasold 1984; Coulmas 1998). From this perspective, language is embedded in broader political, economic, and historical processes, and linguistic behaviour cannot be fully

[10] It should be noted that the term *translanguaging* as used in the literature can refer to different interrelated concepts: a) 'translanguaging as a practical theory of language', as conceptualised by Wei (2018) and García and Wei (2014); b) t*ranslanguaging as practice*, that is, the observable deployment of integrated linguistic and semiotic resources in interaction, sometimes discussed in relation to (but not reducible to) phenomena such as code-switching and code-mixing (e.g., Canagarajah 2011, 2017; Blackledge & Creese 2017); c) *translanguaging approaches to language teaching*, which derive from translanguaging theory and foreground the legitimisation and pedagogical valorisation of translingual communication (e.g. García et al. 2017; Cenoz & Gorter 2021).

understood without taking into account power relations, institutional norms, and dominant ideologies.

While this distinction remains analytically useful, contemporary conditions of intensified mobility increasingly challenge the explanatory adequacy of static micro–macro dichotomies. In contexts characterised by frequent movement, transnational trajectories, and shifting social alignments, linguistic practices are no longer easily mapped onto stable speech communities or territorially bounded social groups. Linguistic diversity in everyday interaction has become the norm rather than the exception, raising a fundamental sociolinguistic question: if variation is no longer anchored to relatively stable social categories or territorially fixed communities, what does linguistic diversity index, and how should it be theorised? This question has been explicitly addressed in sociolinguistic work on mobility, superdiversity, and complexity (e.g. Vertovec 2007; Blommaert & Rampton 2011).

Early attempts to reconceptualise language in dynamic social environments can be traced back to Haugen's notion of *ecology of language*, which framed language as embedded in a broader environment shaped by social, cultural, and material conditions (Haugen 1972). Building on this ecological insight, more recent sociolinguistic theory has increasingly foregrounded mobility, circulation, and complexity as central analytical dimensions. Blommaert (2010, 2013, 2014), in particular, argues that linguistic resources acquire meaning and value not in isolation, but through the trajectories of people who carry them across space, time, and social domains. Linguistic forms do not *travel* autonomously; rather, they travel *with speakers*, and in doing so they are recontextualised, reinterpreted, and transformed.

From this perspective, linguistic practices cannot be analysed independently of the regimes of mobility that shape who can move, under which conditions, and with what degrees of legitimacy and recognition (Blommaert 2010; Glick Schiller & Salazar 2013). Regimes of mobility operate at the macro level – through nation-states, institutions, labour markets, and language ideologies – but they are experienced and negotiated at the micro level, through individual biographies, trajectories, and interactional encounters. The relationship between macro-structures and micro-practices is therefore not linear, but recursive: institutional norms and ideologies shape communicative expectations, while speakers' situated practices may reproduce, adapt, or subtly challenge these norms.

Mobility studies have further reinforced this shift by conceptualising movement not as an exception to social life, but as one of its constitutive conditions. The *new mobilities paradigm* (Sheller & Urry 2006) highlights how flows of people, information, and capital reshape social relations as well as communicative practices. Incorporating mobility into sociolinguistic analysis thus entails moving beyond sedentary assumptions about language, community, and competence, and adopting frameworks capable of accounting for unpredictability, heterogeneity, and translocal connections.

Crucially, mobility involves the crossing of borders that are not only geographical but also symbolic and ideological: boundaries between languages, cultures, registers, and

communicative norms are continuously negotiated and reconfigured in interaction. This idea of border-crossing has a long intellectual tradition that extends beyond sociolinguistics, including literary and cultural theory. Works such as Anzaldúa's *Borderlands / La Frontera* (2012) conceptualise linguistic hybridity as an experiential condition shaped by movement, marginalisation, and identity negotiation. These perspectives resonate strongly with contemporary sociolinguistic approaches to translanguaging and linguistic fluidity, even when they emerge from different disciplinary traditions.

Within this broader theoretical reorientation, concepts such as *languaging*, *transcultural communication*, *plurilingualism*, and *translanguaging* (discussed in detail in Section 3.3) have emerged as attempts to capture the fluid, processual nature of language use in mobile contexts. Importantly, translanguaging was not initially conceived as a top-down theoretical construct, but emerged from pedagogical and ethnographic observations of transliangual practice (García & Wei 2014). This grounding in practice rather than abstraction partly explains why translanguaging has been described as a *practical theory of language* (Wei 2018): it is anchored in what speakers *do* with their resources in situated interaction, rather than in relation to predefined linguistic systems.

Postcolonial and critical sociolinguistic perspectives further complicate this picture by drawing attention to how dominant language ideologies – often rooted in colonial histories – continue to shape expectations about linguistic purity, correctness, and legitimacy (Anzaldúa 2012; Flores & García 2020). Even in contexts where colonial domination is not directly at stake, English remains a historically colonial language whose global circulation is accompanied by powerful norms of standardisation and *golden standards* of correctness. These norms are closely intertwined with monolingual ideologies that position certain varieties and speakers as legitimate while marginalising others (Phillipson 1992; Pennycook 2007).

Because this thesis includes a major focus on English as a lingua franca within translingual settings, it cannot disregard the role of language ideology. ELF communication simultaneously relies on English as a shared resource and destabilises monolithic notions of English by foregrounding variability, adaptability, and communicative effectiveness over native-speaker norms. In this sense, the colonial and elitist ideologies historically associated with English intersect with contemporary monolingual norms, shaping both expectations and evaluations of linguistic behaviour in subtle but consequential ways.

Crucially, incorporating movement into linguistic analysis also requires rethinking methodological approaches. As sociolinguistics has progressively embraced complexity, it has become increasingly clear that no single method can adequately capture the layered nature of mobile communication. Blommaert (2014) explicitly argues that the task of sociolinguistic analysis is not to simplify complexity, but to make it visible: 'The task of analysis is not to reduce complexity […] but to demonstrate complexity, to unfold the complex and multifiliar features and their various different origins that are contained in synchronized moments of understanding' (Blommaert 2014: 13).

As will be discussed in Chapter 4, the FLO project responds to this theoretical orientation through methodological pluralism, combining qualitative interactional analysis, corpus-informed approaches, and statistical modelling. This methodological stance reflects the understanding that linguistic fluidity is not a property of isolated linguistic forms, but an emergent feature of interaction shaped by social relations, trajectories of mobility, and contextual constraints.

In this sense, communication in mobile settings can be understood as a continuous process of negotiation among multi-affiliated individuals who draw on diverse semiotic repertoires to achieve mutual understanding (Zhu 2017). Language use becomes less a matter of selecting from discrete codes and more a matter of strategically mobilising available resources in response to interactional demands, ideological orientations, and processes of self-representation. This reconceptualisation provides the sociolinguistic foundation for the sections that follow.

The next section builds on this perspective by examining how notions such as *languaging*, *translanguaging*, *plurilingualism*, and *transcultural communication* have been theorised within applied linguistics and sociolinguistics, and how their interrelations contribute to a dynamic understanding of linguistic fluidity.

3.3 Languaging, transcultural communication, translanguaging and plurilingualism

As discussed in the previous section, mobility-oriented sociolinguistics requires theoretical tools capable of accounting for language as a dynamic, adaptive, and socially embedded practice. From this perspective, concepts such as *languaging*, *transcultural communication*, *translanguaging*, and *plurilingualism* have emerged across different research traditions to describe overlapping, yet analytically distinct, dimensions of linguistic fluidity. While these notions are sometimes used interchangeably in the literature, this section argues that they are best understood as theoretically related but non-identical constructs, each foregrounding a specific aspect of fluid language use under conditions of mobility.

At a general level, the relationship among these concepts can be described as follows: *languaging* provides the broad cognitive–philosophical foundation; *transcultural communication* captures the interactional and experiential condition in which meaning-making occurs; *translanguaging* describes the intensified enactment of languaging within transcultural conditions; and *plurilingualism* offers a policy-oriented and competence-based perspective on the deployment of heterogeneous resources. Together, they contribute to a multi-layered understanding of linguistic fluidity that moves beyond static models of language.

The concept of *languaging* represents a foundational shift away from viewing language as a bounded system towards understanding it as an activity. Drawing on ecological, dialogic, and distributed cognition perspectives, *languaging* conceptualises meaning-making as an ongoing

process that emerges in interaction rather than as the use of pre-existing linguistic objects with fixed meanings (Love 2017). From this standpoint, languages are not autonomous entities stored in the mind, but historically sedimented abstractions that emerge from repeated social practices, institutionalisation, and ideological stabilisation.

Languaging foregrounds the idea that communication is inherently adaptive, context-sensitive, and relational. Meaning is not transmitted through stable codes but co-constructed through interaction, drawing on linguistic, embodied, material, and contextual cues. This perspective provides the cognitive and philosophical grounding for more specific theories that address multilingual and mobile communication, including translanguaging and transcultural communication.

Within applied linguistics, *transcultural communication* has been theorised most explicitly by Baker (2011, 2015, 2022) as a way of conceptualising communication in contexts characterised by cultural complexity, mobility, and unpredictability. Rather than assuming interaction between clearly bounded cultures, transcultural communication emphasises the fluid, emergent, and negotiated nature of cultural meanings as they are enacted in interaction.

Importantly, transcultural communication is not simply about cultural difference; it is about communication *across*, *through*, and *beyond* cultural boundaries. These boundaries may be national, linguistic, ideological, or symbolic, and they are continuously reconfigured in interaction. From this perspective, culture is not a stable background variable but an interactional resource that speakers orient to, invoke, or transform depending on contextual demands.

In this sense, transcultural communication can be understood as an existential condition of contemporary interaction, particularly in mobile and digitally mediated settings. It creates the social and interactional environment in which languaging takes place under heightened conditions of diversity, asymmetry, and norm instability.

Translanguaging builds directly on this understanding of language as activity, but foregrounds the transcultural dimension of languaging. As articulated by García and Wei (2014) and further developed by Li Wei (2018), translanguaging is not merely a description of multilingual practices, but a theoretical stance on language.

Li Wei (2018) conceptualises *translanguaging* as a 'practical theory of language' grounded in speakers' lived experience. In linguistic research, the translanguaging perspective shifts the analytical focus from discrete languages to speakers' repertoires and practices, challenging monolingual norms and reconfiguring traditional notions of communication as well as linguistic competence.

Translanguaging is inseparable from speakers' biographies, trajectories of mobility, situated in their bodies and cultural histories. As Blackledge and Creese note, 'the translanguaging repertoire incorporates biographies and learning trajectories; […] it is responsive to the places

in which, and the people with whom, semiotic resources may be deployed' (2017: 34). Repertoires, from this perspective, are not static inventories of languages, but relational, adaptive configurations that expand, contract, and reorganise as speakers move across social and geographical spaces. This view aligns closely with mobility- and complexity-oriented sociolinguistics, according to which linguistic resources circulate unevenly and acquire different values depending on context (Blommaert & Rampton 2011).

Using the term *translingualism*[11], Canagarajah further reinforces this reconceptualisation by arguing that translingual practices challenge traditional understandings of multilingualism based on the coexistence of separate linguistic systems:

> Challenging traditional understandings of language relationships in multilingualism, which postulates languages maintaining their separate structures and identities even in contact, translingualism looks at verbal resources as interacting synergistically to generate new grammars and meanings, beyond their separate structures. (2017: 1)

This perspective is particularly relevant for understanding translanguaging not as deviation or mixture, but as the reality of communication.

Within European language policy and research, similar ideas are articulated through the concept of *plurilingualism*, most notably in the Common European Framework of Reference for Languages and its Companion Volumes (Council of Europe 2001; 2018; 2020). Plurilingualism explicitly rejects the ideal of balanced monolingual competence in multiple languages, instead promoting the ability to mobilise partial, uneven, and heterogeneous resources to negotiate meaning across linguistic and cultural boundaries.

While plurilingualism shares important assumptions with translanguaging – particularly the notion of integrated repertoires – it is more competence- and outcome-oriented. It focuses on what individuals can do communicatively with their resources, often within educational and institutional contexts, rather than on the internal and/or interactional languaging process which allow meaning to emerge. In this sense, plurilingualism can be seen as conceptually less radical, but highly effective in legitimising non-monolingual practices within policy and pedagogical frameworks.

---

[11] It should be noted that *translanguaging* is only one of several terms used in the literature to conceptualise fluid translingual practices, alongside notions such as *translingualism*, *metrolingualism*, and *polylanguaging* (e.g. Canagarajah 2013; Pennycook & Otsuji 2015; Vogel & García 2017). These labels reflect different theoretical emphases rather than mutually exclusive views, foregrounding aspects such as urban space, mobility, creativity, or ideological resistance. This thesis adopts *translanguaging* as an overarching term because it offers a theoretically integrated framework that links language practices, speaker agency and learning, while explicitly challenging monolingual and standard-language ideologies (García & Wei 2014; Wei 2018). Unlike other labels, *translanguaging* has developed simultaneously as a theory of language, an analytical stance on interaction, and a pedagogical orientation, making it particularly suited to capturing the disruptive and transformative potential of linguistic fluidity in mobile and transcultural contexts.

The relevance of translation practices further illustrates the interconnectedness of these concepts. Recent work in translation studies has increasingly challenged Western historiographic models that presuppose clear boundaries between source and target languages, texts, and cultures. Pym (2021) argues that such models are grounded in a foundational binarism that fails to capture the diversity of translational practices observed across contexts. He observes that dominant Western models in translation studies are grounded in binary assumptions that conceptualise translation as a transfer between clearly separated languages, cultures, and texts. Recent scholarship has therefore highlighted forms of translation that challenge these rigid distinctions, drawing attention to more fluid, situated, and relational practices that emerge in intercultural communication. Informal situated translation – often oral, partial, and embedded in interaction – aligns closely with translanguaging practices. Rather than treating translation as a specialised, externally regulated activity, this line of research conceptualises it as a relational practice oriented towards trust, collaboration, and inclusion within transcultural encounters (see also Section 3.4.4).

In conclusion, *languaging, transcultural communication*, *translanguaging*, and *plurilingualism* are best understood not as competing theories, but as interconnected analytical lenses operating at different levels. Languaging provides the philosophical foundation; transcultural communication describes the interactional condition; translanguaging captures the complex enactment of meaning-making under such conditions; and plurilingualism offers a competence- and policy-oriented articulation of similar principles. Thereby, these concepts provide the theoretical foundation for analysing communicative competence not as adherence to monolingual norms, but as an adaptive, situated, and socially meaningful practice. The next section moves from theory to practice by examining how these forms of transcultural communication operate in specific communicative modes and how these modes have been approached by linguistic research, beginning with one of the most proliferous field of contemporary applied linguistics studies, namely English as a lingua franca, with a particular focus to mobile and superdiverse interactional settings.

3.4 Transcultural communication in action

While the previous section has outlined the theoretical constructs that underpin a dynamic understanding of linguistic fluidity, this section shifts the focus from theory to practice by examining a diverse body of work dedicated to the investigation of transcultural communication as it is enacted in concrete communicative situations. Hence, rather than treating transcultural communication as an abstract or purely attitudinal competence, Section 3.4 conceptualises it as an interactional achievement, realised through a range of communicative practices that speakers utilise in response to contextual demands, asymmetries of knowledge, and according to the resources they have available, in order to co-construct and/or negotiate meaning for mutual understanding.

In line with the theoretical perspective developed in Sections 3.2 and 3.3, transcultural communication is here understood as unfolding through integrated semiotic repertoires, where linguistic, paralinguistic, translational, and technological resources are dynamically combined. The practices discussed in this Section should therefore not be interpreted as discrete or mutually exclusive categories. On the contrary, they frequently co-occur and overlap in naturally occurring interaction, as will be extensively discussed in Chapter 6. Their analytical distinction serves a heuristic purpose, allowing for a more fine-grained examination of how linguistic fluidity is enacted across different communicative modes.

The section is organised around five interrelated practices that are particularly salient in transcultural communicative environments. Section 3.4.1 examines English as a lingua franca (ELF) as a key shared resource in transnational interaction, focusing on its adaptive, non-normative, and functionally oriented nature. Section 3.4.2 turns to intercomprehension, highlighting how speakers exploit partial mutual intelligibility across related languages to facilitate understanding without full linguistic convergence. Section 3.4.3 discusses translanguaging as a communicative practice, emphasising how speakers strategically mobilise their entire repertoires to negotiate meaning beyond named language boundaries. Sub-section 3.4.4 addresses informal translation, conceptualised as an interactionally embedded practice that challenges traditional distinctions between translation and non-translation. Finally, Section 3.4.5 explores technology-augmented communication, examining how digital tools and platforms reshape communicative practices and expand speakers' semiotic repertoires in transcultural contexts.

The sub-sections thus illustrate how transcultural communication is realised through a constellation of practices that enable speakers to navigate diversity, manage communicative breakdowns, and construct shared meaning. Importantly, these practices do not merely serve immediate interactional goals, but also create conditions for informal language learning and language socialisation, a link that will be explored in detail in Section 3.5.

3.4.1 English as a lingua franca

English as a lingua franca (ELF) has emerged as a central concept for understanding communication in transcultural contexts. A general and agreed-upon definition of ELF is not that of a specific variety of English, but rather its use as a shared communicative resource among speakers for whom it is not necessarily a first language, and often not even a language of primary identification, but instead the only or most convenient option (Seidlhofer 2011; Jenkins 2015). This functional and usage-based understanding of ELF was already foregrounded in early collective work such as *English as a Lingua Franca* (Mauranen & Metsä-Ketelä 2006), which explicitly framed ELF as an emergent mode of communication shaped by interactional needs rather than by codified norms. This stance is undertaken also by most recent threads in ELF research, which do not conceptualise English as a bounded linguistic system to be approximated or mastered according to native-speaker norms, but as a flexible and adaptive means of communication oriented towards mutual understanding and interactional effectiveness.

Early ELF scholarship developed in response to the growing recognition that English was increasingly used in contexts where no native speakers were present, and where communicative success could not be adequately explained by reference to standard varieties alone (Firth 1996; House 1999). This observation challenged long-standing assumptions in applied linguistics and language pedagogy, particularly the idea that native-speaker competence should function as the primary benchmark for successful communication. Instead, ELF research foregrounded accommodation, negotiation, and pragmatic strategies as central features of interaction, shifting attention from formal correctness to communicative intelligibility and cooperation. Over time, however, ELF research has moved beyond early descriptions that implicitly framed English as a lingua franca in terms of simplification or reduction. Subsequent studies have shown that ELF interactions are often characterised by expansion, creativity, and the strategic enrichment of linguistic and semiotic resources, rather than by deficit or loss (Seidlhofer 2011; Jenkins 2015; Cogo & Dewey 2012). Already in the mid-2000s, corpus-based and discourse-analytic work discussed in Mauranen & Ranta (2006) demonstrated that ELF usage frequently involves complex syntactic patterning, pragmatic sophistication, and discourse-level coherence, thereby challenging deficit-oriented interpretations of non-standard English use. English in ELF settings is thus increasingly understood as a flexible and adaptive resource, shaped by speakers' multilingual repertoires, communicative goals, and interactional contexts. This shift aligns with more recent perspectives that conceptualise ELF as a dynamic, emergent practice embedded in broader processes of mobility and linguistic fluidity – an understanding that this thesis adopts and further refines, as will be discussed in Section 6.3.

Corpus linguistics has played a pivotal role in the development of ELF research, providing the empirical foundation necessary to move beyond abstract theorisation and deficit-oriented interpretations of 'non-native' English use. From its early stages, ELF research has been closely associated with the construction of corpora specifically designed to capture linguistic production and/or interaction in contexts where English functions as a shared communicative resource rather than a target language.

One of the earliest and most influential contributions in this respect is the ELFA (English as a lingua franca in academic settings) corpus (Mauranen 2003, 2012), which documents spoken academic English produced by non-native speakers in university contexts. ELFA was explicitly designed to challenge the assumption that academic English must be modelled on native-speaker norms, showing instead that interactional success in academic settings relies on pragmatic negotiation, discourse organisation, and audience design rather than on formal correctness alone. The corpus demonstrated that ELF interaction is not characterised by simplification or communicative deficiency, but by strategic adaptation and functional adequacy.

Alongside ELFA, the VOICE corpus (Vienna–Oxford International Corpus of English) (Seidlhofer et al. 2013) represents a major milestone in ELF corpus linguistics. VOICE was designed to capture a wide range of spoken ELF interactions across domains such as education, professional communication, and informal conversation. A defining feature of VOICE is its principled avoidance of native-speaker benchmarks in transcription and annotation, reflecting

the theoretical stance that ELF should be analysed on its own terms. As Mauranen (2006) notes, the design of ELF corpora involves specific challenges, including speaker heterogeneity, fluid norms, and the absence of a stable reference variety, but these very challenges constitute the empirical strength of ELF research.

Mauranen (2006) explicitly addresses the methodological implications of studying ELF through corpora, by arguing that ELF corpora necessarily depart from traditional corpus design assumptions, as they cannot rely on predefined norms or homogeneous speaker populations. Instead, ELF corpora foreground variability, interactional negotiation, and emergent regularities, requiring flexible analytical frameworks and interpretative sensitivity. Importantly, Mauranen emphasises that ELF corpora are not intended to stabilise a new variety of English, but to document patterns of use that emerge from recurrent communicative needs in multilingual interaction.

Over time, ELF corpus research has contributed to a broader reconceptualisation of English as a dynamic, adaptive, and interactionally shaped resource. Corpus-informed studies have shown that ELF interactions are often characterised by expansion, creativity, and the strategic mobilisation of linguistic and pragmatic resources, rather than by reduction or erosion (Cogo & Dewey 2012; Jenkins 2015). This empirical evidence has supported a shift away from static, system-oriented views of language towards usage-based and practice-oriented models, in which meaning emerges through interaction rather than through adherence to codified norms.

However, ELF corpus linguistics has also highlighted the limitations of corpus-based approaches when dealing with highly fluid, translingual practices. As ELF research increasingly converges with translanguaging and plurilingual perspectives, scholars have pointed out that corpus tools and query systems often remain anchored to monolingual representations of language, making it difficult to fully capture cross-linguistic and multimodal practices (Mauranen 2012; Seidlhofer 2011). This tension has opened new methodological debates about how corpora can be adapted to represent linguistic fluidity without reifying named languages. Within this evolving landscape, the present thesis builds upon the methodological legacy of ELF corpus linguistics while extending its scope to translingual interaction (see Chapter 5 and 7).

From a translanguaging perspective, ELF foregrounds the ways in which speakers mobilise English in conjunction with other linguistic, semiotic, and interactional resources, dynamically adapting their communicative practices to specific interlocutors, domains, and interactional goals (Cogo & Dewey 2012; García & Wei 2014). This practice-oriented view resonates with earlier ELF research that already resisted the reification of ELF as a stable variety, instead emphasising variability, contingency, and interactional emergence (Mauranen & Ranta 2006). Moreover, such a view aligns with broader sociolinguistic reconceptualisations of multilingual communication as repertoire-based and processual rather than code-bound. In contrast to earlier strands of ELF research that implicitly sought to describe or stabilise recurring patterns of ELF usage as an emergent variety, more recent approaches emphasise variability, fluidity, and situated adaptation as constitutive features of ELF communication. ELF usage is thus

inherently variable and context-sensitive, shaped by participants' linguistic biographies, degrees of mobility, and ideological orientations towards English itself. This variability is not a deficiency but a defining characteristic of ELF communication (as will be argued in Section 6.3).

At the same time, the global spread of English cannot be disentangled from its historical and ideological dimensions. English is a traditionally colonial language that has undergone a profound process of normalisation and institutionalisation. Purism and normalilsation towards British and American standards have historically been mechanisms through which colonial power was perpetrated (Bordieu 1991). This has resulted in the persistence of powerful *golden standards* of English that continue to shape language ideologies, even in contexts where English primarily functions as a lingua franca (Jenkins 2007; Holliday 2006; Mauranen & Ranta 2006). In this sense, ELF communication unfolds within a tension between use and ideology: while speakers routinely employ English in flexible, adaptive, and non-standard ways in interaction, these practices are frequently evaluated, policed, or implicitly measured against monolingual, standard-language, and native-speaker-based norms (Canagarajah 2013, 2017). Even in privileged social contexts, like the one on which the present thesis focus, English use remains ideologically loaded: choices about accent, lexical items, or pragmatic strategies can index competence, authority, or legitimacy, revealing how ELF practices are embedded in broader regimes of linguistic inequality.

Importantly, ELF research has increasingly converged with translanguaging and plurilingual perspectives, which reject the idea of English as a self-contained system altogether. From this viewpoint, ELF interactions frequently involve the mobilisation of plurilingual and multimodal resources, including elements from other languages, metalinguistic commentary (Mauranen 2023), informal translation, and digital affordances. English thus becomes one component of a semiotic repertoire, rather than the sole medium of communication. This reconceptualisation aligns with translingual approaches that view meaning-making as distributed across linguistic and semiotic resources interacting dynamically in context (Canagarajah 2017).

In this thesis, English as a lingua franca is therefore not treated as an isolated phenomenon, but as one communicative mode within a broader ecology of transcultural communication. Its analytical relevance lies not in identifying stable linguistic features, but in examining how English is strategically mobilised alongside other resources to manage understanding, negotiate participation, and construct social relations in mobile interactional settings. By building on both foundational ELF scholarship and more recent translanguaging-oriented approaches, the thesis situates ELF within a continuum of research that progressively moves from code-based to practice-based understandings of multilingual communication. This perspective provides the conceptual bridge to the following sub-sections, which examine intercomprehension, translanguaging, informal translation, and technology-augmented communication as complementary practices through which linguistic fluidity is enacted in everyday interaction.

3.4.2 Intercomprehension

Intercomprehension has been theorised over the past three decades as a communicative process that enables mutual understanding between speakers of different languages without requiring full productive competence in any of the involved languages. Rather than presupposing convergence towards a single lingua franca, intercomprehension allows each participant to use their preferred language while actively engaging in the co-construction of meaning across linguistic boundaries. In this sense, it constitutes a paradigmatic example of inclusive *ecological* transcultural communication in action, grounded in plurilingual reception and interaction.

A widely cited definition conceptualises intercomprehension as 'a communicative process that allows the co-construction of meaning by speakers of different languages, in situations of plurilingual reception and/or interaction' (Capucho 2018: 161). This definition foregrounds three aspects that are central to the present thesis: the processual nature of communication, the collaborative dimension of meaning-making, and the mobilisation of plurilingual repertoires rather than discrete linguistic systems.

Research on intercomprehension has developed along multiple, partly overlapping dimensions. First, intercomprehension has been described as an innate or intuitive competence, grounded in speakers' ability to exploit typological proximity, prior linguistic knowledge, and inferential strategies when encountering related languages (Capucho 2004). From this perspective, intercomprehension is not learned ex nihilo, but emerges from the interaction between cognitive dispositions and linguistic experience. Second, intercomprehension has been framed as a competence that can be fostered and developed through educational practices and language policies, particularly within the European context (Escudé 2007). Third, more recent work emphasises intercomprehension as an interactional event, unfolding dynamically in situated communicative encounters, both face-to-face and digitally mediated (Mantegna 2022).

What unites these views is a shared rejection of monolingual norms and of the idea that effective communication presupposes symmetrical competence in a single language. Instead, intercomprehension relies on partial, uneven, and asymmetrical knowledge, legitimising approximation, inference, and strategic adaptation. Intercomprehension, similarly to English as a lingua franca, shifts the focus from linguistic correctness to communicative effectiveness, explicitly challenging ideologies of linguistic perfection that have long dominated language teaching and assessment

This reconceptualisation aligns closely with the plurilingual orientation promoted by European language policy. The Common European Framework of Reference for Languages (CEFR) conceptualises language users as *social actors* endowed with a plurilingual and pluricultural competence, in which languages are not compartmentalised but interact within a single, evolving repertoire. Within this framework, intercomprehension is explicitly recognised as part of plural approaches to languages and cultures, alongside translanguaging and mediation (Council of Europe 2001; 2018). The FREPA / CARAP framework further elaborates this view by identifying competences that cut across languages and cultures, including mediation,

adaptation, decentring, and negotiation—competences that are central to intercomprehension practices (Candelier et al. 2012).

From a sociolinguistic perspective, intercomprehension can be understood as a resource-oriented practice rather than a language-specific skill. Speakers actively mobilise what they already know – lexical transparency, morphosyntactic correspondences, discourse conventions, contextual cues, and world knowledge – to construct meaning collaboratively. As repeatedly emphasised in the literature, understanding is not achieved through word-by-word decoding, but through global, strategic, and inferential processes that tolerate opacity and incompleteness, and, importantly, that can be promoted by developing metatranslinguistic conciousness (Baude et al. 2006; Balboni 2007).

As a matter of fact, intercomprehension does not imply communicative passivity. Research on interproduction highlights how speakers can actively adapt their production in their own language to facilitate comprehension by interlocutors, for instance by slowing down speech, avoiding highly culture-specific idioms, or explicitly paraphrasing key terms (Balboni 2010; Capucho 2018). This adaptive dimension reinforces the idea of intercomprehension as a relational practice, oriented towards the interlocutor and grounded in mutual responsibility for understanding.

In contexts of mobility and superdiversity, like the ones here under investigation, intercomprehension becomes particularly salient. Speakers in such contexts often interact in environments characterised by linguistic heterogeneity, limited shared norms, and temporary alignments, where the adoption of a single dominant code is neither always possible nor desirable. In such settings, intercomprehension functions as a low-threshold communicative strategy, enabling participation without requiring linguistic assimilation. It thus complements, rather than competes with, English as a lingua franca and translanguaging practices.

From the perspective adopted in this thesis, intercomprehension can be situated within a translanguaging semiotic repertoire, in which linguistic resources interact with pragmatic strategies, multimodal cues, and ideological stances towards language. Intercomprehension perspectives share with translanguaging a holistic view of competence and a rejection of rigid language boundaries, while maintaining a stronger focus on receptive and inferential processes. This makes intercomprehension a further crucial analytical category for understanding how linguistic fluidity is enacted in everyday transcultural communication.

The next sub-section turns to translanguaging as a communicative practice, further expanding the analysis of how speakers actively combine and transform linguistic and semiotic resources in culturally-diverse interactional settings.

3.4.3 Translanguaging as a communicative practice

Within the broader framework of transcultural communication outlined in the previous sections, translanguaging can be understood as a situated communicative practice through which speakers actively perform identity and negotiate meaning in communicative conditions

characterised by cultural diversity and asymmetrical linguistic competence. In this sense, translanguaging is best approached as an interactional accomplishment, emerging from concrete communicative and/or performative needs rather than from adherence to predefined linguistic systems.

Building on Li Wei's (2018) formulation of translanguaging as a *practical theory of language*, translanguaging practices are here conceived as strategic deployments of heterogeneous semiotic resources – linguistic, paralinguistic, embodied, and digital – mobilised by speakers to ensure intelligibility, manage interactional contingencies, and achieve communicative goals, while defining and performing their own identity through language. Crucially, thus, agency lies with the speakers themselves, who select, combine, and reshape resources in response to contextual demands, interlocutor repertoires, and perceived communicative risks.

This practice-oriented understanding aligns with Canagarajah's (2017) translingual perspective, which explicitly challenges traditional multilingual models based on the assumption that languages remain structurally separate even in contact. Instead, translingual practice foregrounds the synergistic interaction of verbal resources, through which new meanings, local norms, and emergent grammars are co-constructed in interaction. From this viewpoint, translanguaging is not reducible to alternation between codes (as in classic code-switching accounts), but involves a reconfiguration of available resources that exceeds the boundaries of named languages.

Empirical research in mobility and superdiverse settings has repeatedly shown that translanguaging practices are particularly salient in informal, goal-oriented interactions, such as those occurring in workplaces, online communities, and transient social networks (Blommaert & Rampton 2011; Zhu 2017). In such contexts, speakers often do not share a fully overlapping linguistic repertoire, nor do they orient to stable normative models of correctness. Translanguaging in this sense functions as a resource for managing partial understanding, accommodating diverse linguistic biographies, and mitigating communicative breakdowns without interrupting interactional flow.

Translanguaging as a communicative practice is deeply intertwined with biographical trajectories and mobility histories. As Blackledge and Creese (2017) emphasise, translanguaging repertoires are shaped by speakers' past learning experiences, migratory paths, and social affiliations, and are responsive to the places and interlocutors encountered along these trajectories. In superdiverse communities such as those formed by digital nomads, translanguaging practices are therefore not exceptional or marked, but constitute a default mode of communication, reflecting the ongoing negotiation of belonging, expertise, and legitimacy in fluid social environments.

From a critical perspective, translanguaging practices also carry ideological significance. By foregrounding communicative effectiveness over conformity to standardised norms, they implicitly challenge monolingual and native-speaker ideologies that continue to dominate many institutional and educational settings. This is particularly relevant in contexts where

English functions as a lingua franca: while English often provides a shared communicative baseline, translanguaging practices reveal how speakers adapt, reshape, and supplement it with other linguistic and semiotic resources, resisting a monolithic or prescriptive conception of English use.

Finally, translanguaging as a communicative practice cannot be fully understood in isolation from other closely related practices discussed in this section, such as informal translation and technology-augmented communication. In many naturally occurring interactions, translanguaging co-occurs with spontaneous translation, paraphrasing, reformulation, and the use of digital tools, forming hybrid communicative ecologies that blur analytical boundaries between practices. Recognising this overlap does not weaken the analytical value of translanguaging; rather, it situates it within a broader repertoire of adaptive strategies through which speakers navigate transcultural communication.

Summing up, approaching translanguaging as a communicative practice allows us to move beyond static descriptions of multilingualism and to account for how linguistic fluidity is enacted moment by moment in interaction. This perspective provides a crucial bridge between the theoretical constructs introduced in Section 3.3 and the empirical analyses developed in the subsequent chapters, grounding the study of linguistic fluidity in observable, socially meaningful action.

In this respect, it should be noted that what is defined here as *translanguaging in practice* has traditionally been described in the literature using terms such as *code-switching* and *code-mixing* (e.g. Gumperz 1982; Auer 1998; Muysken 2000). While these notions have been analytically productive, they are grounded in a structuralist and monolingual perspective that presupposes the existence of separate, bounded linguistic systems between which speakers alternate. From a translanguaging perspective, such assumptions are conceptually problematic, as speakers are understood to mobilise integrated and dynamic repertoires rather than switching between discrete *codes* (García & Wei 2014; Canagarajah 2013). Nevertheless, these terms may still be used for operational and descriptive purposes (as done in Chapters 5 and 6), provided that they are detached from monolingual assumptions and reinterpreted as observable linguistic phenomena within a holistic view of speakers' repertoires.

3.4.4 Informal translation

Within the landscape of transcultural communication practices, informal translation occupies a crucial yet often under-theorised position. Unlike institutional or professional translation, informal translation refers to spontaneous, non-regulated translational activities that emerge in everyday interaction to facilitate understanding across linguistic and cultural boundaries. As argued by Antonini et al. (2017), informal translation is not a marginal or deficient form of translation, but a constitutive practice of plurilingual communication, particularly salient in contexts characterised by mobility, diversity, and asymmetrical linguistic competence.

Antonini et al. conceptualise informal translation as a situated, interactionally embedded activity, typically performed by non-professional speakers who do not necessarily self-identify as translators (2017). These practices include ad hoc paraphrasing, summarising, reformulating, explicating, and mediating meaning across languages or varieties, often within the same turn or sequence of interaction. Crucially, informal translation is not oriented towards fidelity to a source text in the traditional sense, but towards communicative adequacy, mutual understanding, and interactional progress.

This functional orientation differentiates informal translation from classical Western models of translation, which have historically been grounded in assumptions of clear boundaries between source and target languages, texts, and communicative roles. As Antonini et al. note, informal translation practices challenge these binaristic assumptions by blurring distinctions between translator and speaker, original and translation, production and mediation (2017). Translation, in this perspective, becomes a relational practice, embedded in social interaction and shaped by contextual constraints rather than by externally imposed norms.

Importantly, informal translation is deeply intertwined with power relations, responsibility, and trust. In many transcultural encounters, speakers who engage in informal translation take on a mediating role that goes beyond linguistic transfer, involving decisions about what to translate, how much to explicate, and which meanings to prioritise. These decisions are often guided by ethical considerations, interpersonal alignments, and assumptions about interlocutors' needs and competencies (Antonini et al. 2017, Pym 2021). Informal translation thus functions not only as a linguistic practice, but also as a social and moral activity, oriented towards inclusion and cooperation.

From a sociolinguistic perspective, informal translation aligns closely with translanguaging practices, although their distinction remains purposeful for analytical depth. Both involve the mobilisation of heterogeneous semiotic resources and challenge monolingual ideologies; however, informal translation is specifically characterised by an orientation to mediation, that is, by the explicit or implicit goal of making meaning accessible across perceived linguistic boundaries. While translanguaging foregrounds the fluid integration of resources within a speaker's repertoire, informal translation highlights moments in which speakers momentarily reconstruct boundaries – only to traverse them strategically in interaction, and to co-create an inclusive communicative space.

Recent developments in translation studies further support this reconceptualisation. Scholars have increasingly called for broader models of translation that account for oral, partial, multimodal, and interactionally distributed forms of mediation, particularly in superdiverse and digitally mediated contexts. In this sense, informal translation exemplifies what Antonini et al. describe as a continuum of translational practices, ranging from highly institutionalised professional translation to everyday acts of meaning mediation embedded in social interaction (2017).

In contexts of mobility and transcultural communication, informal translation should therefore be seen not as an auxiliary or exceptional practice, but as a central mechanism through which communicative continuity is maintained despite linguistic asymmetries. Its prevalence in everyday encounters underscores the need to rethink translation not as a bounded professional activity, but as an integral component of linguistic fluidity under conditions of mobility.

By situating informal translation within a broader ecology of communicative practices, the discussion presented in this sub-section contributes to a more inclusive and process-oriented understanding of how meaning is negotiated in transcultural interaction.

3.4.5 Technology-augmented communication

Technological mediation has become an increasingly salient dimension of transcultural communication. Rather than replacing face-to-face interaction, digital technologies increasingly function as embedded communicative resources, augmenting speakers' semiotic capacities during interaction and reshaping how meaning is negotiated across linguistic boundaries.

At a general level, it is useful to distinguish between two interrelated but distinct perspectives on technology-augmented communication. The first concerns the use of digital tools within co-present interaction, where technology operates as an auxiliary semiotic resource supporting meaning-making in real time. The second, often referred to as technology-mediated communication (Gattiker & Stollemmaier 2012), concerns communication that takes place primarily within digital environments, such as social media platforms, messaging apps, and online communities. While both dimensions are relevant to the present research, this section focuses primarily on the former, including technology as a resource within speakers' translangauging repertoires, whereas the latter dimension is approached by adopting a netnographic perspective on digital nomad communities in Sub-sections 2.2.2 and 4.4.2.

In face-to-face transcultural encounters, smartphones and portable digital devices are increasingly mobilised as aiding tools. Real-time translation applications, image searches, maps, and voice-to-text systems allow speakers to overcome linguistic constraints by supplementing verbal interaction with visual, textual, and auditory resources. As scholars of multimodality have argued, communication is rarely purely linguistic; rather, it unfolds through the orchestration of multiple semiotic modes, including gesture, gaze, spatial orientation, and material artefacts (Kress 2009; Norris 2011). Digital devices extend this multimodal repertoire by enabling the rapid introduction of external semiotic resources into the interactional space.

From this perspective, technology does not merely mediate communication but augments the communicative ecology in which interaction takes place. Hutchins' (1995) theory of distributed cognition is particularly relevant here, as it conceptualises cognitive and communicative processes as distributed across individuals, artefacts, and environments. Real-time translation apps or shared visual displays function as cognitive artefacts that participate in meaning-

making, rather than as neutral tools operating outside the interaction. Meaning emerges through the coordinated use of human and technological resources, blurring the boundary between linguistic competence and technological affordance.

The use of real-time machine translation in interaction further complicates traditional distinctions between speaker, translator, and tool. Studies on translation technologies have shown that even when automated systems are involved, human users remain actively engaged in interpreting, correcting, contextualising, and negotiating output (O'Hagan 2016; Cronin 2013). Machine translation, in this sense, becomes part of an interactional assemblage, rather than an autonomous communicative agent. Its effectiveness depends not only on algorithmic accuracy, but on speakers' ability to integrate technological output into the ongoing flow of interaction.

The use of technology in co-present communication thus aligns closely with translanguaging and informal translation practices. Rather than reinforcing monolingual norms, technological tools often facilitate hybrid, flexible, and adaptive communicative strategies. Speakers may alternate between speech, typing, showing images, pointing at screens, or reformulating translated output, engaging in a form of multimodal translanguaging that transcends purely verbal exchange (Canagarajah 2013; Li Wei 2018). Technology thus expands the semiotic repertoire available to speakers, enabling them to navigate communicative challenges creatively and pragmatically.

At the same time, the increasing role of technology in communication raises questions about access, power, and inequality. Digital tools are unevenly distributed, and their affordances are shaped by commercial interests, algorithmic biases, and dominant language ideologies (van Dijck 2013; O'Hagan 2019). While technology can enhance communicative inclusion, it can also reproduce hierarchies by privileging certain languages, scripts, or communicative norms. These tensions underscore the need to situate technology-augmented communication within broader sociolinguistic and ideological frameworks.

The second perspective on technology-mediated communication concerns interaction that unfolds primarily within digital environments, such as online forums, social media platforms, and messaging applications. In these contexts, technology does not simply support communication but constitutes its primary infrastructure, shaping interactional norms, temporalities, and participant roles. Digital spaces enable new forms of community formation, language socialisation, and identity construction that are not constrained by physical co-presence (Thurlow & Mroczek 2011; Androutsopoulos 2015). While this dimension is closely connected to the present discussion, it will be examined more extensively in Chapter 4, where online communication is analysed as both a site of linguistic practice and a methodological resource.

In conclusion, technology has become an integral component of transcultural communication. Whether embedded in face-to-face interaction or constitutive of online communicative environments, digital tools participate actively in the negotiation of meaning. Technology-

augmented communication thus represents a key dimension of linguistic fluidity, as it enables speakers to mobilise expanded semiotic repertoires in response to the demands of mobility, diversity, and interactional complexity.

3.5 Informal language learning and language socialisation

Research on language learning has long moved beyond formal instructional settings to acknowledge the central role of social interaction in the development of linguistic competence. Within this broader shift, language socialisation theory has provided a powerful framework for understanding how individuals acquire not only linguistic forms, but also the social norms, values, identities, and epistemic orientations associated with language use (Ochs & Schieffelin 1984; Duranti et al. 2012). From this perspective, language learning is not conceived as the internalisation of an abstract system, but as a process of becoming a competent social actor through participation in socially organised practices.

Originally developed in anthropological linguistics and applied primarily to child language acquisition, language socialisation research has increasingly been extended to second / foreign language contexts and adult learning environments (Poole 1992; Duff & Anderson 2015; Duff 2020a). These studies highlight how learners are socialised *through* language and *to* language simultaneously, as they engage in interactional routines that shape both linguistic behaviour and social positioning. Crucially, this process is not limited to classrooms: learning emerges wherever interactional demands require participants to negotiate meaning, align perspectives, and establish shared understandings.

In contexts of mobility, informal and non-formal language learning become particularly salient concepts. Mobile individuals often find themselves in situations where communicative success is necessary for everyday functioning – professional collaboration, social belonging, access to services – yet formal instruction is absent or marginal, and the received school education often insufficient (as survey results have confirmed, see 4.4.3). Under these conditions, learning unfolds through situated interaction, driven by communicative necessity rather than pedagogical design. This aligns with research showing that second language development is deeply intertwined with participation in social practices, especially when learners are required to manage real communicative stakes (Ely & Gleason 1996; Taguchi 2019).

Within this framework, interactional phenomena such as negotiation of meaning, clarification requests, reformulations, and mediation strategies play a central role. In second language acquisition research, these processes have been described as forms of *functional negotiation*, whereby speakers collaboratively resolve communicative trouble by adapting linguistic and semiotic resources in real time (Poole 1992). Such practices resonate with the notion of mediation as articulated in the CEFR Companion Volume (Council of Europe 2020), which distinguishes between solicited and unsolicited mediation, as well as between conceptual and

communicative mediation. Importantly, mediation is not framed as an exceptional skill, but as a routine component of communicative competence in multilingual settings.

From a language socialisation perspective, these mediation practices can be understood as sites of learning, where participants are exposed to new linguistic forms, pragmatic norms, and interactional expectations. Learning, in this sense, is not necessarily conscious or cumulative in a linear way; rather, it is embedded in the repeated experience of communicative problem-solving. Studies in classroom discourse have shown how such processes contribute to the socialisation of affect, authority, epistemic stance, and identity (Poole 1992; Guirado et al. 2025). Similar mechanisms can be observed in informal settings, where norms are less explicit but no less consequential.

The relevance of language socialisation theory is further amplified when considering transcultural communication as a fertile ground for informal learning. As discussed in the previous section, transcultural interaction is characterised by fluid norms, asymmetrical competencies, and hybrid repertoires. These conditions create frequent occasions for explanation, reformulation, translation, and metacommunicative reflection. Rather than being peripheral, such practices constitute a central mechanism through which participants expand their communicative repertoires and develop adaptive competence over time.

Importantly, informal language learning in mobile contexts is also shaped by institutional and ideological dimensions. As Ball (2012) argues, learning processes are embedded in broader policy regimes and power structures that regulate access to legitimate knowledge and valued forms of competence. Even in informal settings, participants orient to dominant language ideologies—such as monolingual norms or native-speaker standards—which influence what counts as successful communication and legitimate learning. This is particularly relevant in contexts where English operates as a lingua franca, carrying both pragmatic utility and ideological weight.

Recent work has therefore emphasised the need to conceptualise language socialisation as a dynamic, multi-scalar process, operating across micro-interactional practices and macro-social structures (Duff 2020b). From this perspective, informal language learning cannot be reduced to incidental exposure or individual cognitive uptake; it must be understood as a socially organised process shaped by participation frameworks, power relations, and communicative ecologies.

In line with the theoretical orientation adopted in this chapter, informal language learning is thus approached as an emergent outcome of linguistic fluidity. It arises at the intersection of translanguaging practices, mediation strategies, and technology-augmented communication, and is sustained through repeated engagement in transcultural interaction. Rather than representing a separate domain of analysis, language socialisation provides a connective lens that links communicative practice to longer-term trajectories of competence development.

This perspective is particularly relevant for the empirical investigation developed in the following chapters. By focusing on naturally occurring interaction among mobile individuals, the analysis can trace how learning opportunities are embedded in everyday communicative practices, without presupposing formal instructional intent. In this sense, language socialisation, paired with metatranslinguistic conciousness, offers a crucial lens on the relationship among linguistic fluidity, transcultural communication practices, and speakers' attitudes, metalinguistic awareness, and agency.

3.6 A theory of fluidity: Towards a practical definitional of language fluidity

This section represents the final step of the chapter by integrating the theories discussed so far for a practical and operational definition of *linguistic fluidity*. The concept is central to the theoretical framework of this thesis, as it allows for a systematic yet non-reductive account of language-in-use in contexts characterized by high mobility, superdiversity, and digital mediation. Rather than referring to instability or lack of structure, linguistic fluidity is here conceptualised as a dynamic mode of language organisation and use, emerging from the interaction between speakers' repertoires, sociolinguistic systems, and communicative ecologies.

Building on mobility studies and the sociolinguistics of complexity (Blommaert, 2010, 2014), linguistic fluidity is understood as a consequence of the deterritorialisation of linguistic resources and of their continuous re-entextualisation across contexts, interactional settings, and semiotic modes. In this sense, fluidity does not imply the absence of norms, but rather the coexistence, negotiation, and contingent activation of multiple norms, values, and expectations within communicative practices.

From this perspective, languages are not conceived as bounded, autonomous systems, but as historically and socially sedimented constructs that are constantly reshaped through use. As argued in translanguaging theory, speakers do not simply alternate between discrete linguistic codes; rather, they draw on an integrated, holistic repertoire of semiotic resources to make meaning, negotiate understanding, and position themselves socially (García & Wei, 2014; Wei 2018). Linguistic fluidity therefore emerges at the intersection of languaging processes, speaker agency, and contextual affordances.

A crucial aspect of this conceptualisation is that linguistic fluidity is not treated as an individual cognitive trait, but as a relational and sociolinguistic phenomenon. It is produced and recognised within interaction, shaped by participants' biographies, language ideologies, and shared assumptions about communication. In line with language socialisation research (Duranti et al. 2012), fluid language practices are learned, reinforced, and transformed through participation in communities of practice, often outside formal educational settings. This makes linguistic fluidity particularly relevant in informal, translocal, and digitally mediated communities, where norms are negotiated rather than institutionally imposed.

According to the integrated theoretical framework here presented, linguistic fluidity can be operationally described as contingent changes and variation emerging through transcultural communicative practices[12].

Conceptualising linguistic fluidity in this way allows for a coherent linkage between translanguaging theories and a usage-based perspective of linguistic analysis. As communities characterised by voluntary mobility, high degrees of agency, and intensive use of digital technologies, digital nomads constitute a particularly revealing site for observing linguistic fluidity. In turn, this concept becomes key to understanding broader processes of social organisation, identity construction, and meaning making in contemporary mobile societies.

[12] This definition is inspired by, among other works, Jasper & Madsen (2019), and Prinsloo (2023).

Chapter 4

Methodology: Triangularity, FAIR data and resources creation

4.1 Introduction to project methodology

The main aim of this chapter is to describe the methodological architecture through which data have been generated, managed and processed throughout the research project, and to illustrate how different forms of evidence – i.e., ethnographic observations, survey and language data – can be structured as reusable language resources, strategically integrated to produce a reliable representation of linguistic fluidity and transcultural communication processes.

Because digital nomadism represents an emergent and rapidly evolving sociolinguistic environment, its study demands a methodological design capable of accommodating novelty, complexity and unpredictability. To achieve this, the chapter combines well-established traditions in sociolinguistic research with recent developments and innovative approaches to research.

In recent years, the role of data within academic research has undergone a profound transformation. At a European level, frameworks such as Horizon Europe explicitly recognise research data as a strategic asset that must be properly managed, preserved and disseminated. The European Research Area now promotes a research culture centred on transparency, openness and reproducibility, encouraging researchers to plan from the outset how data will be stored, documented, shared and re-used.

The datasets collected are conceived as language resources, i.e. machine-readable objects that contain rich linguistic information and are accompanied by documentation enabling future searchability, interoperability and reusability (Godfrey & Zampolli 1997). The creation of such linguistic resources for the study of informal translingual communication must therefore be understood not only as a methodological choice but also as a strategic response to this changing landscape.

Within sociolinguistics, the attention to the creation, curation and dissemination of language resources reflects broader disciplinary trends. Sociolinguistics has always been an interdisciplinary science, combining theories and methods from linguistics, anthropology, sociology, ethnography and statistics. Since the 1950s, the period in which the term *sociolinguistics* gained traction as a label for a scientific field (Spolsky 2011), researchers have emphasised the necessity of integrating linguistic and social data. Early sociolinguistic work already required forms of structured evidence that crossed disciplinary boundaries: recordings of natural speech, interviews, ethnographic observations, demographic metadata and contextual annotations.

Throughout the second half of the twentieth century, the field expanded to include experimental, statistical and qualitative approaches, allowing researchers to establish correlations between linguistic forms and social categories (gender, class, mobility, style) and to refine the methodological apparatus needed to handle increasingly complex data. Variationist sociolinguistics (Labov 1992, 2001) formalised rigorous procedures for data collection and quantitative analysis, while ethnographically oriented approaches (Hymes 1974; Gumperz 1982; Duranti 1997; Rampton et al. 2004) insisted on the need to examine communicative behaviour within its social, cultural and interactional context.

The contemporary study of mobility, superdiversity and globalisation (Blommaert 2010, 2013; Pennycook 2010, 2012) pushes these methodological traditions further by foregrounding complexity, unpredictability and fluidity. As Blommaert argues, the social environments in which communication unfolds today require researchers to 'employ several types of ethnographic and textual materials' to do justice to organisational constraints and contextual dynamics (2013: 309). Simple correlations are no longer sufficient; they must be grounded in detailed ethnographic analysis (Hymes 1996). Moreover, mobility as a central concept has three methodological consequences: (a) it introduces unpredictability in what we observe; (b) it requires close ethnographic inspection of communicative practices; and (c) it highlights the limitations of existing methodological vocabularies, demanding new conceptual tools. These insights are central to the present thesis, which investigates linguistic fluidity in mobile, transcultural, digitally mediated contexts.

Against this background, this chapter situates the development of the FLO project, the European Fluid Languages Observatory within contemporary debates around data management, triangulation and mixed methods. The project responds both to sociolinguistic and open science requirements by designing a methodology capable of generating reliable, ethically grounded and richly documented language resources.

A key guiding principle of this research is triangularity (Denzin 1978), understood as the integration of multiple types of data and methods to capture a multifaceted social and linguistic phenomenon. In the present research triangulation is conceived as:

- Data triangulation: using different types of data (spoken interactions, fieldnotes, interviews, surveys).
- Methodological triangulation: combining qualitative and quantitative tools (ethnography, netnography, corpus analysis).
- Investigator triangulation: incorporating perspectives from participants or collaborators within the community.
- Theory triangulation: bridging together different theoretical traditions (see Chapter 3) interpreting data using complementary theoretical frameworks.

Digital nomad communities are mobile, fluid, and culturally heterogenous. Triangulation allows the research to embrace complexity, producing a richer and more accurate representation of linguistic practices in transcultural settings.

The remainder of this chapter is organised into three main sections. The first section engages with FAIR and Open Science principles to discuss project design, data management planning, ethical and legal aspects of working with human participants. In the second section the results from the preliminary research phase are reported, as the basis on which the final methodological protocol was developed. Then, the third section describes the main methodological strands of the project, namely ethnographic fieldwork, survey research and corpus linguistics. It should be noted that the methodological design was progressively refined through continuous interaction with the data. For this reason, selected empirical observations are introduced in this chapter as illustrative elements that informed and shaped methodological decisions, and references are provided for the datasets resulting from each methodological step. A in-depth analysis of these findings is then presented in Chapter 6.

4.2 Starting the FLO: The European Fluid Languages Observatory

4.2.1 FLO project overview

The European Fluid Languages Observatory (FLO) is the conceptual and methodological framework developed to guide the collection and analysis of linguistic data for this project. Not a formally established observatory, but with the longer-term ambition of evolving into a fully-fledged observatory that supports knowledge co-creation on transcultural communication, FLO has a twofold mission: (i) to provide a systematic methodology for collecting, curating, and sharing language resources representing transcultural communication, including but not limited to natural language data, in accordance with FAIR and Open Science principles; and (ii) to enhance knowledge co-creation and sharing about transcultural communicative practices and translanguaging beyond the boundaries of academia.

The observatory originates from the recognition that the study of transcultural communication in superdiverse mobile contexts requires both new methodological tools and innovative forms of resource creation. Traditional sociolinguistic data collections have focused on communities defined by geographical and cultural stability. By contrast, digital nomadism presents a scenario of deterritorialised subjectivities and superdiverse cultural settings, where speakers deploy plurilingual repertoires in ways that disrupt conventional concepts of *native speaker* (Bloomfield 1933; Rampton 1990), or *speech community* (Gumperz 1968; Labov 1972). FLO thus addresses a double gap: the lack of systematically collected and documented linguistic resources on digital nomads, and the scarcity of research infrastructures in fields of linguistics tailored to translocal – and translingual – communities (Odijk & Van Hessen 2017; Calzolari et al. 2020).

The early stages of the project (2022–2023) combined literature review (including academic publications, blogs, marketing campaigns, and forum discussions, an account of which is provided in Chapters 2 and 3) with pilot ethnography and netnography (here discussed in Section 4.3) to capture the phenomenon of digital nomadism from within. Taking advantage of institutional scholarships like the Marco Polo Programme active at the Department of

Translation and Interpreting of the University of Bologna and the CIUTI Travel Scheme, FLO evolved into an international research project, hosted by – besides its main promoter, the University of Bologna – the University of Madeira, the University of Las Palmas de Gran Canaria, and KU Leuven.

The vision behind the whole project is that research on digital nomad communicative behaviours can both contribute to scientific progress in the field of linguistics and produce a positive impact on language education and policies. FLO tries to do so by investigating communicative behaviours and the practices, like the use of linguae francae, intercomprehension and mediation strategies, employed by highly mobile individuals in transcultural community settings, here specifically represented by digital nomad communities.

The project fully embraces the digital turn that has characterised humanities studies in the last decades, integrating data-intensive methods with grounded ethnography and participatory approaches. The outputs include open-access FAIR resources (survey datasets, the NomadLingo corpus), analytical insights, dissemination through digital media, and civic activities.

In practical terms, the observatory has the role of setting up methods to manage multiple kinds of research data. These include ethnographic notes and interviews, survey responses, and recordings of naturally occurring conversations. Each data type comes with specific challenges which are treated in greater detail in the dedicated sections; for instance, spoken data require careful transcription and anonymisation; survey data require privacy-compliant software; and netnographic material must be archived with explicit consent and appropriate licences. By explicitly mapping the heterogeneity of its datasets, FLO aims to propose a dynamic – and surely improvable – model for contemporary sociolinguistic research, which requires hybrid protocols of data collection and analysis.

Summing up, FLO is organised as a complex net of research and social impact activities, mainly developed around the following strands:

1. Community-based participatory ethnographic and netnographic fieldwork, primarily in Madeira and the Canary Islands, documenting everyday communication and community dynamics (discussed in Section 4.4.2);
2. Survey research, producing a quantitative dataset (*Where's your language at?*) mapping repertoires, mobilities, and attitudes towards translanguaging, with a particular attention to English used as a lingua franca in transcultural communicative contexts (see Section 4.4.3);
3. Corpus creation, i.e. NomadLingo, a spoken corpus of naturally occurring translingual interactions among digital nomads, annotated for features such as code-switching, translanguaging, negotiation of meaning, and repair strategies (which will be treated in greater detail Chapter 5);
4. Dissemination of translanguaging theories for the general public, through both online social interaction tools like thenomadlinguist.eu website, and workshops, radio

interviews, and non-acedemic educational events (like the European Youth Event organised in Forlì in 2024).

These perspectives converge in a triangular design (ethnography, survey, corpus), which underpins FLO's claim to methodological innovation and transdisciplinarity. However, before delving into the discussion related to data collection procedures (see Section 4.4 and Chapter 5), the reminder of the present section (i.e, Sub-section 4.2.2) provides an overview on general data management and project specific features, including research personal branding through the website *The Nomad Linguist* and its associated social media channel. The latter dimension reinforces the participatory ethos of the project, further engaging digital nomads not merely as research subjects but as interlocutors and co-constructors of knowledge.

4.2.2 Creating FAIR & OPEN language resources

Data management within the FLO project follows the FAIR principles – Findability, Accessibility, Interoperability, and Reusability – first formulated by Wilkinson et al. (2016) to ensure that scientific data can be easily discovered, cited, and reused by both humans and machines. These principles, today adopted across all scientific domains, constitute one of the cornerstones of Open Science, a movement that seeks to make scientific research more transparent and collaborative.

The roots of Open Science can be traced back to the Open Access movement of the late 1990s and early 2000s, which advocated for free and unrestricted access to scholarly publications (Suber 2012). Over time, the focus expanded from publications to all forms of scientific output – data, software, protocols, and even peer review – leading to what the European Commission (2020) defines as 'a systemic change in the modus operandi of science and research.' FAIR principles emerged in this context as a technical and ethical framework that translates the ideals of Open Science into concrete operational criteria.

According to the European Commission Recommendation on Access to and Preservation of Scientific Information (2018/790/EU), FAIR data management should guarantee that data are *findable* through persistent identifiers and rich metadata; *accessible* through standard protocols; *interoperable* via shared vocabularies and formats; and *reusable* thanks to clear licences and documentation. Importantly, FAIR does not imply that all data must be fully open: rather, data should be 'as open as possible, as closed as necessary' (European Commission 2018). In sensitive fields such as linguistics and social sciences, where data often derive from human participants, this principle enables a balanced approach between openness and ethical responsibility.

In Italy, the Open Science agenda has gained institutional momentum through initiatives led by CNR**,** CRUI**,** and the national plan PNSA – Piano Nazionale per la Scienza Aperta (2022), which explicitly aligns national research practices with the European Open Science Cloud (EOSC) and the FAIR data framework. Italian universities, including the University of Bologna, have adopted internal Open Science policies promoting data management planning,

repository use, and FAIR-aligned training for doctoral researchers. FLO thus positions itself within this evolving ecosystem, aiming at contributing to the practical implementation of these policies in the field of linguistics.

In linguistics, the concept of FAIR data intersects with the long-standing tradition of language resource creation, defined as the systematic collection, annotation, and dissemination of linguistic data for research and technological development (Godfrey & Zampolli 1997; Monachini & Frontini 2005). However, whereas early corpora were often proprietary or distributed in isolated formats, the FAIR approach calls for integrated, interoperable, and reusable datasets that can support replication, cross-study comparison, and long-term preservation.

Throughout the FLO project, Open Science is understood not just as a policy framework but as a methodological stance. As underlined in the EU guidelines, Open Science promotes transparency, inclusion, fairness, and sharing (European Commission 2020). In this project, corpora and datasets are created with the intention of promoting their use in institutional policy planning, in applied linguistics, including foreign language teaching and plurilingual competence development, in research settings, for sociolinguistic analyses as well as for linguistic tools development, and as a methodological example of triangular methods for language documentation (see Chapter 7 for a comprehensive discussion on the foreseeable uses of FLO language resources).

For clarity reasons, this section is organised into thematic paragraphs which link together general considerations about FAIR and OS principles with the practical organisation and workflow adopted in the FLO project. First, the role of research infrastructures, general and disciplinary repositories is presented; then, the Data Management Plan is introduced as the tool supporting the overall project organisation and data management; finally, legal and ethical concerns of working with data produced by human participants is analysed.

Research infrastructures, repositories and CLARIN ERIC

The principles of FAIR and Open Science cannot be realised without appropriate research infrastructures – the technological, organisational, and legal frameworks that enable research and its output to be developed, properly documented, shared, and preserved. In linguistics, infrastructures play an essential role in ensuring that resources are not only open but also responsibly managed and sustainably maintained (Science Europe 2021). Infrastructures can thus be understood as the overarching systems that make Open Science possible, and repositories as their most tangible manifestation.

A repository is a trusted digital archive where datasets and related documentation are stored and made findable through persistent identifiers (such as DOIs), searchable metadata, and standard access protocols. Repositories serve a dual function: they safeguard research outputs and make them interoperable and citable, allowing other scholars to build upon them. From a FAIR perspective, repositories operationalise the *four pillars*: (i) findability, through metadata

indexing and persistent identifiers (PID); (ii) accessibility, by applying open licences and standardised retrieval interfaces; (iii) interoperability, via shared schemas (e.g., Dublin Core, CMDI, TEI) and by promoting the use of open over proprietary file formats; (iv) reusability – through detailed documentation and versioning (CoreTrustSeal 2023). In the humanities and social sciences, repositories also embody the ethical dimension of responsible openness, balancing transparency with participants' rights. This is particularly relevant for the FLO project, whose data include personal interactions, field notes, and spoken corpora derived from spontaneous communication among digital nomads. Repositories allow the project to share metadata and derived datasets (e.g., anonymised transcripts, codebooks) while restricting access to sensitive materials (e.g., raw audio or identifiable images). This selective openness aligns with Horizon Europe's policy on open data (European Commission 2018).

Repositories can be institutional, disciplinary, or general-purpose. Institutional repositories (such as AMSActa at the University of Bologna) ensure local accountability and compliance with university Open Science policies. Disciplinary repositories (such as CLARIN repositories for languages/linguistics) maintain domain-specific standards, metadata schemas, and preservation workflows. General-purpose repositories (such as Zenodo) facilitate cross-disciplinary sharing and provide technical tools for versioning, collaborative editing, and DOI assignment.

In the context of European research policy, infrastructures are not only storage systems but strategic instruments for coordination and knowledge circulation across Member States. The European Strategy Forum on Research Infrastructures (ESFRI), established in 2002, defines them as 'facilities, resources or services of a unique nature that are used by research communities to conduct top-level activities in their fields' (ESFRI 2021). Within the European Research Area (ERA), these infrastructures, called ERIC, are designed to promote excellence, interoperability, and sustainability across disciplines. The shift towards digital and distributed infrastructures has been a defining feature of European science policy since the early 2000s, culminating in the creation of the European Open Science Cloud (EOSC), a federated environment for sharing scientific data in accordance with the FAIR principles (European Commission 2020). The ERIC legal framework – European Research Infrastructure Consortium, established by Council Regulation No. 723/2009 – represents the EU's most advanced model for managing such initiatives. An ERIC provides legal personality and governance mechanisms that enable long-term collaboration among states, ensuring that research infrastructures operate beyond the temporal and financial limits of individual projects. ERICs represent the highest form of legal and organisational recognition for cross-national research infrastructures. They allow Member States to jointly fund and govern large-scale facilities that provide services to scientific communities across Europe. Each ERIC is a public legal entity that operates under a distributed model, integrating national nodes into a common European infrastructure. ERICs are essential to the European Research Area (ERA) because they guarantee both scientific excellence and administrative stability. They ensure that resources, expertise, and standards are shared across borders. In the field of linguistics and language technology, the most relevant ERIC is CLARIN – the Common Language Resources

and Technology Infrastructure – which since 2012 has served as the European backbone for digital language data.

CLARIN ERIC (Broeder et al. 2010; Odijk & van Hessen 2017) was established to give researchers in the humanities and social sciences easy and sustainable access to language resources and tools. It federates dozens of repositories and service providers – called CLARIN Centres – across Europe. These centres host data collections (corpora, lexicons, metadata), offer processing tools (for annotation, speech-to-text conversion, concordancing), and ensure interoperability through a shared technical and legal framework. CLARIN's structure is fully aligned with the FAIR principles. Each dataset deposited within a CLARIN Centre is assigned a persistent identifier (PID) and described using the Component Metadata Infrastructure (CMDI). Licensing and access are handled through the CLARIN Licence Category System, which clearly distinguishes between open, restricted, and academic-use-only resources. The federated search interface, the Virtual Language Observatory (VLO), enables users to find resources across all national centres, while the Language Resource Switchboard supports direct processing through online tools.

CLARIN's significance lies not only in its technical standards but also in its community governance model: national consortia (e.g., CLARIN-NL, CLARIN-D, CLARIN-IT) manage local centres, ensuring that national priorities and research cultures are integrated within a common European vision. Italy joined CLARIN ERIC in 2015, forming CLARIN-IT, coordinated by the Istituto di Linguistica Computazionale A. Zampolli (ILC-CNR, Pisa). The Italian consortium brings together universities and research institutes (CNR, FBK, University of Trieste, University of Bologna, among others) that contribute data, tools, or methodological expertise. CLARIN-IT's mission is to promote the FAIR management of linguistic data in Italy and to connect national projects to the European infrastructure. In particular, the University of Bologna recently became an active institutional member of CLARIN-IT. This inclusion was directly linked to the research activity developed through the COLITEC research group, which has been the fundational environment where the Fluid Languages Observatory was developed. From a technical standpoint, the Bologna node contributes expertise in TEI-XML annotation, spoken corpus design, and FAIR-compliant metadata management, while benefiting from CLARIN's certified repositories for archiving and dissemination. The case of the University of Bologna illustrates how bottom-up research needs – including those identified by the FLO project – can drive institutional and infrastructural innovation.

CLARIN has its own repository network which consists of a distributed federation of certified data centres, each responsible for the long-term preservation, curation, and dissemination of language resources. All CLARIN repositories adhere to shared standards for metadata (CMDI), persistent identifiers (PID), and licensing (CLARIN Licence Categories). These repositories host a wide range of data types – written and spoken corpora, multimodal collections, lexica, metadata sets, and language processing tools – and are regularly audited according to the CoreTrustSeal certification scheme (2023) to guarantee data integrity, accessibility, and sustainability. The Italian CLARIN consortium (CLARIN-IT) operates several certified repositories distributed across partner institutions, including the Fondazione Bruno Kessler

(FBK), the University of Trieste, and the Istituto di Linguistica Computazionale "A. Zampolli" (CNR-ILC), which hosts the national repository ILC4CLARIN. The latter plays a pivotal role as the primary deposit site for the NomadLingo corpus, a product of the FLO project described here in Chapter 5. ILC4CLARIN provides a stable preservation environment and full integration with the CLARIN federation, ensuring that the datasets produced within the FLO project are findable and reusable across Europe through the VLO.

The advantages of depositing in CLARIN repositories are manifold and include:

- Standardisation – data are formatted and documented using shared schemas (CMDI, TEI);
- Interoperability – datasets can interact with other European collections through VLO and Switchboard;
- Legal clarity – access and reuse conditions are transparent and machine-readable;
- Visibility and citation – every dataset receives a persistent identifier (PID/DOI) and a recommended citation.

For FLO, this means that its outputs – particularly *NomadLingo* – can be integrated into a European catalogue, accessible not only to linguists but also to scholars in sociology, anthropology, and communication.

While repositories ensure long-term preservation, collaborative platforms such as GitHub and the Open Science Framework (OSF) support day-to-day project management and open workflow sharing**.** GitHub, originally designed for software development, provides version control and collaborative editing features, making it ideal for managing evolving datasets, transcription guidelines, and code scripts. As will be described in greater detail in the next sub-section, GitHub hosts FLO's public documentation, metadata templates, and data processing scripts. The Open Science Framework (OSF), developed by the Center for Open Science, offers a complementary function. It serves as a project management hub where data, DMPs, ethics forms, and supplementary materials can be stored and shared under adjustable privacy settings. OSF supports both pre-registration and dissemination, helping researchers track changes and maintain transparency throughout the project lifecycle (Hett & Hett 2013). OSF provides a further access point to resources and publications related to FLO: it links repositories, GitHub resources, and institutional archives into a coherent structure, while providing a project DOI which allow citation and traceability. With such functionalities, these two tools make it possible to trace the evolution of datasets from field collection to open publication.

Summing up, research infrastructures like CLARIN ERIC and collaborative platforms like GitHub and OSF form the structural backbone of FLO's data management strategy. They ensure compliance with FAIR principles while supporting methodological transparency and cross-disciplinary collaboration. The next sub-section describes in detail how these infrastructures are operationalised through the FLO Data Management Plan (DMP), which defines data flows and ethical protocols across all project phases.

FLO Data Management Plan

In line with Horizon Europe guidelines, FLO has developed a Data Management Plan (DMP) as a living document that evolves together with the research. The DMP does not only ensure compliance with research standards and transparency but also methodological consistency. In general, the DMP can be considered as a working document where researchers plan and record decisions about what data will be collected and how, the standards and formats adopted, how and under what conditions data will be shared, privacy and ethical considerations, storage and long-term preservation strategies (European Commission 2016a; Science Europe 2021). In the context of the FLO project, the DMP has been a fundamental instrument allowing for the implementation of FAIR and Open Science principles.

The whole FLO's DMP is included in Appendix 2; what follows is a summary of the main pieces of information contained in the document and an account of the reasoning processes that led to specific choices related to the above-mentioned dimensions of project planning and data management.

A first step in this process was the classification of data according to type, sensitivity, and potential reuse. For example, survey data were planned to be anonymised and stored in structured formats (.csv, .xls) with accompanying documentation, while spoken interactions, which may contain personal data, were planned to be stored in encrypted folders and audio files (.wav) to be released for research purposes and upon request only.

The DMP explicitly addresses the repository question. Depositing data implies the transfer of responsibility from the researcher to the repository. Repositories were selected according to privacy, accessibility, and domain standards. For data resulting from pilot and preliminary phases of research – building blocks for the final methodological protocol – AMSActa (University of Bologna's institutional repository) was considered as the best option. While providing less visibility than more popular repositories like Zenodo, AMSActa has a service for data and documentation checking which guides researchers, in particular those who are new to depositing producers, to embrace FAIR practices. The second dataset deriving from survey research was instead deposited on Zenodo. Finally, naturally occurring language data composing the NomadLingo Corpus were deposited, as planned, onto a domain-specific repository among those provided by CLARIN ERIC (see Chapter 5).

An essential component of the DMP is documentation and metadata. FLO adopts a layered approach, where each dataset is accompanied by a *readme file*, describing its content, structure, and reuse conditions, and metadata regarding authors, contributors, and affiliation. *Readme files* provide essential contextual information that enables other researchers to understand, reproduce, and reuse the data in accordance with FAIR principles (Wilkinson et al. 2016). They typically include details such as data collection methods, file organisation, variable descriptions, data formats, licensing conditions, and version history (AMS Acta 2023). Within the field of corpus linguistics, *readme files* have been recognised as crucial complements to metadata records, offering interpretive guidance about corpus design, annotation conventions, and representativeness limitations (CLARIN 2020; Ide & Romary 2004). As an example, the

readme file of the NomadLingo corpus (more details about this research output are given in Chapter 5) is included in Appendix 10.

As for licensing, all data are by default protected by copyright. FLO therefore associates explicit licences to each dataset (typically Creative Commons Attribution or Attribution-NonCommercial). When no licence is present, users must contact the author for permission (this issue, together with other legal considerations, will be discussed in the dedicated sub-section following the present one).

In Figure 9 a summary of FLO's data management is provided.

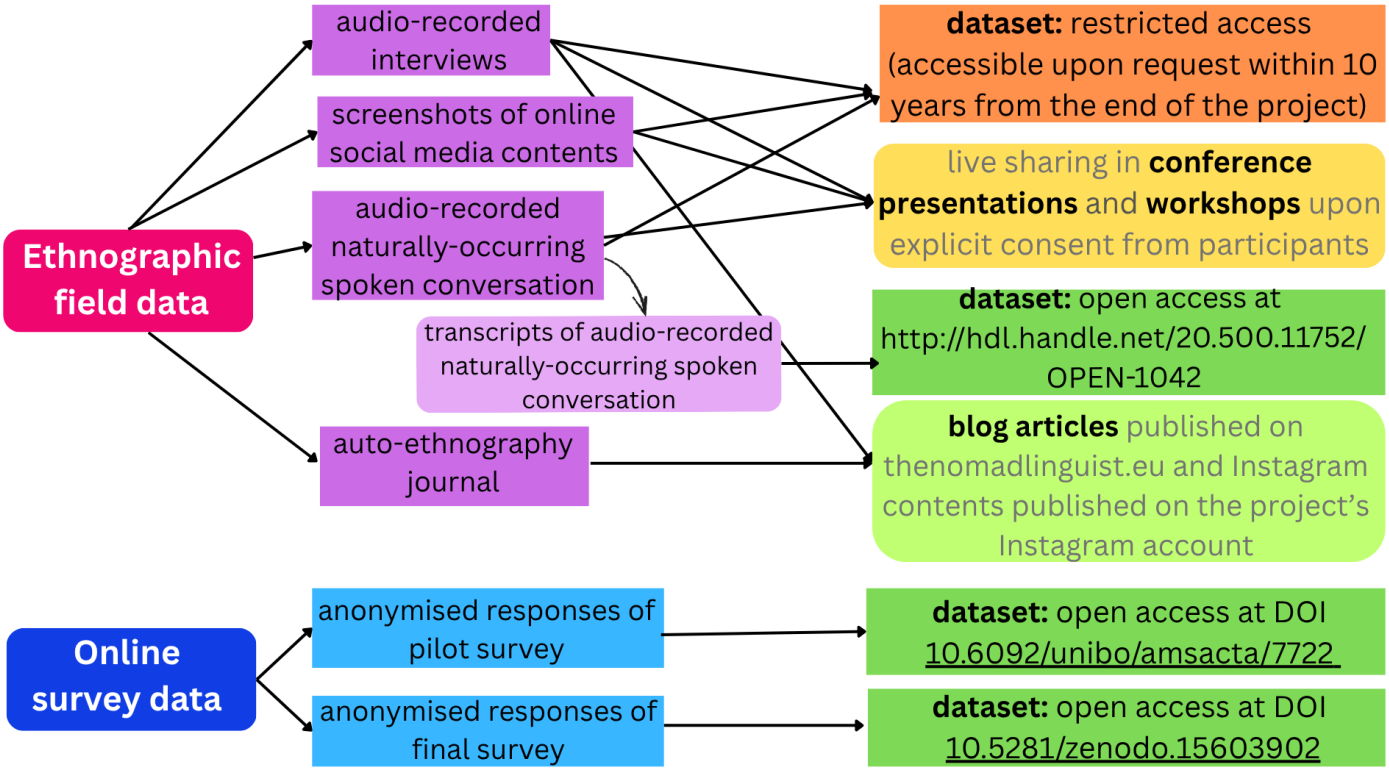


Figure 9. FLO's data overview and accessibility.

The two main sources of data, i.e. field research and online survey, led to the creation of different types of datasets, some of which are published in open access and fully adhere to the FAIR data principles (in the green boxes at the right), while access to some data is restricted to guarantee respect of participants' privacy. Nevertheless, these types of data have informed blog articles and live presentations (which in the figure appear enclosed in the rounded boxes in light green and yellow, respectively).

Summing up, the FLO DMP guarantees that data are not only safely stored during and after the end of the research process, and that they comply with legal requirements, but also that these

resources become accessible and reusable by the broader research community. By doing so, FLO aims to provide amodel for managing sensitive and heterogeneous linguistic data in ways that are FAIR, open, and socially impactful.

Research involving human participants: Legal and ethical aspects

All the data collected within the FLO project derive from direct and/or indirect interaction with and/or observation of human participants. Working with languages from a usage-based perspective, as is the case with corpus methods, inevitably involves handling data that are socially and personally meaningful. As Bucholtz, Liang, and Sutton (2014) remind us, 'language data are not neutral artefacts: they are communicative acts tied to identifiable speakers and communities'. Consequently, researchers must address a range of legal and ethical questions concerning consent, intellectual property, privacy, and copyright before collecting, analysing, or sharing linguistic material.

Informed consent is the cornerstone of ethical research involving human participants. It ensures that individuals understand the nature, purpose, and potential consequences of their participation. In sociolinguistics, this extends beyond experimental tasks to include the use of naturally occurring speech or written interaction (Bucholtz 2007; Bowern 2018), as well as the use of ethnographic and netnographic data, like survey responses, screenshots of contents posted on online social media, interviews and any other data deriving from observation. Importantly, informed consent in ethnographic research presents specific challenges compared to experimental or clinical settings, as participation often unfolds through prolonged observation, evolving relationships, and contextual interaction rather than discrete data-collection events. As Hammersley and Atkinson (2019) note, ethnographic consent is necessarily processual and negotiated, requiring researchers to revisit and reaffirm consent as the fieldwork progresses. In participant observation and linguistic ethnography, consent is frequently situational and contingent, as interlocutors may shift between public and private modes of communication (Tusting & Maybin 2007). When recording naturally occurring interactions, participants should be informed about recording methods, anonymisation procedures, data storage, and potential reuse of their contributions in open datasets. In netnographic research, issues arise due to blurred boundaries between public and private online spaces, making it essential to assess whether online posts or exchanges can be ethically observed or quoted (Kozinets 2020). For survey research, by contrast, informed consent tends to follow a more standardized protocol, often formalized through participant information sheets and digital consent forms, though clarity about data storage and reuse remains crucial (BERA 2018). Across these modalities, consent in the FLO project has been conceived as an ongoing ethical dialogue balancing transparency, participants' autonomy and freedom of choice, and contextual sensitivity.

Specifically, throughout the research processes of data collection consent was obtained in accordance with the European Code of Conduct for Research Integrity (ALLEA 2017) and the University of Bologna Ethical Guidelines. Participants were briefed both orally and in writing, and consent was documented through paper or digital forms (consent form samples

are reported in Appendix 2 alongside the DMP). In informal ethnographic situations – such as fieldwork observations or spontaneous interactions – participants were reminded orally of the project's aims and their right to withdraw consent at any stage.

Furthermore, the issue of intellectual property (IP) in linguistic research arises because the data produced by participants – especially in the case of written texts, or online posts – constitute original expressions protected by copyright law. According to the Berne Convention for the Protection of Literary and Artistic Works (WIPO 1979, art. 2), any text fixed in a tangible form is a copyrighted work. However, as Ide and Pustejovsky point out, most research reuse of language data qualifies as *fair scholarly use* when datasets are anonymised and redistributed for non-commercial purposes (2017). In FLO, the principle of IP respect is operationalised through controlled access and licensing. Data contributed by participants remain their intellectual property, but researchers obtain permission for reuse through consent forms specifying the nature of data sharing (as shown in Appendix 2).

Because linguistic data often contain personal data (e.g. voice, name, opinions), FLO operates under the legal framework of the General Data Protection Regulation (EU) 2016/679. Abbreviated as GDPR, the Regulation defines personal data as 'any information relating to an identified or identifiable natural person,' which clearly applies to most speech data. Following the guidelines of the European Data Protection Board (2020), FLO ensures privacy through pseudonymisation, and secure storage in institutional servers. Metadata and transcriptions are released only after de-identification, in compliance with Article 89 of the Regulation, which allows for scientific processing of personal data under safeguards for subject rights. Audio files are not accessible by the general public since the voice constitutes personal data.

Finally, ethical data management in linguistics requires attention to copyright and licensing, which determine how datasets can be redistributed and reused. As highlighted by Wynne (2017), explicit licensing (e.g. Creative Commons) is crucial to clarify usage conditions and prevent ambiguity in downstream research. Within FLO, each dataset is accompanied by a licence statement and a README file specifying the permitted level of reuse, following CLARIN and Open Science Europe standards. Sensitive or non-anonymisable data are stored under restricted access, while derived or synthetic datasets are shared under open licences (CC BY-NC or CC BY).

All these principles are integrated into the FLO Data Management Plan and have been operationalised across the project's three main datasets:

- the corpus *NomadLingo*, which includes transcribed and pseudoanonymised spoken interactions collected with participant consent;
- the dataset *Where's Your Language At?*, which collects voluntary responses to the survey without identifiable data;
- the netnographic dataset, which has not been deposited as a research output on any repository but has been used to produce descriptions of the communities under study and to inform the analysis of the other types of data. In the very few cases when images

are used, explicit consent has been asked of every person present or involved in its production.

The entire research protocol was reviewed and approved by the Bioethical Committee of the University of Bologna on June 27th, 2023.

Artificial intelligence tools were employed in a highly controlled and ethically informed manner. No full datasets or materials containing personal data were ever uploaded to external large language models or commercial platforms. Instead, automatic transcription was conducted locally using Whisper and WhisperX on the institutional server and through the application WhisperGUI, developed and maintained by the Department of Interpreting and Translation (DIT) at the University of Bologna (DIT 2023). These tools were run exclusively on university servers, ensuring compliance with the General Data Protection Regulation (EU 2016/679) and institutional data policies. AI was thus used as a supportive research aid. In no ways, it replaced human analysis or comprised participant privacy (European Commission 2020; Taddeo & Floridi 2018).

In conclusion, by developing a methodological protocol with strong attention to ethical and legal issues, the FLO project aligns with the FAIR principles and with the European and international frameworks that define responsible research with human participants.

4.2.3 Resource access and project promotion through personal branding

In contemporary research, dissemination activities extend well beyond the publication of final products. Increasingly, research fairness and transparency are achieved by making the processes of investigation – that is methods, datasets, protocols, reflections on foreseeable outcomes and applications – openly available (Burgelman et al. 2019; Pescosolido 2007). This shift is part of the broader movement toward Open Science, which seeks to democratise knowledge by fostering reproducibility, collaboration, and inclusivity (Vicente-Saez & Martinez-Fuentes 2018), on which the preceding sections have focused in more detail. In linguistics and digital humanities, open infrastructures have been shown to increase not only data reuse but also transdisciplinary dialogue and innovation (Scholtz et al. 2024). Further discussion about research outreach and social impact is provided in Chapter 7.

Sharing materials in open repositories ensures that data are findable, citable, and reusable, aligning with the FAIR principles (Wilkinson et al. 2016). Furthermore, disseminating ongoing work through academic portals and social platforms allows peers and non-specialist audience to follow the evolution of the research and contribute feedback – a practice that enhances accountability while also accelerating knowledge transfer (Borgman 2015). Academic events such as conferences, workshops, and public talks also remain central to this ethos, fostering face-to-face dialogue, professional networking, and interdisciplinary cross-fertilisation (Rowe 2018).

Within this framework, personal branding emerges as a complementary practice that strengthens the visibility and social resonance of research. While the term originates in

marketing and psychology (Labrecque et al. 2011), it has recently entered academic discussions as a strategy for communicating one's scholarly identity and for translating complex research into accessible narratives (Graffigna & Barello 2018; Scatena & Gaggioli 2020). Developing a coherent personal brand does not imply self-promotion in a commercial sense, but rather the ability to convey the values, methods, and goals of one's research to diverse publics. This approach was explored during the Summer School 'Transferable Skills: Comunicare e Finanziare la Ricerca' held at Università Cattolica del Sacro Cuore, in which I participated in 2023. The courses addressed topics such as strategic communication, digital storytelling, and research and highlighted how digital presence and narrative consistency can enhance the credibility, approachability, and societal impact of research projects.

The online dissemination strategy of the FLO project operates on two complementary levels:

(1) academic sharing, through open repositories and research infrastructures such as OSF and GitHub;

(2) public communication, through the website The Nomad Linguist and its associated Instagram channel.

This section illustrates how these tools contribute to transparency, collaboration, and accessibility promoting a view of research as a collaborative participatory process.

OSF and GitHub

To ensure methodological transparency and reproducibility, FLO materials, are progressively shared on two open academic platforms, the Open Science Framework (OSF) and GitHub, already introduced in Sub-section 4.2.2.

OSF is a multidisciplinary infrastructure designed to support the entire research lifecycle – from project development, methods and processing codes to data sharing and archiving (Hett & Hett 2013). It allows for version control, integration with institutional repositories, and the assignment of persistent identifiers (DOIs), thereby ensuring long-term accessibility and citation. The FLO OSF project[13] includes documentation such as the Data Management Plan, conference presentations, and links to associated datasets, serving as the project's institutional record. There, you can also find the Appendices to this thesis in their original format.

GitHub, by contrast, is primarily a software-development platform based on Git version control but widely used across the digital humanities for managing collaborative projects and open-source tools (Perkel 2016). The GitHub repository[14] hosts FLO scripts and documentation related to corpus building, data processing, and annotation workflows, as well as Jupyter Notebooks for linguistic analysis. The repository is organised according to open-science best practices, with a clear folder structure, README files, and metadata describing dependencies,

[13] Project DOI: 10.17605/OSF.IO/UK9NY.

[14] https://github.com/novella-tedesco/FLO.

licences, and usage instructions. OSF ensures the archival stability and citation of research outputs, while GitHub facilitates the continuous, dynamic development of resources and methods.

The Nomad Linguist

Complementing these academic infrastructures, a website named The Nomad Linguist[15] was created as the project's public-facing platform. Conceived as both a professional portfolio and a dissemination hub, it integrates three main areas:

- research – including downloadable outputs such as conference presentations, bibliographic references, and methodological resources linked to the FLO project;
- academic activity – highlighting collaborations, institutional partnerships, and updates from fieldwork conducted in Madeira and Canary Islands;
- participation within the online community of digital nomads – through a blog with posts exploring themes of linguistic fluidity, mobility, and identity, written in accessible English for both scholars and digital-nomad readers.

The site follows accessibility standards and integrates feedback channels allowing readers to contact the researcher or access open materials via online links.

Parallel to the website, the Instagram page[16] functions as a micro-communication environment, in line with the project's participatory nature. Posts and stories (three examples are shown in Figures 10a, 10b and 10c) present visual excerpts from fieldwork, short educational videos on language fluidity, and reflective commentary on research but also nomadic life, blending academic and experiential perspectives. In developing this communication style, inspiration was drawn from digital nomad content creators who use Instagram to share professional journeys and community narratives. Adopting similar visual and narrative conventions, the project could be embedded within the communicative ecology of the communities under study, fostering reciprocity and cultural understanding.

[15] https://www.thenomadlinguist.eu.

[16] https://www.instagram.com/the.nomad.linguist.

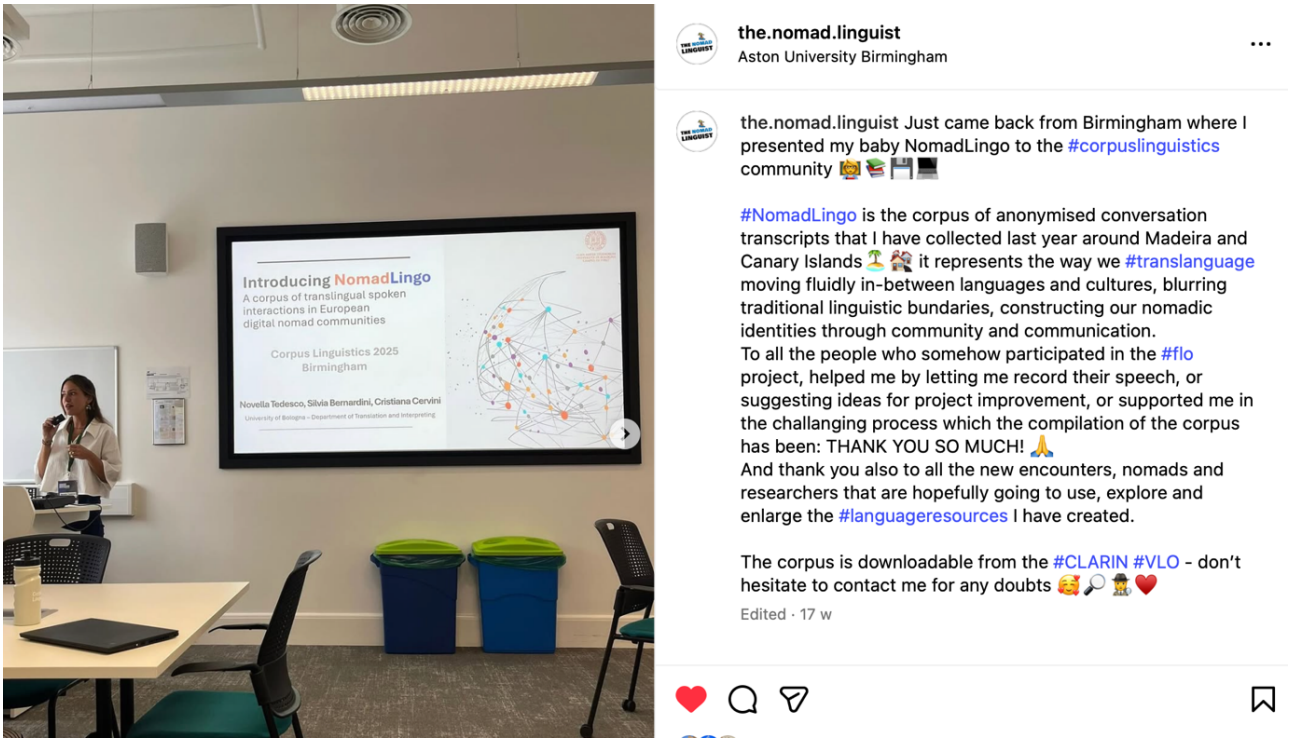


Figure 10a. An Instagram post sharing news about research outputs and professional goals.

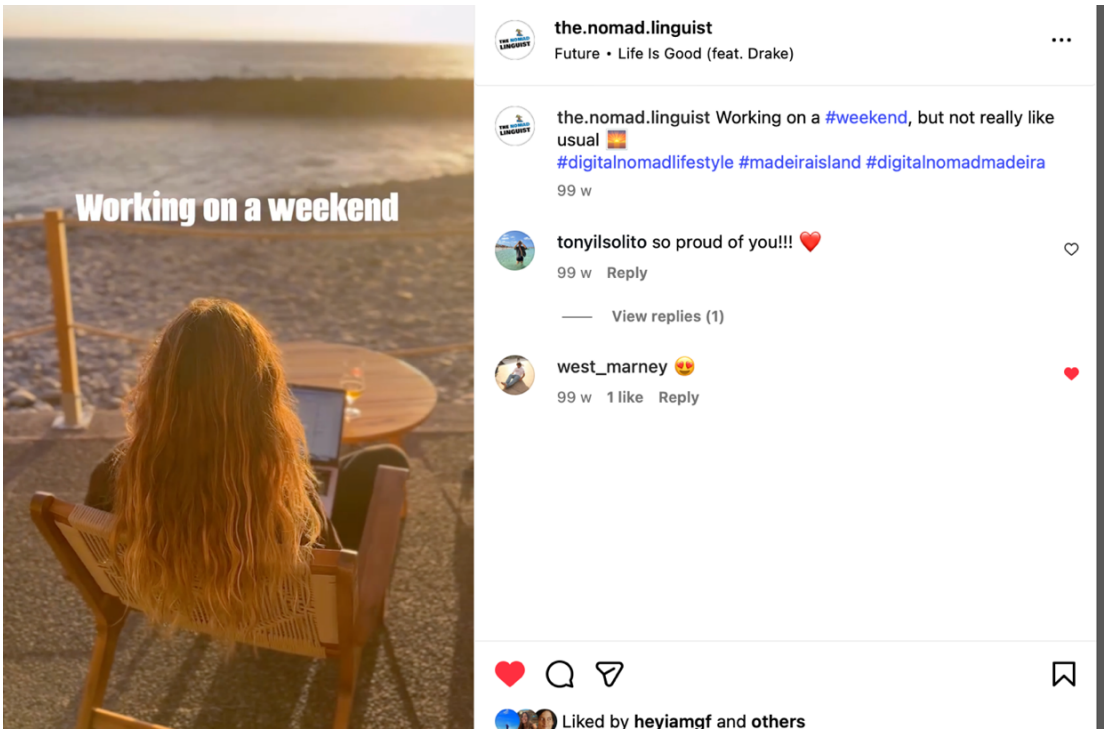


Figure 10b. An Instagram post sharing a personal picture of life as a digital nomad in Madeira.

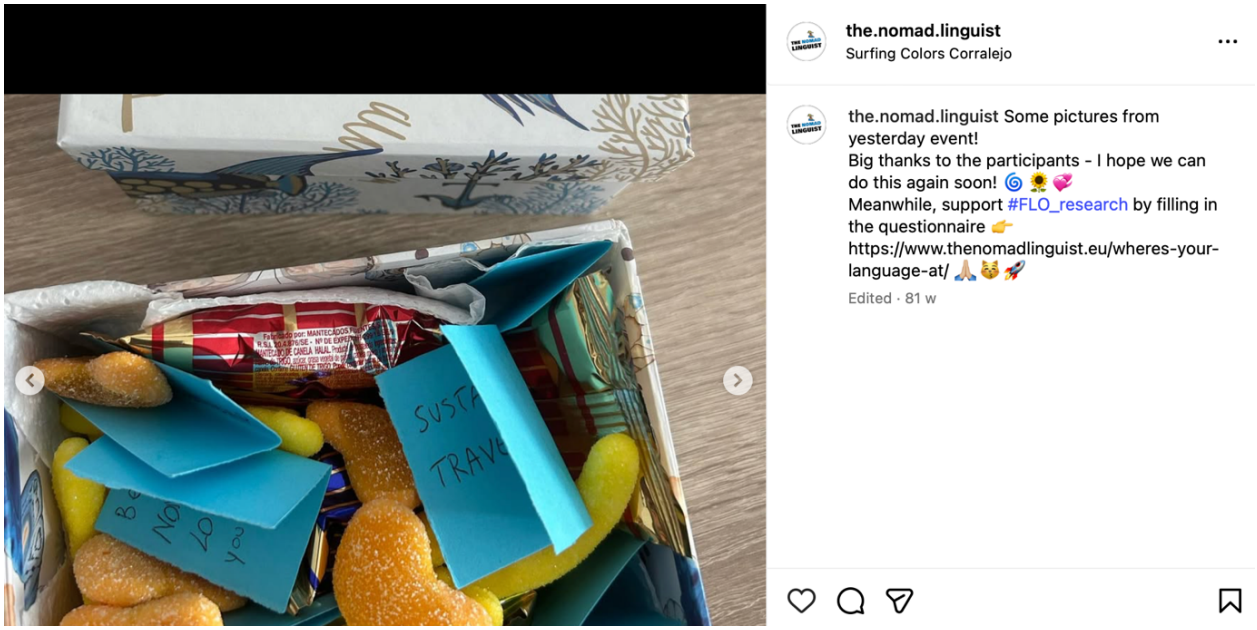


Figure 10c. An Instagram post to thank participants after a data collection event and promote participation in the online survey.

Hence, in the context of FLO, personal branding functions as a bridge between academic communication and social engagement. By adopting accessible language and visual formats, The Nomad Linguist channels enable the project to dialogue with both scholarly audiences and the communities under study. In doing so, personal branding strategies have reinforced FLO's commitment to knowledge co-creation and mutual learning.

The integration of academic and social media platforms underlines one of the project's central premises: that transparency, reflexivity, and social engagement are complementary dimensions of ethical research. Making both data and processes visible encourages responsible critique, while presenting them through accessible channels invites participation and democratizes access to knowledge (Fecher & Friesike 2014). By committing to such communicative activities, FLO shows how personal branding, when consciously applied, becomes not a self-promotional exercise but a strategy for research fairness – acknowledging the labour, contexts, and communities that make scientific knowledge possible.

The alignment between academic infrastructures for open access and participatory dissemination tools, like the blog and the Instagram profile, illustrates how academic communication can evolve toward greater inclusivity and social relevance. In this sense, the project contributes to the debate aro innovative practices, ethical values and social engagement in contemporary linguistic research, moving in a world characterised by always growing openness and interconnectedness.

4.3 Foregrounding Research

4.3.1 Digital nomadism and language

Digital nomadism is a relatively new phenomenon and, until very recently, few studies had been published on the topic. When approaching a topic this new and unexplored, preliminary research is needed in order to identify and locate the most representative research participants, to set clear project goals, and to determine the most interesting factors to investigate and the most effective research practices from a linguistic perspective. Preliminary research was carried out in the first six months of the doctoral project following three main paths: literature review (discussed in Chapter 2), an online survey piloting phase, and field research in two islands of the Canary archipelago – Fuerteventura and Gran Canaria.

This research phase had both a conceptual and an empirical purpose: to verify whether communication could be regarded as a significant issue within digital nomad communities, and to assess how phenomena such as superdiversity, transculturality, and linguistic fluidity – as identified in the literature presented in Chapters 2 and 3 – manifested in actual communication practices of the community.

Foregrounding research is a practice aligning with the exploratory and grounded nature of the entire FLO research design. As Stebbins (2011: 9) explains, exploratory research is by definition 'inductive, empirical, and open-ended,' aiming to develop theoretical understanding through observation and reflection rather than hypothesis testing. Similarly, grounded approaches (Glaser & Strauss 1967; Tusting 2020) emphasise the iterative relation between data and theory, where field encounters progressively shape conceptual frameworks and methodological choices. The foregrounding phase therefore operated as a pilot and diagnostic investigation, confirming the relevance of linguistic inquiry within digital nomadism and helping to identify the most effective methodological tools and entry points into these translocal, loosely structured communities. More specifically, from a theoretical perspective, this preliminary phase exemplifies what can be described as a 'light grounded' approach (Tusting 2020): empirical exploration guided by theoretical sensitivity but open to emergent insights. Rather than imposing predefined categories, the study proceeded from observation to conceptualisation, allowing the key notions discussed in Chapter 3 to be refined through interaction with the phenomenon under study. This process also resonates with participatory and community-based research traditions (Chevalier & Buckles 2013), which stress the importance of co-creating knowledge with participants and adapting research design to the social ecology of the field.

At the outset, some broad guiding questions oriented this first exploration phase:

- Do these communities indeed display a high degree of *superdiversity* (Blommaert & Varis 2015)?
- Is communication a salient issue in digital nomad communities?
- How can digital nomad linguistic attitudes and underlying ideologies be investigated?
- How do they perceive their communicative practices?
- Is translanguaging visible in their interactions?

- How can plurilingualism, translanguaging, and English as a lingua franca be addressed in ways that are relevant and meaningful for linguistic research and for the nomads themselves?

Answering these questions required both mapping the field and understanding its logics of interaction. The preliminary phase thus became a form of situated learning about how to do research among highly mobile, digitally mediated, self-organised groups.

In practice, the foregrounding research unfolded from late 2022 to mid-2023 and combined four main strands:

1. A systematic literature review, comprising over 160 academic publications on digital nomadism and related phenomena such as mobility, lifestyle migration, and the gig economy (see Sub-section 2.2.1);
2. Online exploratory research, analysing blogs, forums, Slack and WhatsApp groups, and dedicated web platforms (e.g. *DigitalNomads.World*, *RemoteTribe*, *NomadList,* also accounted for in Chapter 2);
3. Field research in the Canary Islands (Fuerteventura and Gran Canaria) focusing on direct interaction, observation, and preliminary interviewing;
4. Identification and piloting of research tools (including the extended pilot of the survey *Where's your language at?*).

Each layer informed the next. The literature review clarified theoretical and disciplinary gaps, while online research provided empirical insights into how nomads present, organise, and connect themselves digitally. Fieldwork then allowed these mediated representations to be confronted with lived realities. All the knowledge acquired in the first three steps allowed for the identification of purposeful research tools and methodologies for data collection, which were also tested.

It should be underlined that the fieldwork design of the FLO project originated from both experiential familiarity and considerations of accessibility, diversity, and representativeness within European digital nomadism. The first field site identified for foregrounding exploration was Corralejo, in Fuerteventura (Canary Islands)**,** where I had previously lived for one year as a remote worker. This prior immersion provided an insider's understanding of the community's structure and routines, and facilitated access to local networks. There, I conducted semi-structured interviews with digital nomads of different nationalities, ages, and occupational profiles, focusing not only on their linguistic backgrounds and communicative needs, but also on their mobility trajectories, everyday social practices, and motivations for adopting a nomadic lifestyle. The interviews were conducted in very informal places, some at the beach before and after surf sessions, some others at community social or sportive events.

In the following sub-sections, the main results of the foregrounding research phase are described concerning the issue of language and communication within digital nomad communities, the identification of the most convenient sites for field research, and the

development and piloting of methodological tools, with a particular attention to the online survey which was piloted during this phase.

4.3.2 Developing an operational definition of digital nomad and digital nomad communities

A first essential aim of the preliminary research phase was to develop an operational definition of the object of investigation. As discussed in Chapter 2, the notion of digital nomadism defies easy categorisation, occupying the intersection between work, lifestyle, mobility, and technology (Cook 2023; Dreher & Triandafyllidou 2023). As Riccardo, one of the interviewed nomads based in Fuerteventura, explained, 'I call myself a digital nomad in my bio[17], it represents who I am. Being a digital nomad is not the same as being a remote worker.' For him, this distinction is not only professional but also ideological, associated with values of autonomy, travel, and transcultural awareness.

The term digital nomad broadly refers to individuals who use information and communication technologies to perform remote work while engaging in geographically mobile lifestyles (Birtchnell 2019; Hannonen 2020). As discussed in Chapter 2, scholarly attempts to define and delimit this category reveal persistent ambiguities, with digital nomadism encompassing a wide spectrum of practices and socioeconomic positions. Insights emerging from the present research confirm this heterogeneity, showing how mobility-driven remote work ranges from highly privileged forms – such as tech professionals employed by transnational firms – to more precarious trajectories pursued by freelancers seeking economic or existential stability (Thompson 2019; Mancinelli 2020).

As delineated in Chapter 2, for the purposes of this study, digital nomadism is understood as a voluntary form of mobility characterised by location-independent work and the extensive use of digital technologies for both professional and social purposes. This conceptualisation aligns with Hannonen's (2020) distinction between work-related and lifestyle mobility, and with Moret's (2020) notion of mobility capital, which captures the economic, cultural, and social resources enabling individuals to move according to their own will. In contrast to traditional migration, digital nomadism entails a form of transnational deterritorialised subjectivity (Blommaert 2010), where belonging and identity are negotiated through networks rather than fixed territorial or national affiliations. Moreover, since self-narratives constitute an essential characteristics of digital nomad discourse (Benneau et al. 2023), anyone who defines themselves as digital noamd is regarded as one for the purposes of the present research, which tries to engage with the phenomena under study from an insider perspective.

Summing up, an inclusive operational definition of digital nomads was adopted in the present work, as (1) individuals working in the digital labour market and choosing a traveling lifestyle, often outside their country of residence, or (2) anybody who identifies as a digital nomad. This description, particularly through the first point, allows for the inclusion of other subjectivities besides *canonical* digital nomads, such as expats, digital nomad *wannabes*, remote-working

[17] The term *bio* refers to the Instagram biography provided in the profile.

back-packers, and local young digital entrepreneurs. Indeed, field observations have shown that they do contribute to the community everyday life.

The definition of digital nomad community is equally complex and, to some extent, paradoxical. As Woldoff and Litchfield (2021) note, nomads appear by definition as individualistic and mobile subjects; yet, they actively construct communities that provide stability, belonging, and shared meaning. In Chapter 2 the concept of *light communities* by Blommaert and Varis (2015) has been adopted to describe the social structure of digital nomad communities. Indeed, such communities coalesce around shared practices rather than shared origins or stable membership, and their cohesion derives from intentional participation (Thompson 2019) and shared values of flexibility, collaboration, and autonomy.

In operational terms, this study adopts a working definition of digital nomad communities as networks of mobile professionals who, through online infrastructures, coordinate co-located living and working arrangements in specific destinations, combining virtual interaction and situated participation in local spaces.

The first ethnographic data gathered in Fuerteventura revealed a socially and professionally heterogeneous community. Participants included salaried professionals – such as project managers, financial consultants, and IT specialists – whose stable economic situation allowed them to afford co-working spaces and frequent travel, but in some cases also to buy or rent long-term accommodation, called in their lingo *home-base*. In contrast, younger and more precariously employed individuals, often freelancers in digital marketing, web development, translation, or content creation, articulated their mobility in more existential terms, describing nomadism as both a search for autonomy and a response to instability (see Reichenberger 2018; Hannonen 2020). This internal diversity aligns with recent studies that interpret digital nomadism as a continuum between privilege and precarity (Thompson 2019; Mancinelli 2020), shaped by access to economic and mobility capital (Moret 2020).

The preliminary ethnographic research conducted in the Canary Islands thus confirms that digital nomad communities are:

- digital, maintained through online communication infrastructures, which facilitate networking, knowledge sharing, and coordination across global hubs;
- intentional, where membership and participation are agentive and self-selected, grounded in shared values rather than pre-existing cultural or institutional bonds;
- communities of practice (Lave & Wenger 1991), where learning and socialisation occur through shared engagement in both professional and leisure activities, from co-working and start-up events to informal gatherings and local collaborations.

This multidimensional characterisation allows for an understanding of digital nomad communities as fluid, unstable, and superdiverse (Tedesco 2025a), where belonging is continuously negotiated through everyday social and communicative practices.

Within the FLO project, these operational definitions refine and enlarge the ones proposed in the literature (discussed in Section 2.2), thus providing a framework for identifying, locating, and engaging with the field sites and subjects on which the research focuses.

4.3.3 Identifying sites for field research: Madeira and Canary Islands

This research phase also helped determine the sites for structured ethnographic field research.

Geographically, the Canary Islands emerged as a particularly strategic context. Within the archipelago, Gran Canaria and Fuerteventura host the largest and most active digital nomad communities (Sequera & Nofre 2019), supported by extensive infrastructures of co-living, co-working, and event-based networking**.** These two islands were selected as primary field research sites also for practical and logistical reasons: Gran Canaria was able to host the project through a formal research collaboration with the University of Las Palmas de Gran Canaria, while Fuerteventura was already known to the researcher and offered continuity of access. The geographical proximity between the two islands and frequent transport connections further facilitated research logistics.

Contact with participants was achieved through the use of the Slack and WhatsApp communication channels of the Fuerteventura community, which subsequently led to connections with Gran Canaria online networks. Discussions with local community managers and nomad coordinators suggested that another major hub within the European space is represented by Madeira, a Portuguese archipelago that has become internationally recognised for its Digital Nomad Village initiative in Ponta do Sol (Startup Madeira 2021; Sardinha et al. 2023). Through an initial exploration of the web-based infrastructures of the Madeira Friends community and direct contact with the managers of the Digital Nomad Village start-up, collaboration opportunities emerged. The presence of an academic partner interested in the project, the University of Madeira, further strengthened the suitability of Madeira (main island) as a complementary research site. Additionally, the neighbouring island of Porto Santo, also part of the Madeiran archipelago, was identified as a location hosting numerous nomad retreats and networking events.

For these reasons, the four islands of Madeira, Porto Santo, Gran Canaria and Fuerteventura were chosen as the contexts of investigation for this thesis. Summing up, they were chosen because they host internationally visible nomad hubs with infrastructures that support collaboration, learning, and digital entrepreneurship. They exemplify what Orel (2019) terms the 'touristic-residential continuum', where mobility contributes to place-making and hybrid on/offline communities. Their relevance also aligns with recent analyses identifying Southern European islands as key contexts for studying the intersection of mobility, work, and transcultural community building (Webb 2024; Dreher & Triandafyllidou 2025).

4.3.4 Developing a methodological framework

Conducting foregrounding research had important methodological consequences. First, it showed that the sociolinguistic investigation of digital nomadism required a mixed-methods design combining qualitative insights into both qualitative and quantitive data, such as information gathered through survey and naturally occurring language data collected and analysed in a systematic manner. This triangulation – which will be presented in greater detail in the next section – mirrors Creswell's (2014) notion of sequential exploratory design, where qualitative findings inform quantitative instruments (and vice versa).

In this phase of the research, reflection on my position within the community and the role that my personal experience played in shaping the project became increasingly relevant. Being a partial insider, already familiar with the digital nomad environment, required an ongoing awareness of how my dual position – as both participant and observer – and of how this could affect data collection and interpretation (Davies 2008; Clifford & Marcus 1986). This reflexive dimension, further discussed in the next section, led to the identification of autoethnographic journaling and a first-person presence on social media as fundamental methods for ethnographic inquiry. These tools served not only as records of field engagement but also as means of accessing community discourses and building ethical, transparent relationships with participants (Ellis et al. 2011).

Then, the preliminary research clarified which data types would be feasible, ethical, and informative. Early consideration was given to collecting chat or social-media data; however, privacy concerns, intellectual property issues, contextual incompleteness, and the difficulty of obtaining consent from all the participants in group chats (which, in some cases, are over 500) led to the decision of focusing on naturally occurring spoken interactions which could be recorded during community events, in co-living and co-working spaces. However, netnographic insights were regarded as highly informative so a decision was taken to keep this type of data in the project, as qualitative instances collected through participation in online social networks, an important part in the broader participatory ethnography framework.

Moreover, the intensive use of digital infrastructure to share knowledge within the community and organise community life, as observed in the first phases of the preliminary research, led to the decision of developing an online survey aimed at investigating linguistic attitudes and perceived communicative practices. Thus, the survey named 'Where's your language at' was developed.

What follows is a comprehensive description of the different methods employed during the foregrounding research phase.

Preliminary ethnographic observation

As already mentioned, a preliminary explorative ethnographic phase was carried out at the beginning of 2023 in Gran Canaria and Fuerteventura. This phase included participation to community events and unstructured interviews which were not recorded but accounted for

through written notes. Interviewed subjects included the creator of the app *Diginomad*[18] and Riccardo Molaro, who is a long-standing member of the European digital nomad online network and often visits Fuerteventura digital nomad community. Insights gather through preliminary ethnography were crucial for a better understanding of community dynamics and confirmed the relevance of issues related to communication and languages within digital nomad communities.

Where's your language at? Extended pilot

At this phase of the research project, a survey appeared to be a useful instrument because it could provide better understanding of the demographic characteristics of the population under study (such as, profession, level of education, gender). For this reason, survey research was chosen as one of the FLO methodological strands (see Section 4.4). While more details about survey methodology, developmental phases, final survey administration procedures and results is provided in Sub-section 4.4.3, here the piloting phase of the survey is described and its results are reported. As a matter of fact, the results of the pilot survey, deposited in the University of Bologna instiutional repository AmsActa (Tedesco 2024b), were used to:

- apply all needed changes to improve focus and readability of the actual survey;
- gain first insights into the link between the use of certain communicative practices and specific social contexts of communication;
- gain first insights into hidden ideologies and attitudes towards the use of linguae francae, translanguaging, monolingualism and plurilingualism.

The survey was distributed online through digital nomad networks, consistent with the digital nature of these communities and their high mobility, between October 2023 and April 2024, and gained 24 valid responses. The analysis of these data informed the first scientific contributions deriving from FLO research (e.g., Tedesco 2024; Tedesco 2025a). More specifically, this first version of the questionnaire contains 58 questions divided into three groups – questions on personal information related to digital nomading, linguistic background and needs, and linguistic attitudes respectively. The whole questionnaire accompanies the dataset named *Where's your language at? Extended Pilot Study* (Tedesco 2024b).

Testing and field pre-piloting (van Teijlingen and Hundley 2002, Aithal and Aithal 2020, or Kopper et al. 2023) were carried out to assess question relevance, scale reliability, conceptualisation adequacy, questionnaire readability, and administration efficacy. After testing, carried out by experts, and structured field pre-piloting, with participants from the population under study, the questionnaire underwent many changes, which improved coherence in scales and readability, and limited bias in questions (for example, through the addition of options including a larger number of nationalities). Finally, in October 2023, the pilot version of the survey was launched online to collect responses from digital nomads around the world. The questionnaire was closed after two months.

[18] https://www.diginomad.app/.

Given the openness of digital nomad communities, the most effective sampling method seemed to be convenience sampling (Teddlie and Yu 2007; Bornstein et al. 2013), also employed in the few other surveys conducted on digital nomadism, such as the Flatio Report (2023, 2024). The survey was thus open to voluntary participation, but it was promoted on the Whatsapp and Slack channels of the different local communities, on the Instagram profile *The Nomad Linguist*, and at popular digital nomad spots in Madeira and Canary Islands (Tedesco 2024b). Responses were tracked using IP locations. In Figure 11, the red spots on the map are larger on locations with more responses.

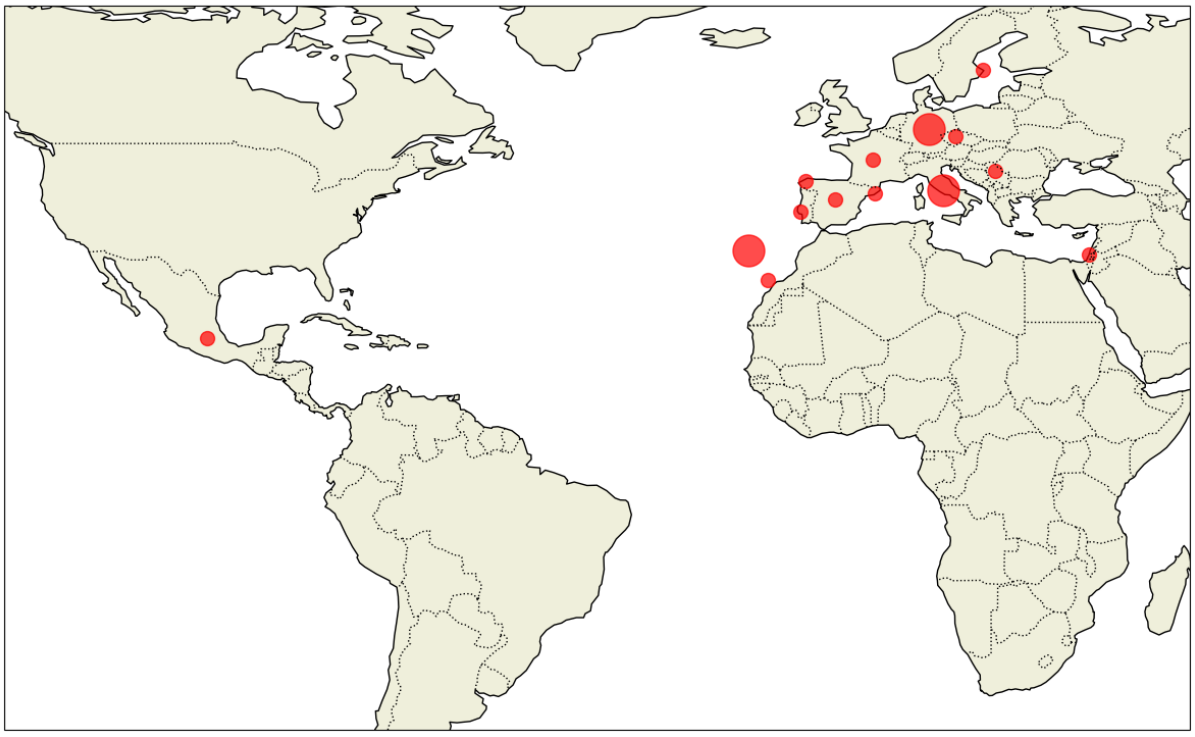


Figure 11. Response location in Where's your language at? PILOT (Tedesco 2024b).

As previously mentioned, the pilot survey provided first informative results on the demographic profiles of digital nomads. While the respondents cannot be considered statistically representative of the whole digital nomad population because of the chosen sampling procedure and on the limited number of participants, still responses confirm the diversity in socio-economical and cultural background of the population under study.

After a first screening, 24 respondents out of 25 were considered as representative of the population under study, all of whom are European residents and only 4 do not have European origins[19].

Drawing on the operational definition discussed in Sub-section 4.3.2, seven criteria were established, based on European statistics on tourism as well as on the literature about digital

[19] The only excluded respondent, despite showing a positive attitudes towards travelling, does not regard herself as a digital nomad, is a student, and spend a lot of time in her country of residence, which is also the country where she was born.

nomadism. The first criterion is the difference between nationality and country of residence. The second and third criteria relate to the number and length of trips made in the last two years. The fourth criterion is related to the motivations for travelling. Although during research some cases of Ph.D. and Master's students conducting a digital nomad lifestyle were reported, the general definition of digital nomads refers to 'professionals' (Reichenberger 2018; see Section 2.2), therefore main occupation (not study) was set as a further criterion. Despite acknowledging the strong diversity (Cook 2023), studies on digital nomad occupation suggest that common areas of work are ICT, marketing and digital programming industry (Cook 2023; Flatio Report 2023), hence these fields of work were adopted as sixth criterion. Lastly, a neutral or positive attitude towards constant movement was regarded as a possible indicator of belonging within the digital nomad community. We considered any responses meeting four or more of these criteria as representative for the purposes of our investigation. The results presented from now on refer to this subset of participants, which are called 'respondents'. Among these, a subset of 18 participants regarded themselves as digital nomads; however, as the Flatio Report (2023) suggests, not all digital nomads identify with this label for both professional and personal reasons.

It should however be noticed that the process here outlined for checking participants' belonging to the social group under study was found to be difficult and time-consuming, hence it was replaced by three direct questions on self-definition, remote work and travel habits in the final version of the questionnaire, as will be explained in the next section.

In Table 4, Figures 12, 13a and 13b a general characterisation of the above-defined subset of respondents is provided.

Table 4. Demographic information about respondents in *Where's your language at? Extended pilot*.

Male	Female	Age (median)	Countries visited in the last 12 months (median)
9	15	33 StDev: 5.87	6 StDev: 3.33

The country of residence of the respondents is shown in Figure 12; red spots are larger on countries where more participants reside.

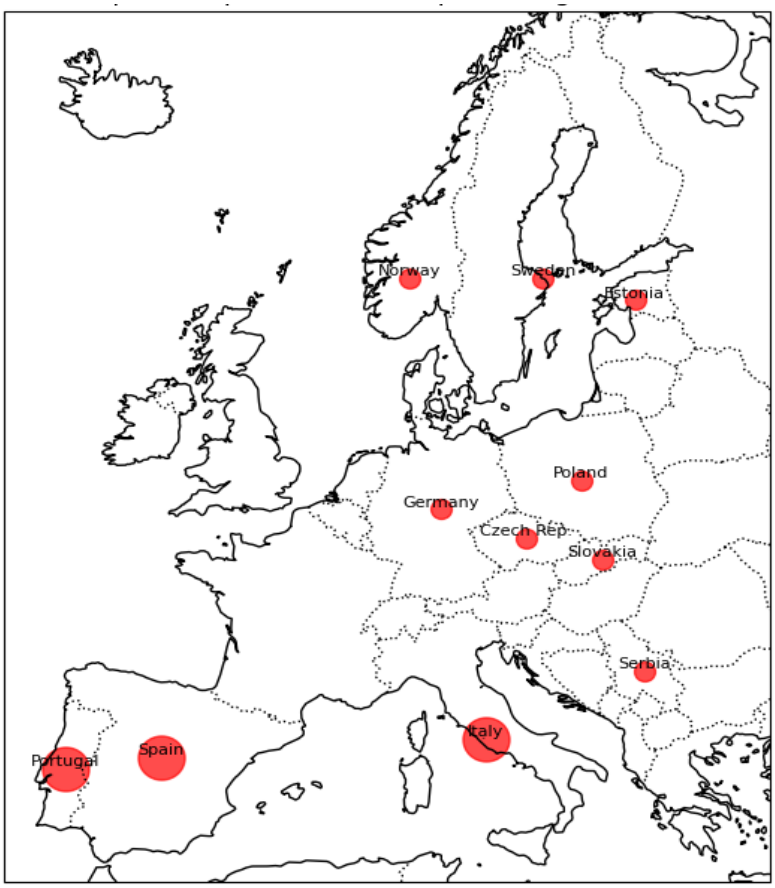


Figure 12. Countries of residence of respondents to *Where's your language at? Extended Pilot*.

In Figures 13a and 13b information about respondents' educational background is provided.

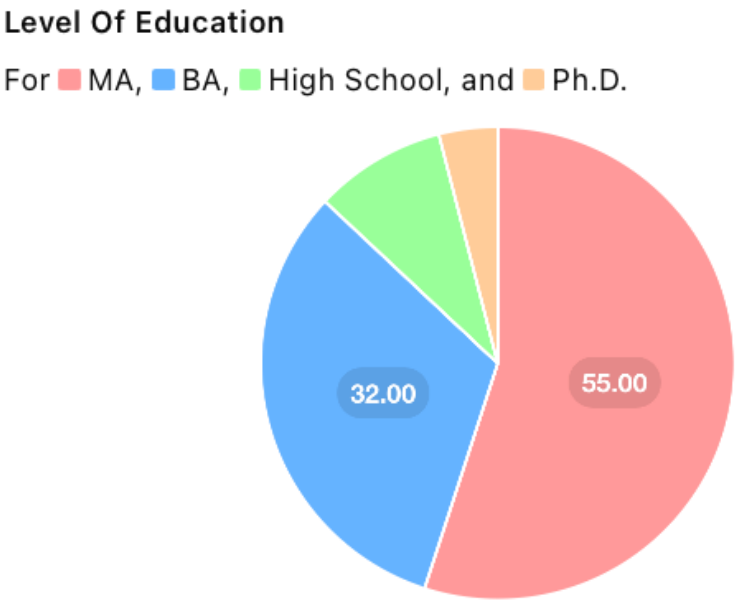


Figure 13a. Respondents' level of education in *Where's your language at? Extended pilot*.

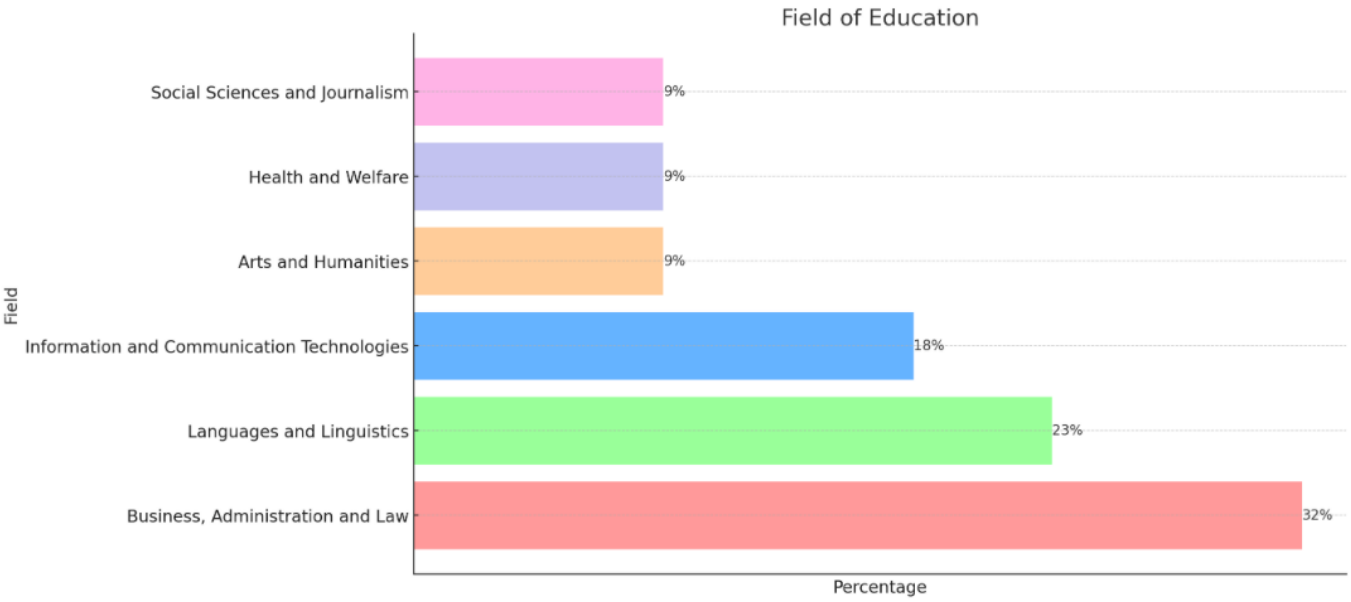


Figure 13b. Respondents' field of education[13] in *Where's your language at? Extended pilot*.

As can be noticed, the majority of the participants have some kind of university education, an issue to take into account for a better understanding of their plurilingualism.

4.3.5 Initial findings

Out of 24 respondens, 18 regard themselves as bilinguals, but the median value of the languages currently spoken is 4 (StDev = 1.57). 17 use English in their daily social and/or work life, thus outlining a plurilingual community whose communicative practices are strongly based on the use of English as a lingua franca.

More specifically, Figure 14 provides a graphical representation of languages spoken in different communicative situations, based on a non-linear mapping of the frequency scale (given that the distance between frequency adverbs in the scale is not always the same) and on response weighted averages.

[13] Linguistics is set apart from arts and humanities because a background in this field may be relevant for the interpretation of results concerning language use.

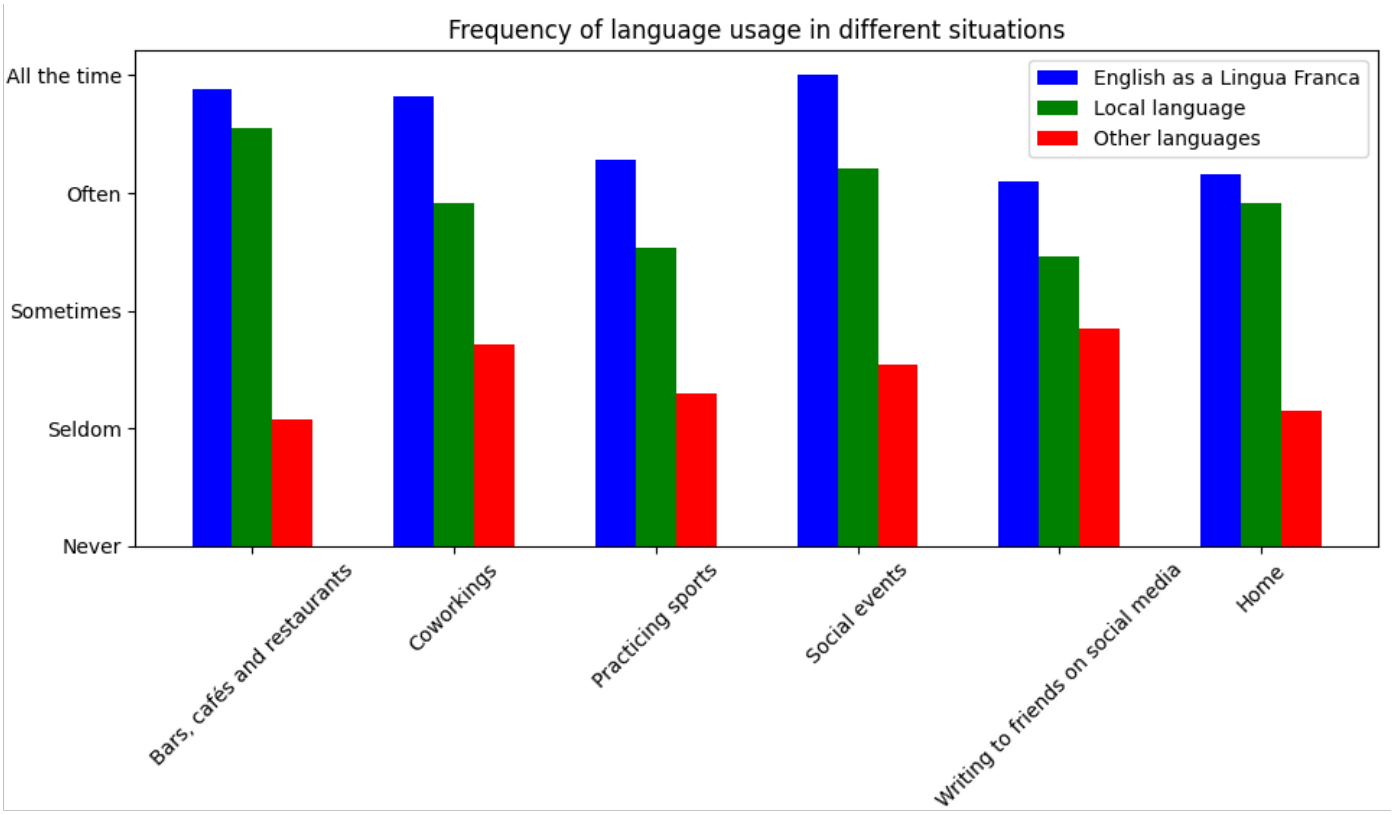


Figure 14. Frequency of language usage in digital nomad everyday life.

As Figure 14 shows, more than half of the respondents to the extended pilot state to frequently use English as a lingua franca at social events and coworkings. However, such an intensive use of English is not always regarded positively, as indicated by Table 5, where the results about attitudes towards English are reported.

Table 5. Digital nomad attitudes towards English.[20]

positive attitude towards speaking more and more English	6
negative attitude towards speaking more and more English	9
preference towards speaking languages other than English more frequently	13
thinking that English is the easiest means of communication	14

[20] The table is particularly based on responses to questions Q50c, Q50d, Q51a, part of the block named 'Linguistic behaviours and attitudes'. The questionnaire's pilot version is not reported in the Appendices but the dataset can be downloaded from AMS Acta (Tedesco 2024b).

The tendency towards using English as a major communicative resource applies also to social media interactions: during netnography, it has been observed to be the most used language on Slack and Whatsapp, and 15 respondents stated to often/always post their contents in English.

For a more precise assessment of the communicative strategies consciously employed by the respondents, the frequency of use of translanguaging, intercomprehension and English as a lingua franca was investigated using the categories shown in Figure 15. The figure presents the average answer for each communicative behaviour.
An overview of Figure 15 thus suggests that English is generally the most used means for communication. However, 11 respondents stated that they use both English and another language often and/or all the time, thus suggesting a plurilingual context of communication, particularly at cafés, coworkings and social events.
As for the respondents speaking languages other than English or the local language(s), results from branching questions show that these are used by participants also to practice their plurilingual skills. Moreover, some of them speak their first language(s) with people of their same nationality when abroad, which can be explained also by the partial overlap between digital nomad and expat communities (the latter being less pluricultural than the nomad ones). Such results are confirmed by interviews and observation in both Madeira and the Canary Islands. For example, Corralejo in Fuerteventura, currently one of the nomads' favourite destinations in the Canary Islands, has a large Italian expat community. Here expats and digital nomads mix, as often happens, and Italian is a language heard as frequently as Spanish and English.
But the category 'other languages' also represents the use of linguae francae other than English, a practice confirmed by ethnographic research in Madeira, where Spanish was often used as a means of communication (especially between Portuguese and Italian people) (Tedesco 2024b).

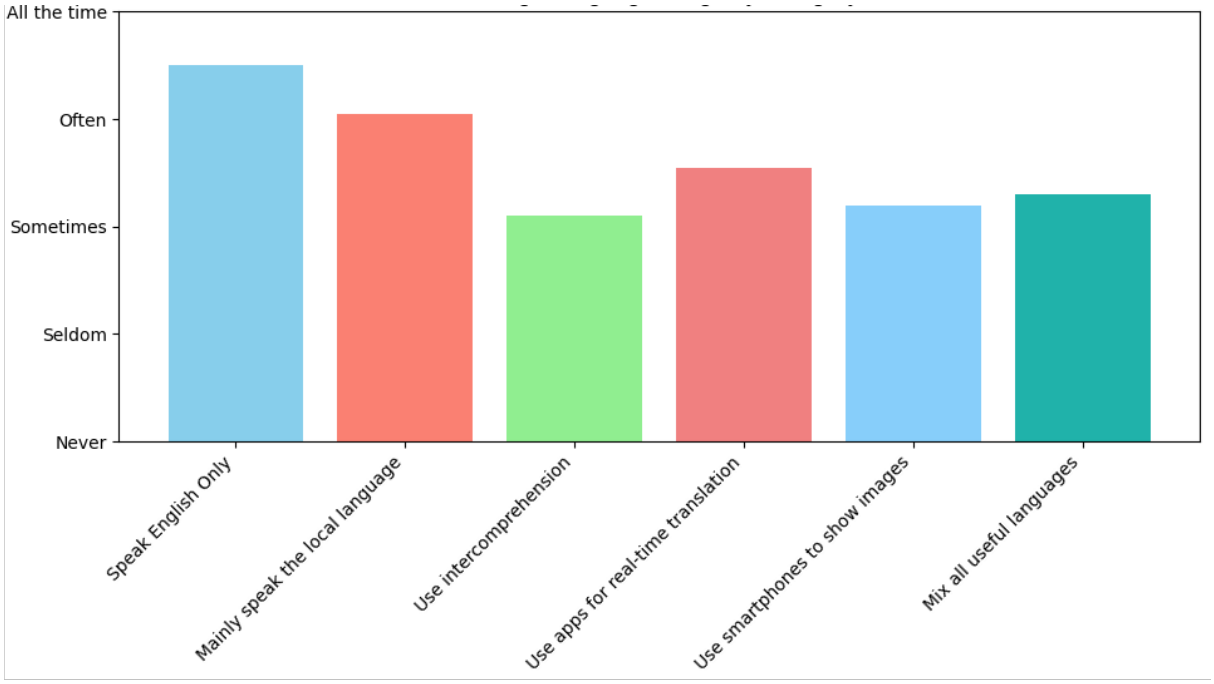


Figure 15. Transcultural communicative strategies employed by digital nomads.

Translanguaging is more used in informal spoken interactions than written or work-related communication. Through an open question, respondents could provide some examples of translanguaging from their own experiences. Answers are shown in Table 6.

Table 6. Examples of translanguaging given by the extended pilot respondents with explanations.

Examples	**Explanations**
‘Portungles, Swinglish’	These are names to call varieties of English enriched by a large presence of Portuguese and Swedish, respectively.
‘Mi sono anmeldata. (Anmeldung: registrazione nella città di residenza in Germania)’	As the respondent explains in parenthesis, ‘anmeldata’ comes from the German word Anmeldung. Here it is mixed with Italian words.
‘Sabar - "Patience" in Bahasa Indonesia’	The respondent provides an example of an Indonesian word also used by speakers who do not know Indonesian.
‘I often use English curses with foreign-language morphology. If I'm trying to say a situation is messed up in French, for example, I'll say "Ça c'est fucké," or switch it to English, as in "You are all jodido" to someone who is drunk.’	The respondent provide examples of phrases where English and French mix.

The survey also revealed information about perceived linguistic needs. Results suggest a certain degree of dissatisfaction with educational programs. English appears to be the only language for which seemingly satisfactory education is provided, which confirms the generally acknowledged tendency towards the global adoption of English as a lingua franca (Smokotin et al. 2014) and as a medium of instruction (Lasagabaster 2022). Even so, further results show that respondents are (at least partially) dissatisfied with their competence in English. Moreover, some respondents have stated that they sometimes feel misunderstood when speaking English, which suggests that English proficiency may not be enough in all the situations but also reveals a possible bias towards hypercritical responses.

Informal language learning (Duranti et al. 2012) was also investigated and obtained generally positive responses. English, Spanish, Portuguese, German and French are the most common languages learnt outside school. This finding aligns with interview data gathered in the Canary Islands. For example, Riccardo, a young digital nomad based in Corralejo, described his communication experience on the island as one of constant informal language learning through exposure to new contexts and peers, often engaging in *Spanglish* or hybrid forms of communication that, for him, reflect adaptation and creativity.

The notion of English as a *lingua franca* was recurrent across interviews. Participants often displayed a metalinguistic awareness of the role of English as a shared resource, while simultaneously engaging in translanguaging practices that integrated other linguistic repertoires, particularly Spanish, Italian, Portuguese, and German. As observed during informal interactions and social events (e.g., *Nomad Grooves* and coworking meetups in Corralejo), code-switching and pragmatic negotiation of meaning were routine features of communication, contributing to a sense of inclusion and mutual intelligibility.

Furthermore, reflections gathered through interviews and online observation indicated that linguistic choice and self-representation were central to community discourse. On platforms such as *Diginomads* (created by Giacomo Bastianelli, an Italian entrepreneur interviewed during the preliminary research phase), language functions as both a practical medium and a branding strategy – a way to construct legitimacy and belonging. English dominates these platforms, yet local languages appear in hybrid forms, suggesting a process of ongoing linguistic negotiation. Riccardo's comment that 'if you are a true digital nomad, you need to know English' encapsulates this ambivalence of English playing both an accessibility and gate-keeping role.

As explained in the following sub-section (4.3.6), the preliminary results hre presented served to develop the final methodological tools, of which results are analysed and discussed in greater details in Chapter 6.

4.3.6 Foregrounding research: General remarks

Foregrounding research uncovered communication as a central concern within digital nomad communities, sometimes even articulated by community members in explicit linguistic terms. Interviews and informal conversations revealed recurring reflections on accents and variaties, intercomprehension practices, code-mixing, and adaptation strategies, indicating language fluidity as both a communicative necessity and a means for symbolic belonging. As one interviewed community member remarked, 'you don't need to speak perfect English, you just need to make yourself understood.'

The observations carried out, particularly within the community based in the northern area of Fuerteventura, supported the view that digital nomads constitute a superdiverse social formation in Blommaert's (2013) sense: individuals of diverse national, linguistic, and professional backgrounds, linked through overlapping mobilities, digital infrastructures and lifestyle rather than shared ethnicity or origin. Communication within such networks is necessarily fluid and adaptive, in sociolinguistic terms they seemed to well exemplify what García and Wei (2014) describe as *translanguaging*, the dynamic mobilisation of semiotic resources across perceived linguistic boundaries.

At the same time, early findings from survey and interviews suggest ideological tensions: while participants valued linguistic diversity and creativity, many still oriented to native-speaker norms and to English as the default lingua franca of work and community life. These insights

later informed the design of survey questions addressing attitudes towards plurilingualism and monolingual ideologies (see 4.4.3).

Summing up, the foregrounding research phase contributed to defining FLO's triangular design (discussed in Section 4.4), establishing the rationale for a multi-layered documentation of linguistic fluidity. Each component emerged as a response to the empirical realities observed during this early exploratory work: ethnography to grasp the context, survey to measure attitudes and perceived behaviours, and corpus creation and analysis to ground interpretation in naturally occurring language data. In line with Stebbins's (2011) notion of exploratory research as 'the systematic search for ideas and insights', this phase provided the intellectual and practical foundations for FLO's data collection. It also reflected the project's commitment to reflexivity, adaptability, and collaboration, core principles in both grounded and participatory methodologies. Ultimately, by foregrounding research practices in real contexts and communities, this phase ensured that the investigation of linguistic fluidity among digital nomads would remain empirically grounded, ethically sound, and socially relevant.

4.4 Mixed methods for linguistic ethnography: From *grounded* theory to data-driven research

4.4.1 General methodological stance

Section 4.4 describes how linguistic ethnography principles were operationalised in the FLO project through a mixed-methods design that integrates qualitative and quantitative approaches within a single, coherent epistemological structure.

FLO's methodology rests on the assumption that complex, multi-scalar social realities require multiple lenses of observation (Creswell 2014; Dörnyei 2016)[21]. Thus, the project adopts a mixed-methods design combining qualitative and quantitative procedures in a single coherent framework. Qualitative methods provide interpretive depth, while quantitative ones ensure systematicity and transparency. Rather than being sequential, the two are mutually informing, each shaping and refining the other through iterative cycles of observation, reflection, and verification. This dialectical movement is inspired by what Blommaert (2010) calls a *sociolinguistics of complexity*, in which empirical unpredictability can only be addressed through close attention to the communicative (micro)context, complemented by analysis of the (macro)context where communication takes place through systematic mapping of larger-scale patterns.

Furthermore, the triangulated design described in the next sections aligns with FLO's commitment to participatory and ethical research. The project was conceived not as a study *on*

[21] The expression *multi-scalar social realities* refers to the idea that social phenomena unfold across several interconnected scales — individual, interactional, local, institutional, and global (see Brenner 2001). A multi-scalar lens highlights how practices are shaped not only by personal choices but also by broader structures such as mobility regimes, digital infrastructures, and local community dynamics.

digital nomads but *with* them. Following the principles of participatory-action research (Lewin 1946; Reason & Bradbury 2008) and community-based knowledge co-creation (Chevalier & Buckles 2013), further discussed in the next sub-section, participants are regarded as collaborators contributing to data collection, interpretation, and dissemination.

The epistemological foundation of FLO's methodology lies in *grounded theory*, originally formulated by Glaser and Strauss (1967) as a method for building theory inductively from empirical observation. Grounded theory rejects the linearity of hypothesis-testing models, proposing instead that categories and explanations emerge from data through iterative cycles of coding, comparison, and theoretical integration. In this thesis, grounded theory is interpreted through what Tusting (2020) calls a 'light grounded' approach. This means that while data collection remains open to emergent insights, the research is also guided by existing relevant studies (accounted for in Chapters 2 and 3). For example, notions such as linguistic fluidity (Blommaert 2014, Prinsloo 2023), mobility capital (Moret 2020), and transcultural communication (Baker 2018) provided conceptual orientation, but their operationalisation was shaped through ongoing field engagement. This flexible inductive logic resonates with participatory and community-based traditions (Heron & Reason 1997; Chevalier & Buckles 2013), which emphasise knowledge co-creation and the epistemic value of experiential insight.

The adoption of a light-grounded perspective thus allowed FLO to evolve iteratively: the design of each methodological component (ethnography, survey, corpus) was informed by findings from the previous one, generating a continuous feedback loop between observation and conceptualisation.

While grounded theory is traditionally qualitative, it can be effectively enhanced by quantitative and computational methods. A special issue of the journal *Frontiers in Sociology* focused on methodological shifts in mobility studies highlights how digital tools can extend the analytical scope of ethnography by enabling the systematic collection, processing, and visualisation of large-scale interactional data (Jasso 2023).

Such integration is coherent with corpus methodologies. Corpus linguistics provides techniques, such as frequency analysis and keyword extraction, that reveal recurrent patterns across vast amounts of text or speech (Baker 2006; McEnery & Hardie 2011). Within this paradigm, corpus-driven approaches (Tognini-Bonelli 2001), which let linguistic categories emerge from the data itself, seem to complement from a quantitative perspective, qualitative grounded approach to research. FLO's data-intensive component, namely the creation of the NomadLingo corpus, illustrates this convergence: transcription, annotation, and statistical processing are grounded in ethnographic knowledge of community practices, ensuring that quantitative outputs remain socially meaningful. Moreover, by linking the interpretive depth of qualitative inquiry with the scalability of computational analysis, the project aligns with recent calls for a digital ethnography that unites interpretive and quantitative paradigms (Varis & Blommaert 2020; Kendon 2022).

Participatory ethnography also builds on grounded theory's constructivist foundations, emphasising experiential knowledge as a legitimate and necessary form of evidence (Heron & Reason 1997). Closely related is netnography (Kozinets 2020), a recent development of ethnography adapted to online environments. As will be explained in Sub-section 4.4.2, netnography uses digital affordances, social-media platforms, messaging channels and collaborative spaces to observe and archive naturally occurring interaction.

FLO research design thus rests on three interdependent pillars:

1. Community-based participatory ethnographic and netnographic observation, capturing the lived experience and communicative practices of digital-nomad communities through on-site participation, interviews, reflective journals, and online engagement;
2. Survey research, producing quantitative evidence on language repertoires, mobility trajectories, and ideologies through the dataset *Where's Your Language At?* (Tedesco 2025b);
3. Data-intensive linguistics, culminating in the *NomadLingo* corpus , a systematically annotated collection of spoken interactions processed with state-of-the-art corpus linguistics and computational tools.

This design ensures triangulation maximising depth, breadth, and replicability (Teddlie & Tashakkori 2009). Some examples of how these data can be interpreted according to specific research questions is provided in Chapter 6. The remainder of Section 4.4 details how this methodological framework was implemented across the project's three components. Sub-section 4.4.2 outlines the ethnographic backbone of the research, including active participation, self-ethnography, interviews, digital ethnography (i.e., netnography) through hashtags and online platforms, and the co-creation of knowledge with community members. Building on the foregrounding research (see 4.3.4), Sub-section 4.4.3 discusses the design and administration of the survey *Where's Your Language At?*. Finally, Sub-section 4.4.4 provides general methodological information on corpus linguistics and spoken corpora, while a more detailed discussion on the specific methods and techniques that were employed for the creation of NomadLingo is presented in Chapter 5.

4.4.2 Community-based participatory ethnography and netnography

Ethnography has long been recognised as one of the most flexible and interpretively rich approaches for investigating social life, communication, and culture. In its classical form, ethnography was developed within anthropology as an immersive method aimed at producing thick, contextualised accounts of human practices and beliefs (Hyme 1974). The traditional ethnographer's task was to go to the field and to observe, describe, and interpret the everyday lives of a relatively bounded community, engaging directly with participants in order to understand their worldviews. Over time, however, social life has become increasingly fluid, mobile, and digitally mediated, requiring the discipline to adapt both conceptually and methodologically. As Pink et al. (2016) observe, 'ethnography is not a fixed technique but a sensibility', one that must evolve alongside the forms of human life it seeks to study.

In contemporary sociolinguistics and communication studies, ethnography has thus expanded beyond the study of geographically delimited groups to embrace mobile, translocal, and networked forms of sociality (Marcus 1995; Falzon 2009; Blommaert 2010). This shift has given rise to new types of ethnography, such as multi-sited ethnography, mobile ethnography, and netnography, reflecting the need to capture communicative practices as they move across contexts, media, and communities. As Blommaert and Rampton (2011) argue, linguistic ethnography, in particular, must now account for the complex sociolinguistics of mobility that defines late modern societies, where meaning is constantly renegotiated in migration and continuous connectivity.

As introduced in Chapter 1, linguistic ethnography, as described by Rampton et al. (2015), represents a hybrid tradition situated at the intersection of ethnography, linguistics, and discourse analysis. This approach assumes that 'linguistic form, communicative practice, and social process are inseparable' (Rampton et al. 2015). Thus, rather than imposing pre-defined analytical categories, linguistic ethnography invites researchers to explore how meaning emerges through situated interaction, combining micro-analytic attention to language with a broader ethnographic sensitivity to context, identity, social and power dynamics.

The ethnographic dimension of FLO was designed as community-based and participatory, meaning that the writer did not study digital nomads from an external position but *with* them, as a partial insider who lived, worked, and interacted within the same spaces of the communities under study. This aligns with the epistemological stance of participatory-action research (Lewin 1946; Reason & Bradbury 2008), which views participants as co-producers of knowledge rather than passive subjects. The digital nomad communities of Madeira and the Canary Islands, characterised by a strong collaborative nature, sharing, and creative entrepreneurship, provided an especially fertile environment for this approach. Indeed, as Jenkins (2009) observed, digital nomad culture itself is built around participatory logics of co-creation and mutual learning.

Drawing on these principles, ethnographic fieldwork was conceived not only as a method for data collection but as a process of knowledge exchange that could benefit both scientific advancement and the communities involved. This vision also reflects the European Commission's emphasis on research valorisation – that is, the transformation of scientific outcomes into socially and economically valuable results, including open-access resources, training, and potential enterpreneurship spin-offs (see Chapter 7). In the FLO project, valorisation was achieved through workshops, collaborations with local institutions, and the creation of open datasets that support applied research in plurilingual education and transcultural communication.

From a methodological standpoint, the ethnographic design drew on several theoretical traditions that informed how participation, observation, and data collection were implemented. Classical ethnography (Hymes 1974; Spradley 1980) emphasised long-term immersion and thick description, while linguistic ethnography introduced analytic precision to the study of communicative events (Rampton et al. 2015; Tusting 2020). In turn, digital ethnography (Pink

et al. 2016; Hjorth et al. 2017) expanded the field to online environments, highlighting how digital traces, platforms, and media contribute to meaning-making. Finally, netnography (Kozinets 2015; Costello et al. 2017) provided tools for studying communities that exist both *on* and *through* the Internet – a key methodological reference in this thesis, given the hybrid on-/offline ecology of digital nomadism.

Ethnography in FLO therefore operated across three interrelated scales:

1. The experiential scale, corresponding to the researcher's participation in community life both online and on the field (e.g., co-living, attending events, contributing to online forums);
2. The discursive scale, encompassing interactions in both physical and virtual spaces;
3. The reflexive scale, comprising field diaries, self-ethnographic notes, and autoethnographic reflections documenting the researcher's evolving positionality.

Thus, ethnography became not merely an observational technique but a mode of engagement – a way of inhabiting the research context and making sense of it through the lived experience. As Clifford and Marcus (1986) already argued, contemporary ethnography must be understood as a form of writing culture: an act of representation that requires self-awareness, negotiation, and ethical care.

The specific methodological articulation of this approach is detailed in the following paragraphs. First, I will focus on self-ethnography and my position within the community under study. Then, discoursive engagement with community members and activities will be presented through the discussion of interviews and the daily journal, where notes focus on linguistic variation, language attitudes and transcultural consciousness. This discussion is followed by the description of netnographic data collected on digital platforms such as Slack, WhatsApp, and Instagram, exploring how hashtags, community channels, and online networks mediate the construction of belonging. Finally, a reflection on community-based knowledge co-creation outlines the role of participants through the research process.

A note on positionality and situated ethics

Reflexivity and positionality constitute central tenets of contemporary ethnography. As argued by Hammersley and Atkinson (2019:15), 'ethnographic knowledge is always situated: it arises from the particular social position of the researcher, from specific interactions, and from the interpretive frames through which experience is made sense of.' In the case of FLO, the researcher's position cannot be easily reduced to the traditional dichotomy of *insider* versus *outsider*. Having lived and worked as a digital nomad already before the start of the project, I entered the field as a member of the community that I aimed to study. This dual status, being simultaneously participant and observer, enabled privileged access to the community's infrastructures, networks, and affective dynamics, yet also required heightened awareness of interpretive bias.

Such hybrid positioning aligns with what Dwyer and Buckle (2009) describe as the 'space between': an intermediate stance where the researcher is neither fully external nor fully immersed, but moves reflexively across boundaries. In digital nomad settings, this fluidity was particularly visible. The field was not a bounded village or institution but a constellation of transient networks – Slack channels, co-living spaces, surf hostels, yoga studios, and coworikings – each with its own rhythms of arrival and departure. My life as a researcher of digital nomadism has been characterised by continuous travels for the last three years. Interactions during community breakfasts, surfing sessions, or co-working events were simultaneously social encounters and ethnographic observations. When discussing topics such as work–life balance or linguistic adaptation, interlocutors often shifted between addressing me as a peer and as a researcher. This emotionally demanding oscillation required constant negotiation of identity and transparency about the research aims, which were communicated both verbally and through accessible online materials such as the project website (*thenomadlinguist.eu*) and Instagram page (*@the.nomad.linguist*). In this sense, ethical engagement became an ongoing dialogue with myself as well as with the participants in the research, rather than a one-time procedural requirement.

Ethnographic ethics, as Holmes, Fay, Andrews and Attia (2022) emphasise, are never fully codifiable. They unfold through situational judgments that depend on local norms, relational dynamics, and the researcher's moral awareness. FLO adopted what can be defined as a situated ethics approach, aligning with the principles of the European Code of Conduct for Research Integrity (ALLEA 2017) but extending them through reflexive practice. Consent was understood as a process rather than a static signature: participants were informed orally and in writing of their rights, the use of data, and their freedom to withdraw at any time, yet these discussions were revisited as relationships evolved. Moreover, the informal and mobile nature of digital nomad communities demanded a flexible interpretation of traditional consent models. Conversations and interactions often took place in spontaneous, translingual settings characterised by a high level of informality – beaches, terraces, or during community dinners. In these situations, recording was only possible when all the people that were there explicitly agreed. For this reason, the procedures of data collection for the NomadLingo corpus, which will be addressed in greater details in Chapter 5, were particularly challenging. In interviews and everyday conversations, verbal consent was often preferred, followed by immediate anonymisation and reflective fieldnotes rather than verbatim transcription when privacy could not be fully guaranteed. Such practices respond to what Pink et al. (2016) define as ethnographic ethics in practice: the capacity to make context-aware decisions that prioritise participants' agency and dignity over data exhaustiveness.

An additional challenge concerned the visibility of the researcher's role within online environments. In Slack and WhatsApp groups, participation involved reading messages, occasionally contributing to discussions, and analysing patterns of communication. To ensure transparency, I introduced myself in each group as a linguist conducting a project on language and mobility, clarifying that any data used for research purposes would be anonymised and aggregated. This disclosure fostered trust and, in some cases, prompted meta-linguistic reflections by participants themselves, discussions about accents, code-switching, or 'how

English is used in the nomad world' – using the words of a participant in the study – which later informed analytical categories applied to the interpretation of survey as well as corpus data.

In other words, the combination of mobility, plurilingualism, and digital mediation brings specific ethical tensions that go beyond traditional fieldwork dilemmas. As Markham (2018) notes, in networked ethnography the boundaries between public and private, participant and audience, are constantly negotiated. A post on Instagram or Slack may be intended for a restricted circle yet accessible to hundreds of viewers; an online chat may blend personal and professional communication. These conditions make binary distinctions such as consented vs. non-consented data inadequate.

Therefore, as already mentioned in the discussion of the DMP, ethical decision-making in the FLO project followed a precautionary principle: any digital content was included in a closed dataset and only if it was publicly available. Private conversations were used solely as background knowledge or were paraphrased with participants' authorisation. Whenever screenshots or visual materials were considered for presentation, written permission was obtained from all identifiable individuals. This approach respects both the General Data Protection Regulation (EU 2016/679) and the Association of Internet Researchers (AoIR) guidelines (Franzke et al. 2020), which emphasise contextual integrity and relational accountability.

Plurilingualism itself introduced further interpretive challenges. Many participants alternated between English, Spanish, Portuguese, and other languages within the same interaction. Following Blommaert (2010) and García & Wei (2014), the project adopted the notion of *translanguaging spaces* – contexts where linguistic boundaries are fluid – not only in the analysis of corpus data but also during ethongraphy. This means that all semiotic resources were treated as equally valid analytical materials.

As previously mentioned, fieldwork among mobile, self-reflective professionals also entailed a substantial degree of emotional labour. I was often confronted with the same dilemmas expressed by participants – precarious work, constant relocation, and the pursuit of belonging in transient networks. Rather than suppressing these resonances, I embraced them as part of the reflexive process. As Ellis and Bochner (2006) argue, autoethnographic honesty lies in acknowledging vulnerability as epistemic strength: emotions are not distractions but data that reveal the affective texture of social life.

To manage this complexity, a reflexive diary was maintained throughout fieldwork, combining descriptive entries with analytic memos. Notes captured not only what participants said or did but how interactions felt – the atmosphere of co-working spaces, the pace and attitudes in translingual conversations, the moments of misunderstanding or connection. These records later informed the analysis of communicative practices, besides being fundamental in shaping ethical awareness.

Importantly, reflexivity also extended to temporal positionality. The researcher's role evolved over time: from an active organiser of events in Madeira and Porto Santo to a more observant presence in the Canary Islands, where previous familiarity had already established relations. This shift mirrored the gradual transition from participant to analyst, from *doing with* to *reflecting on*. Yet, even during analysis, the relational bonds created during fieldwork remained active, requiring careful boundaries between professional collaboration and personal friendship. Conversely, in some cases, the emotional closeness to participants allowed for a better understanding of their cultural and linguistic backgrounds and on the discourse dynamics put into action within certain conversational environments. These insights greatly helped with data processing and analysis.

Ethnography inevitably involves asymmetries of power. The ability to represent others' experiences in academic writing confers authority that participants do not equally share. FLO sought to mitigate this by enhancing peer discussion and incorporating participant voices wherever possible – through blog posts on *thenomadlinguist.eu*, direct quotations with consent, and feedback sessions where participants reviewed interpretations. This participatory validation process contributes to edge what Clifford and Marcus (1986) defined as the crisis of representation: the recognition that ethnographic texts are always partial, positioned, and dialogic.

Ultimately, positionality and ethics in FLO are inseparable from its participatory ethos. The project's methodological foundation – community-based ethnography – presupposes that ethical research is not achieved through compliance but through relationship. By living, working, and communicating with participants, the researcher enacted co-operative inquiry (Heron & Reason 1997): knowledge generated through shared experience and mutual reflection. In practical terms, this meant engaging with participants as co-thinkers, co-organisers, and sometimes co-authors, recognising their expertise as insiders in the digital nomad world.

This collaborative ethics continues beyond the end of fieldwork. Through open data practices (see Section 4.2.2) and dissemination via public platforms, FLO aims to sustain an exchange that contributes to both academic and community learning (as discussed in Chapter 7). Ethical reflexivity, therefore, is not confined to consent forms or debriefings; it extends to the entire life cycle of research, from design to publication, from data collection to societal impact – and as such as been accounted for at different points in this thesis to mirror the recurring reflections that characterised the whole research process.

Active participation & self-ethnography

Traditionally, ethnography was associated with long-term immersion in a delimited field, often geographically bounded and culturally cohesive. The researcher went to the field, observed and participated in local activities, and then returned home to write. However, the increasing digitisation and mobility of social life have led to new forms of continuous and multi-sited ethnography (Marcus 1995; Falzon 2009), in which researchers follow people, artefacts, and

discourses across online and offline environments. Today's ethnographers often combine short-term immersions in multiple sites with long-term thematic engagement and sustained familiarity with their research communities (Hine 2015).

Within this shift, the distinction between observer and participant has progressively blurred. Participatory and action-oriented approaches (Lewin 1946; Reason & Bradbury 2008) place emphasis on collaboration, reflexivity, and the co-production of knowledge. The active participation characterising FLO's ethnographic approach is consistent with community-based participatory research (CBPR) (Chevalier & Buckles 2013; Denzin & Lincoln 2018). CBPR emphasises shared ownership of research processes, mutual learning, and the production of results that are meaningful for the communities themselves. Hence, what follows is a report of the ethnographic activities and self-ethnographic reflections carried out within the communities of Madeira and Canary Islands during the FLO project, including the presentation of a few salient examples.

It should be noticed that the lifestyle and mobility patterns of digital nomads require an adapted notion of *field*, because the communities observed are neither static nor circumscribed. Preliminary findings led to the planning of 'multi-sited' observations and participation (Marcus 1995). Visiting a community appeared not be enough; I decided to engage in a mobility trajectory similar to those actually followed by the nomads. Continuous travel has proven to be a necessary condition to understand the digital nomad culture, honestly engage with the communities and really capture the diversity characterising their linguistic and cultural landscapes. This process enabled a form of *continuous ethnography*, in Burrell's terms an ongoing, iterative participation in a community of practice that transcends any single phase of data collection (2009). In Figure 16 a timeline of the most significant movements I made from the development of the Ph.D. research project to its final moments is shown.

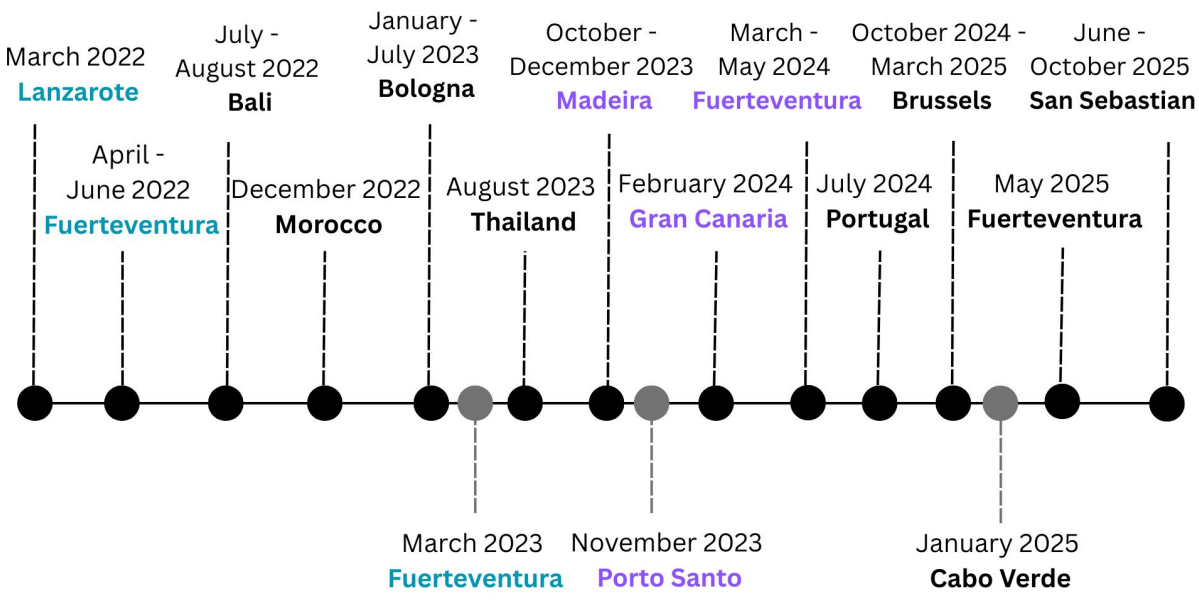


Figure 16. Researcher's mobility trajectory from March 2022 to November 2025.

Through a nomadic lifestyle, constant engagement with online communities, and continuous presence in co-workings, coworking cafés and community events at all the digital nomad hotspots to which I travelled (listed in Figure 16; all highlighted as top digital nomad destinations in Flatio Report 2024 and/or nomads.com website), I embraced the digital nomad philosophy all round.

Indeed, self-ethnography became an essential methodological tool of the project. Sometimes also referred to as autoethnography, it is a form of inquiry in which the researcher's personal experience becomes both the lens and the data for understanding broader cultural phenomena (Ellis et al. 2011). As Anderson (2006) notes, autoethnography requires maintaining an analytic balance between subjectivity and systematic observation: the self is not merely narrated but interrogated as part of a collective experience. In the FLO project, self-ethnography was particularly relevant because I shared lifestyle, values, and routines with the participants. Having lived as a digital nomad before and during the research, I was both insider and observer, a dual positioning that enriched interpretation while demanding rigorous reflexivity.

Self-ethnography also shaped the project's temporal structure. As summarised in Figure 16, rather than a single, delimited fieldwork period, data collection evolved through cycles of participation and reflection. In this process, the ethnographic journal served as both memory archive and analytic space, where impressions, conversations, and small linguistic episodes were documented immediately after they occurred. Over time, these micro-observations accumulated into a composite portrait of how language, mobility, and sociality interlace in the everyday life of digital nomads.

The colors in Figure 16 represent sites for preliminary research, in turquoise, and final research, in purple. Grey points are trips which happened within longer stays. The labels used in the figure refer to geographical regions, countries, or cities; however, longer stays in Madeira and Canary Islands included movement around the regions. The time spent in Madeira was split into a stay in Funchal, the major city, home to one of the largest digital nomad communities of the island, and Ponta do Sol, where the digital nomad village project is based (see Chapter 2). In Funchal I attended community events, I worked at Sangha coworking where I also participated in workshops (Figure 17), I joined the nomad dancing, surfing and hiking groups (Figure 18) and I shared my knowledge of yoga by leading free yoga practices in a coliving in the city centre (Figure 19).

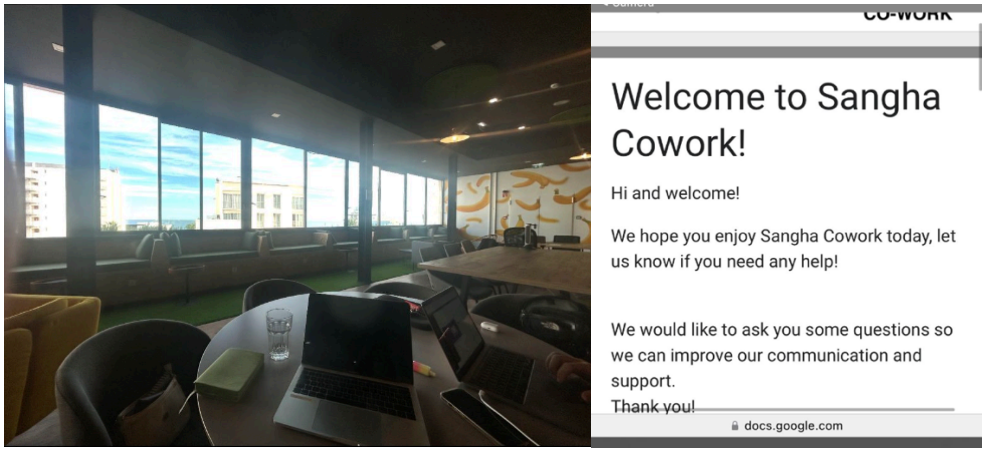


Figure 17. Sangha coworking, and its welcome page with information after subscription.

Figure 18. A hiking activity organised by Madeira Digital Nomad Community.

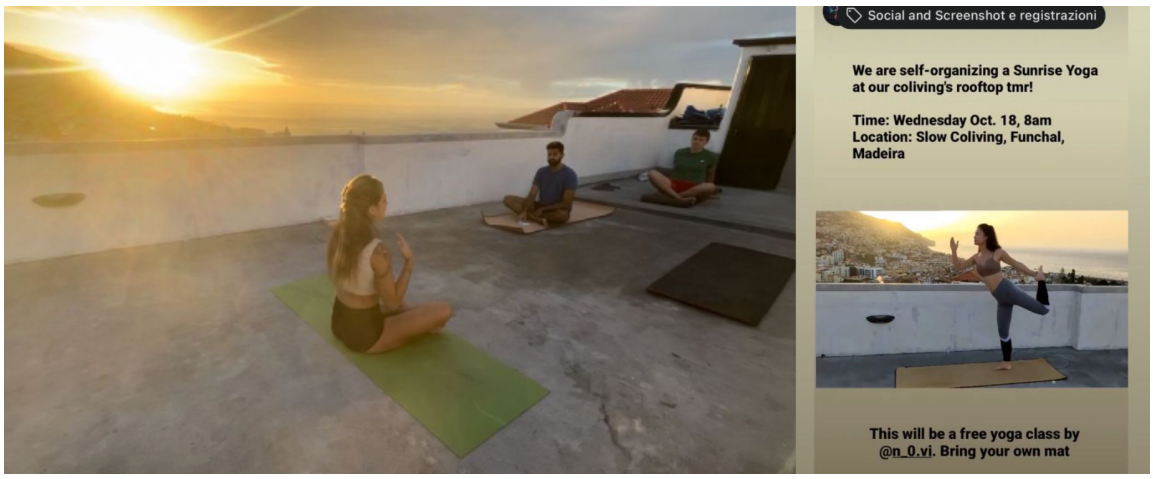


Figure 19. The researcher leading a yoga class on the rooftop terrace of a coliving in Funchal, and the flyer posted on Instagram by one community member to promote the activity.

Then, in Ponta do Sol I lived in a shared house with three more nomads very close to the small village centre. There I attended community events daily. Unlike in Funchal, where community events are mainly hosted by cafés and restaurants, here community life develops around the coworking, which was also the place where the project Digital Nomad Village and its business spin-off NomadX were born. Figure 20a and 20b show some details of the location.

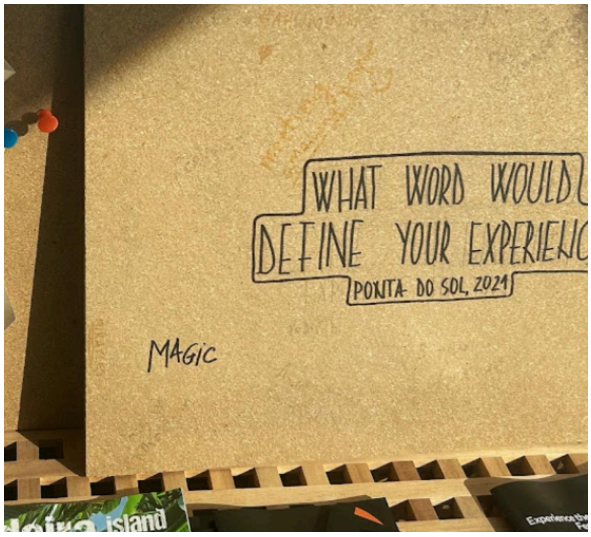


Figure 20a. A board where people could write their impressions about living within the Madeiran digital nomad community.

Figure 20b. A couple of nomads sharing biscuits and Dutch cultural tradition with the other members of the community.

During my three-months stay in Madeira, I also joined a nomad retreat in Porto Santo (Figure 21), which was also site for natural language data collection (see Chapter 5). The retreat lasted four days; the coordinator (Ganzalo Hall, from NomadX) was helped by one of Porto Santo's community managers, Sofia. We stayed at a beachfront hostel and worked at the coworking in the village centre, opened thanks to the efforts of the Nomad Village project and Start Up Madeira. The activities there included hikes, lunches and dinners, kayaking, diving, and a workshop on how to make Poncha (a traditional Madeiran drink).

Figure 21. A group of nomads during the nomad retreat organised by NomadX in Porto Santo.

Seemingly, the time spent in Las Palmas de Gran Canaria also included many trips to Maspalomas to participate in community events in that area. However, the majority of the events were based in Las Palmas de Gran Canaria. There, I changed three accommodations: first, I stayed three weeks in a coliving named Kite&Surf Nomad House. Figures 22a & 22b show how activities were organised and sponsored by coliving managers both on place and online.

Figure 22a. The board at the entrance of Kite&Surf Nomad House in Las Palmas de Gran Canaria.

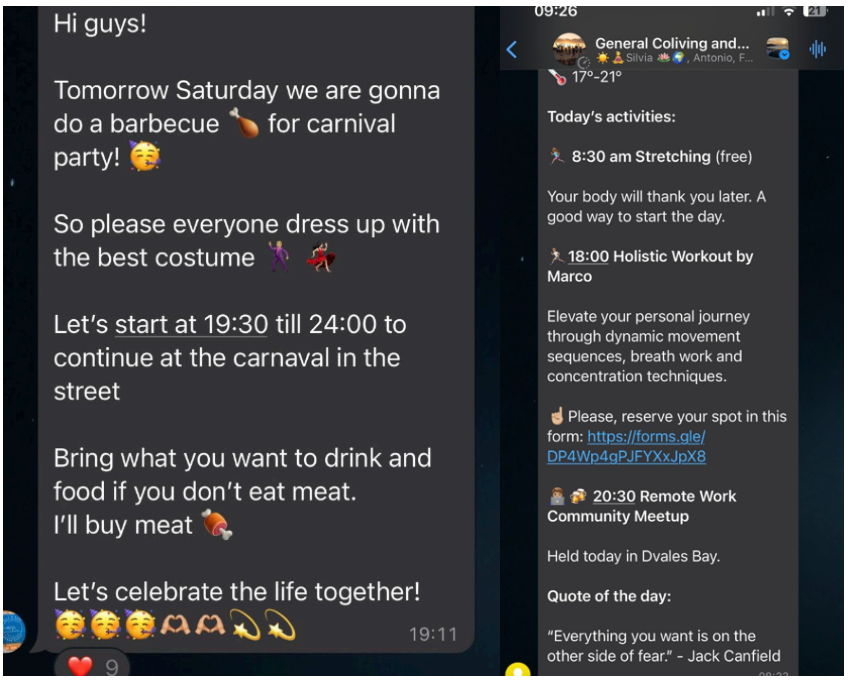


Figure 22b. Messages on the Whatsapp group of Kite&Surf Nomad House.

The community at the Nomad House was very active and pluricultural. The screenshot in Figure 23 indeed shows a translanguaging message on the Whatsapp group where one community member was thanking another (from Central America) for having prepared tacos for the whole group.

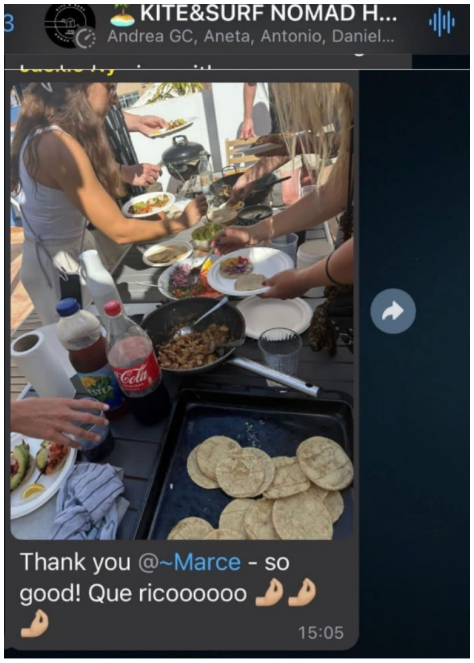


Figure 23. A thanking message on the Kite&Surf Nomad House Whatsapp Group with evidence of translanguaging.

Besides the events organised by the coliving managers, the community, which in Las Palmas is named 'Live it up Las Palmas', promoted a variety of events through the Slack Channel, including language exchange happy hours (as shown in Figure 24a, 24b and 24c).

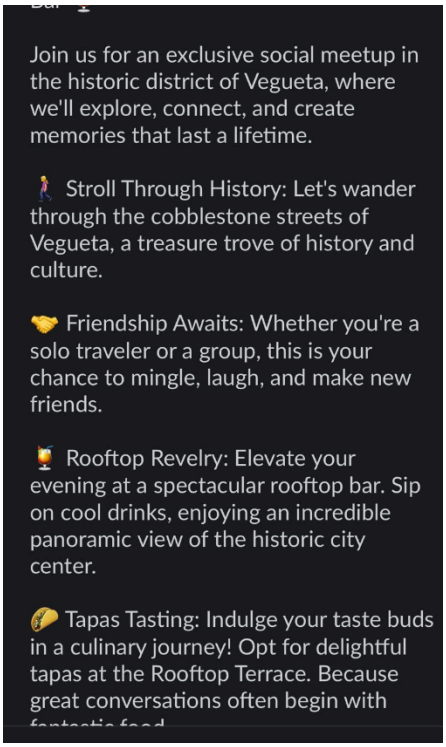


Figure 24a. Screenshot of a message on the Slack Channel for the promotion of a social event at Las Palmas de Gran Canaria.

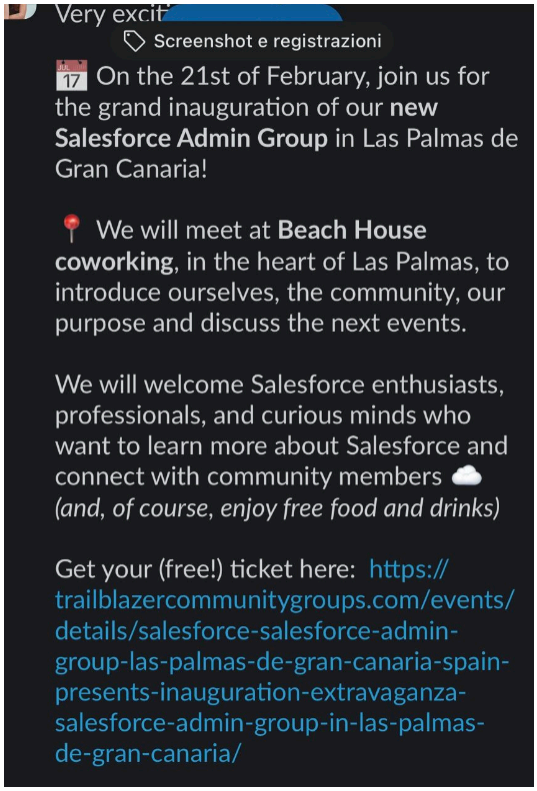


Figure 24b. A workshop promoted through the Gran Canaria community Slack channel.

Figure 24c. Flyer of the event 'Bilingual Aperitivo' in Las Palmas de Gran Canaria.

The last ten days in Las Palmas de Gran Canaria were spent in two more coliving accommodations. While in Las Palmas I attended a coworking named Talleres Palermo (Figure 25) and many coworking cafés (Figure 26).

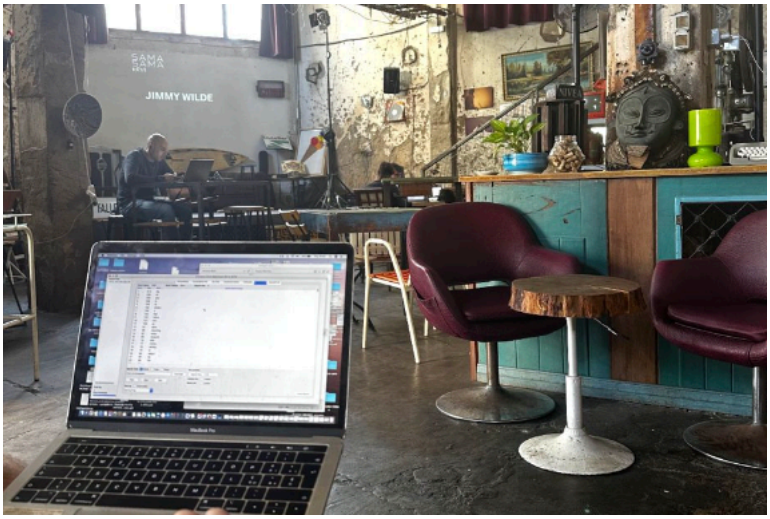


Figure 25. Talleres Palermo, a coworking in Las Palmas de Gran Canaria.

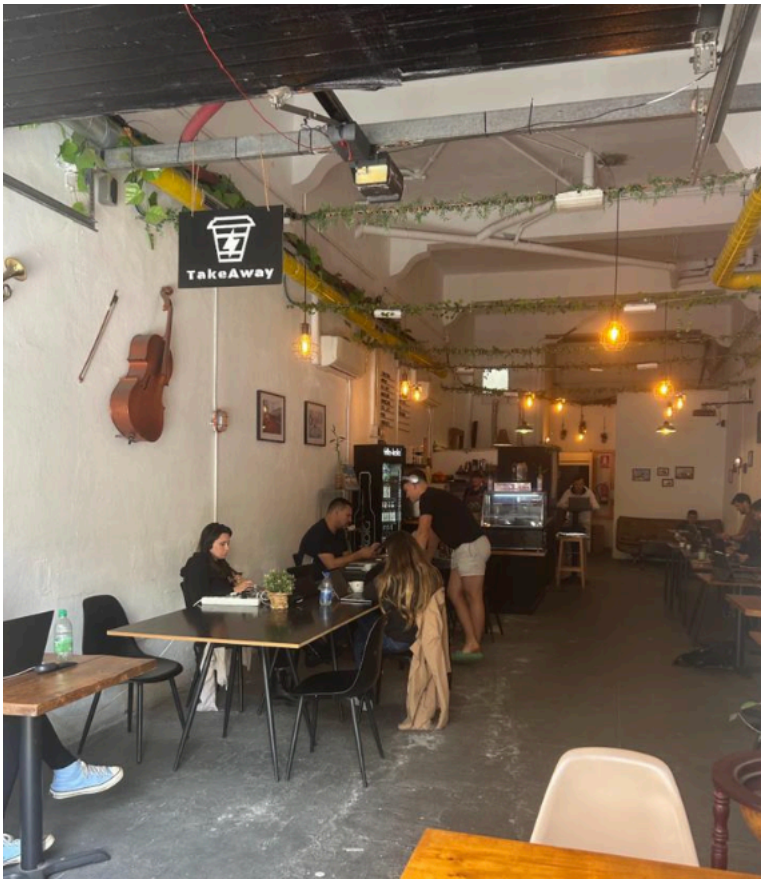


Figure 26. A coworking café in Las Palmas de Gran Canaria.

In March, I then flew to Fuerteventura. In Fuerteventura, I took part in community life and social events around Corralejo, Los Lajares and El Cotillo, the three main digital nomad hotspots on the island. A picture from the welcoming dinner organised by one of my friends on the island is shown in Figure 27. Nomads of eight different nationalities participated.

Figure 27. Welcome dinner party in Corralejo, Fuerteventura.

In Fuerteventura I mainly participated in the life of the coworkings Surfing Colors, in Corralejo and Surfescape, in Los Lajares.

As can be noticed in Figure 28, seemingly to the one in Las Palmas, Surfing Colors coliving & coworking has its own Whatsapp group where events and activities are organised, not only by community managers but also by members proposing to go out for dinner, go surfing or for a hike.

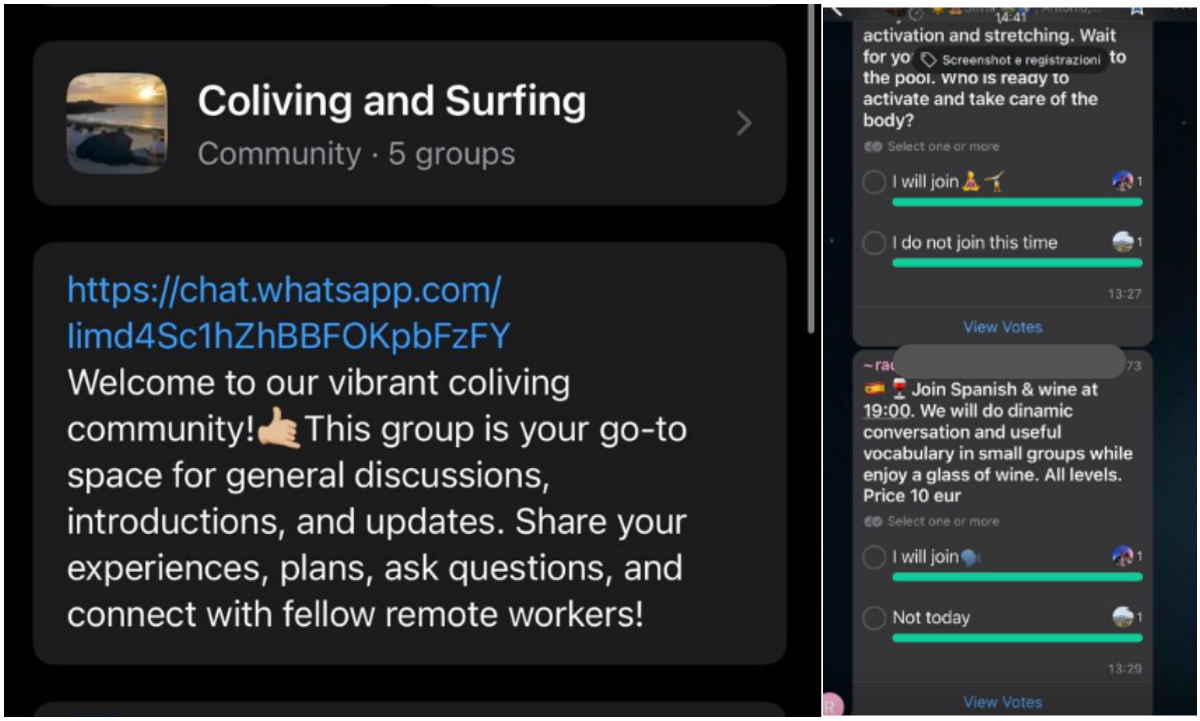


Figure 28. Screenshots from the Whatsapp group of the coliving and coworking Surfing Colors in Corralejo, Fuerteventura.

The first screenshot is the group description featuring typical keywords of digital nomad online discourse. In the second screenshot, once again the activity proposed is a Spanish class, which points to nomads' curiosity towards *local* languages (as will be discussed in Chapter 6). It should be noted that this is not the only Whatsapp group for the Fuerteventura community in which I took part, as will be discussed in the sub-section on netnography.

Besides activities strictly linked to coworking life, I also joined spontaneous meetings and activities with other community members like beach dinners and hikings (Figure 29). These were fertile settings for observing people interacting in very informal contexts, hearing about their stories and trajectories and interviewing them.

Figure 29. A dinner on the beach after a hike with the digital nomad community in Fuerteventura.

A summary of all the locations explored during foregrounding and main research phases is provided in Figure 30.

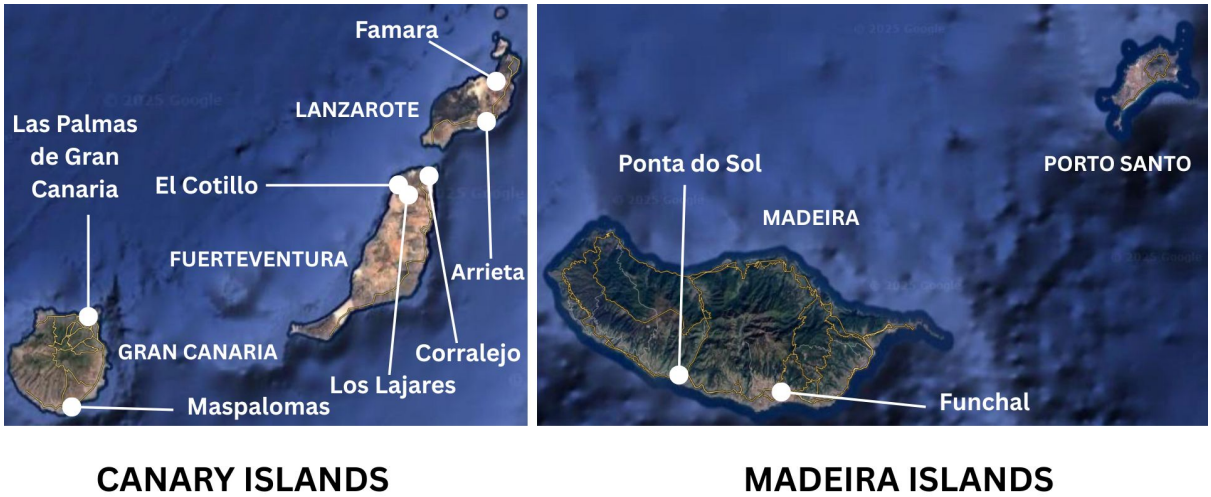


Figure 30. Field ethnography locations.

Through the long-term immersion in the communities of Madeira, Gran Canaria and Fuerteventura, and the short nomad retreat in Porto Santo I was exposed to daily translingual practices, in which English, Spanish, Portuguese, and Italian alternated fluidly depending on participants, context, and affective dynamics. Such recurrent exposure provided both *emic* (insider) and *etic* (analytic) insights into how nomads employ their plurilingual repertoires.

The multisite design of FLO required maintaining ethnographic continuity across different locations. In practice, this meant transferring certain research routines, such as daily journaling, language observation, and informal interviewing, from one site to another while adapting them to local conditions. For example, in Madeira my integration into community life was facilitated by institutional support from Startup Madeira and the University of Madeira; in the Canary Islands, access relied more on peer networks and informal gatherings. Across all sites, the ethnographic journal served as the 'portable archive' of the project, ensuring comparability and coherence. Its structure also allowed temporal layering: revisiting earlier notes generated new insights, highlighting how understanding evolved alongside the researcher's participation.

Moreover, adopting a self-ethnographic perspective also entails recognising that data are co-produced through interaction. Every conversation, interview, or shared activity is simultaneously a social event and a research moment. This awareness informed the ethics of engagement throughout FLO. Participants were not anonymous informants but collaborators who contributed to shaping research questions and interpreting findings. For instance, after informal interviews, summaries were often shared with the interlocutors for feedback, ensuring accuracy and fostering transparency.

Interviews and daily journal

In general, interviews and ethnographic journals constitute two of the most valuable tools within the methodological framework of linguistic ethnography, as they offer complementary insights into participants' lived experiences and the researcher's own interpretive process. In

FLO, these instruments were not treated as separate techniques but as intertwined forms of reflexive documentation, supporting what Rampton et al. (2015) describe as the ethnographer's 'movement between participation and analysis'. Both methods were essential for tracing the social and linguistic complexity of digital nomad communities and for grounding the mixed-methods design in lived communicative realities.

Both interviews and fieldnotes were processed through iterative cycles of analysis that mirrored grounded theory principles (Glaser & Strauss 1967; Charmaz 2014). Initial open coding was followed by axial coding, during which categories were related to emerging theoretical concepts such as linguistic fluidity, mobility capital, and transcultural communication (Charmaz 2006). No software was used to analyse these types of data so interpretation remained anchored in ethnographic immersion rather than computational abstraction.

Fieldnotes often alternated between first-person descriptions of events and analytic commentary, reflecting the dialogic movement between lived experience and theoretical abstraction. Over time, these ethnographic notes evolved into public-facing blog articles, where the reflexive stance of the researcher intertwined with narrative accounts of community life. For instance, in an article derived from reflections on the experience of the nomad retreat in Porto Santo (November 2023), the fieldnote becomes a situated reflection on community belonging and transcultural encounters:

> It's eight a.m., I came to the beach for some morning stretching. The dunes hide the road so now that I'm writing it's only me, the sand and the sea. […] I can't help but think about the people I've had dinner with last night. Total strangers until two days ago, we come from different angles of the world, we have lived different experiences, our eyes do not see things in the same way. And yet somehow we manage to communicate.

With a more practical tone, in the blog article resulting from the interview with Fabi, the creator of the network Remote Work Community, the analytic attention to interactional practices emerges from the recollection of personal experiences:

> In the Nomad Cruise, for example, the organisers were always asking to speak English. They were afraid that we would start speaking German preventing engagement of people from other countries". For Fabi the English language embodies both a practical requirement and a symbol of inclusivity. "But at the same time it is not nice if you are in a place, let's say here, for six months and do not speak a word of Spanish. You are like an alien" – we laugh, and he tells me about the many language course flyers he has seen in the language channels of the SLACK communities.

This intertwining of observation, narration, and interpretation is characteristic of linguistic-ethnographic work, where fieldnotes operate as both empirical records and sites of self-reflection (Tusting 2020). In this sense, writing about the community required continuous attention to my own language practices, translanguaging repertoire, cultural and linguistic trajectory.

Field notes were coded into a daily ethnographic journal, that was maintained throughout all fieldwork phases (2023–2024). The journal had multiple functions: (1) a descriptive record of events, (2) a reflexive tool for tracking my own role in the community (as above discussed), and (3) an analytic bridge between field experience and theoretical insight. As Tusting (2020) notes, writing fieldnotes is an act of analysis in itself, 'a process through which ethnographers construct the data they later interpret.'

Entries were written in a hybrid format combining narrative observation and analytic commentary. Each day was typically divided into three parts:

- descriptive notes**,** capturing places, interactions, and events;
- analytic notes, reflecting on linguistic phenomena observed;
- reflexive notes**,** documenting the researcher's emotions, assumptions, and positionality in relation to participants.

An excerpt from March 2024 illustrates this interplay:

> Morning at the coworking in Las Palmas: group conversation about projects. English dominant but some Italian and Portuguese jokes exchanged. One person apologised for her accent, another replied, 'Don't worry, we all have accents here.' → Striking example of linguistic egalitarianism as ideology, though still oriented toward English as default lingua franca.

To pay attention to communicative details, many journal entries included micro-analyses of speech events, often written immediately after interactions. These short analytical notes captured phenomena such as overlapping talk, discourse markers, self-repair, or humor-based alignment, elements that later informed the corpus annotation categories (discussed in Section 5.5).

Particular attention was paid to moments of breakdown or repair, where mutual understanding was negotiated. For instance, one recurring pattern was the use of reformulation strategies ('What I mean is...') or metalinguistic checks ('You know what I mean, right?'), which indicated high levels of communicative awareness and cooperative alignment, and also unveil the role of trust and emotions between participants and researcher. These phenomena were catalogued under the tag interactional negotiation, following Gumperz's (1982) framework on conversational inference.

Through systematic comparison across field sites (Madeira, Porto Santo, Gran Canaria, Fuerteventura), the diary revealed recurring communicative features:

- reliance on English as lingua franca with creative translingual insertions;
- alternation between more formal and highly informal registers within the same event;
- the emergence of shared idioms specific to digital nomad culture (e.g., 'slownomad,' 'rebase,' 'workation');
- and a recurring awareness of linguistic repertoire as an element of identity definition, more than nationality or cultural belonging.

These findings directly informed the later design of the *NomadLingo* corpus, which aimed to capture and quantify similar linguistic patterns across a systematised dataset.

Given the mobility and diversity of digital nomad communities, designing interviews required balancing structure and spontaneity. The first stage of fieldwork, particularly in Fuerteventura and Madeira, revealed that traditional semi-structured interviews were often inadequate for capturing the fluid, informal nature of daily communication with participants in the community. Moreover, many community members worked asynchronously across time zones, changed accommodations weekly, and preferred to converse in relaxed environments rather than formal settings. Consequently, FLO adopted 'opportunistic interviewing' (Hammersley & Atkinson 2019): flexible, situational encounters that arise naturally in the flow of everyday participation. Conversations were conducted in cafés, coworking spaces (as the one shown in Figure 31), or beaches, often following group activities such as yoga, surfing, or networking events. The tone was conversational rather than interrogative, reflecting the ethnography of communication perspective (Hymes 1974; Saville-Troike 2003), where language use is observed as both a topic and a means of interaction. While this flexibility precluded complete standardisation, it allowed for greater ecological validity. As Kvale and Brinkmann (2009) emphasise, interviews in qualitative inquiry are 'social productions of knowledge through interaction.' Each exchange became a co-constructed narrative shaped by rapport, setting, and the linguistic repertoires of the interlocutors. In many cases, participants mixed English with Spanish, Portuguese, or Italian, a natural manifestation of the translanguaging practices that FLO sought to document.

Figure 31. Interview with one of the community managers of Digital Nomads Madeira at the Coworking space in Ponta Do Sol.

Interview guides were prepared before the start of the ethnographic immersion to ensure thematic coherence across encounters. Questions revolved around four major domains:

1. Mobility trajectories (e.g., 'How did you start living this lifestyle?' 'What places have you lived in and why?');
2. Language repertoires and learning (e.g., 'What languages do you use in your everyday and at work?' 'Have you learned new languages since becoming a digital nomad?' 'How important is linguistic competence in your life?');
3. Work and community practices (e.g., 'What is your relationship with coworking and coliving spaces?' 'How do you connect with other nomads?');
4. Language ideologies and cultural belonging (e.g., 'Do you feel English is necessary to be part of the nomad community?' 'Have you ever felt excluded because of language?').

These themes drew on the exploratory findings of the *Where's Your Language At?* pilot survey (Tedesco 2024b) and on the theoretical frameworks discussed in Chapter 3 (translanguaging, linguistic fluidity, plurilingualism, and language ideologies). They were designed to elicit reflections on both language use and the symbolic value of linguistic competence within digital mobility cultures. The final sample comprised 3 in-depth recorded interviews and dozens of informal conversations, involving participants aged 24–55 from more than 10 different national backgrounds.

Because FLO's main interest lies in spoken interaction, interviews were approached not merely as sources of reported data but as sites of communication to be analysed in their own right. Following Copland and Creese (2015), this approach treats interviews as interactional events where both participants and researcher co-construct meaning through discourse. Linguistically, interviews provided rich evidence of pragmatic features typical of translingual communication, like pauses, reformulations, back-channeling, and metalinguistic comments, thus bridging the gap between ethnography and corpus data analysis.

For example, during an interview in Fuerteventura, a Spanish-speaking participant reflected on her hybrid linguistic identity: 'Sometimes I start speaking in English, then switch to Spanish if I see someone doesn't get me. It's like… we all speak a bit of everything, right?' Such utterances are not merely statements about language but enactments of translanguaging, illustrating how participants negotiate and validate mutual understanding in superdiverse contexts (Blommaert & Rampton 2011; García & Wei 2014).

The few recorded interviews were conducted with participants' informed consent, following the University of Bologna's ethical guidelines and GDPR (EU 2016/679) (see 4.2.2). Thematic coding focused on domains such as identity, mobility, language use, *and* community participation, while linguistic coding addressed language alternation, and metalinguistic awareness. The resulting dataset served both as qualitative evidence and as an empirical guide for corpus annotation.

The three interviews were conducted in Madeira and Fuerteventura with individuals whose professional roles position them at the centre of digital-nomad community building, governance, and everyday social organisation. Mikaela, interviewed at the coworking space in Ponta do Sol, is the marketing and programme manager at Startup Madeira. As one of the architects of the Nomad Village project, she offered the institutional perspective of someone who helped design the ecosystem from scratch. Nuno, interviewed at the Estalagem (in Figure 32) – a venue that hosts many of Madeira's digital nomad community emblematic events– organises *Purple Fridays*, a weekly gathering that has become a cultural symbol for nomads across Europe. His account shed light on community rituals, event culture, and the social atmosphere that holds the group together. Finally, Diego, interviewed in Corralejo, manages the Surfing Colors coworking space in Fuerteventura and provided a critical, grounded account of nomadism in the Canary Islands, with a particular focus on linguistic barriers, gentrification, and cultural challenges.

Figure 32. A picture taken at the Estalagem hotel in Ponta do Sol.

Identity and community building from the perspective of community members

A close analysis of the interviews reveals that a first shared theme concerns the emergence and consolidation of digital nomad communities during and after the pandemic. Mikaela traces the roots of the Madeiran project back to long-standing local entrepreneurship and hospitality infrastructures: Startup Madeira, she explains, has been active for '26 years… always helping entrepreneurs, start-ups, and innovation on the island.' When COVID-19 hit, the island's reputation as a safe destination intersected with the growing remote work movement. She recalls how the arrival of NomadX founder Gonçalo in 2020 triggered the idea of building a village specifically for nomads: 'He said, "Why not create a nomad village? A place where people can stay, work, eat, and find community."' The Centro cultural de Ponta do Sol – empty

during the lockdown – was transformed into a coworking space, monthly accommodation packages were negotiated with local property managers, and a dedicated website was launched to centralise information. The rapid response resulted in an unexpected global resonance: 'We opened in February 2021… and very soon we had 200 registrations overnight. Now we have nearly 18,000 from 138 countries.'

Nuno's account complements this narrative by emphasising the cultural foundations that pre-dated the nomad boom. The Estalagem and the Centro Cultural had already hosted festivals and artistic residencies for over a decade. When nomads started arriving en masse during the pandemic – 'more than a thousand everywhere on the island… because the world was closed' – existing cultural infrastructures adapted. Purple Fridays emerged from this convergence, becoming, in Nuno's words, 'a networking event, a teaching event, and without realising, a very good party with amazing energy.' The event's international reputation reflects the connectivity of nomad networks: 'We receive messages from New York: "I want to come to Madeira just to experience Purple Fridays."'

In contrast, Diego's narrative describes a far more precarious trajectory in Fuerteventura. His coworking–coliving initiative flourished during the 2020 lockdown because Surfing Colors stayed open when most hotels closed, offering one-month stays for €600. 'It was a success… the hotel was full while others were empty.' But as tourism resumed, the structural tensions resurfaced: housing shortages, dependency on tour operators, and a lack of institutional mediation. Diego bluntly summarises the current situation: 'We are not making social impact. We are creating bubbles.' His perspective highlights the internal fragmentation of many nomad hubs and the absence of cohesive governance comparable to Madeira's.

A second major theme is community structure and self-organisation**.** Mikaela explains that the Madeiran model works because it is collaboratively built: 'I am just representing many people… hotels, property managers, restaurants, communities, the government.' Digital nomadism in Madeira is described as a coordinated ecosystem where top-down infrastructure (Startup Madeira, NomadX) meets bottom-up communities (Madeira Friends, NomadEscape, various local initiatives). Nuno confirms this dynamic, describing the early period (2020–22) as 'crazy… two activities per day, masterclasses, events in the village, yoga, coaching, language lessons… everything.' The weekly programme published on Slack ensures constant opportunities for integration: 'Anyone arriving solo will never feel alone. Every day there is something.'

Diego's description of Corralejo, however, reveals the fragility of community building in contexts where organisation is decentralised and often commercialised. He notes that many event organisers 'ask for a lot of money… but don't give service,' which frustrates nomads and fuels local resentment. Without institutional coordination, he argues, the ecosystem lacks cohesion and tends to produce short-lived or exclusive enclaves.

Community linguistic practices from the perspective of community members

A third transversal theme investigated during interviews is, of course, that related to language practices and linguistic backgrounds. All three interviewees agree that English is the central lingua franca, but their interpretations differ. For Nuno, the linguistic diversity is striking: 'You hear German, Polish, Italian… but official communication must be in English. There's no way we can create another common language.' Attempts to create subgroups by nationality 'didn't work, because they moved away from the main community.' Mikaela attributes Madeira's community communicative success partly to its long hospitality tradition: 'We have 100 years of hospitality… people here always find a way to communicate, even older people who don't speak English, gestures, mimic, friendliness.' The island's multilingual adaptability is presented as a culturally embedded behaviour rather than a learned skill. Diego, however, is much more critical: 'Most nomads speak only English. This is the biggest barrier. If you don't speak the local language, you live in your bubble. It's impossible to connect.' He gives a concrete example: 'Even making queso majorero… with a local farmer who doesn't speak English… it's impossible, misunderstandings everywhere.' His interpretation highlights asymmetrical multilingualism and the risk of cultural detachment in nomad enclaves.

Finally, the theme of socio-economic impact and sustainability emerged in all interviews. Mikaela framed nomads as agents of economic diversification and social innovation: 'Technology represented €600 million last year… combining tourism and tech is the future.' She believes digital nomads bring local consumption, international visibility, and long-term opportunities, although she acknowledges that housing remains a challenge. For her, the model is scalable: 'With the right people, the right partners, and the right values, it can work elsewhere.' Nuno sees culture as the heart of 'integration' and describes the community as 'flowing now… self-organised, but still very strong.' Diego, by contrast, warns that without careful planning, nomadism risks exacerbating local tensions: 'You are in a paradise, but it's your paradise, not the locals.' He stresses that without linguistic bridges and shared spaces for interaction, 'it's impossible to connect. Impossible.'

Overall, the interviews portray digital nomadism as a multilayered sociocultural ecosystem shaped by governance, cultural heritage, infrastructural choices, and language practices. Madeira emerges as a coordinated, socially embedded model where hospitality, institutional support, and multilingual openness sustain a thriving community. Fuerteventura illustrates the opposite tendency: fragmented, and socio-economically uneven. However, results from ethnographic participation in this island helped smoothing what emerged as two opposite poles in interview results.

As Tusting (2020) argues, grounded theory within linguistic ethnography functions best when 'light' – that is, empirically driven yet sensitive to existing frameworks. As already noted, FLO's analytic cycle exemplified this balance: observations generated provisional concepts that were subsequently tested against data and refined through comparison with relevant literature (e.g., Blommaert 2010; García & Wei 2014; Vertovec 2007).

Importantly, insights from interviews and diary entries also guided the design of quantitative instruments. Questions for the *Where's Your Language At?* survey (Section 4.4.3) were directly

derived from recurring ethnographic themes, particularly those concerning perceived linguistic competence, belonging, and attitudes towards monolingualism. In this way, qualitative and quantitative components remained in constant dialogue, providing a sequential exploratory design (Creswell 2014).

Finally, both interviews and journals raised questions of representation – i.e., whose voices are heard and how they are mediated. In line with the project's participatory ethos, excerpts from interviews were shared with participants whenever possible. The journal also documented the affective dimension of this reciprocity. Writing about people one has lived and worked with entails emotional investment; yet acknowledging this investment enhances rather than undermines validity. As Davies (2008) reminds us, 'reflexivity is not self-indulgence but methodological awareness of how knowledge is produced.' The FLO diary thus functioned as a reflexive mirror where personal involvement and scholarly distance coexisted, allowing a richer, more human account of the cultural and linguistic dimension of digital nomadism.

Netnography

Netnography's power lies in its ability to reveal the traces of social interaction within digital artefacts. During the observation period, several recurrent communicative genres that structure digital nomad online discourse were documented:

I. Announcements and Coordination Threads: posts or Slack messages organising meetups, volunteer days, or professional talks. These often included linguistic politeness markers ('Hey lovely people!' 'Anyone up for…?'), reinforcing community warmth.
II. Reflexive Posts: Instagram captions reflecting on the emotional and existential aspects of nomad life ('Sometimes freedom feels lonely'). These functioned as narrative rituals of self-disclosure, strengthening authenticity.
III. Advice and Knowledge Exchange: typical of Slack and Whatsapp groups, where participants share visa information, tax guidance, or remote-work tools. This genre displays an epistemic economy (Myers 2010) based on reciprocity and trust.

One of the first findings of the netnographic observation was that the linguistic ecology of digital nomad communication mirrored the spatial ecology of mobility. Online spaces like Slack channels, Instagram ashtags, and Whatsapp groups functioned as 'translocal assemblages' (Greiner & Sakdapolrak 2013): flexible infrastructures connecting geographically dispersed actors through shared semiotic repertoires. Within these spaces, the same communicative resources – words, emojis, memes, photographs – were repeatedly mobilised to perform belonging and authenticity.

The Slack platform with separate channels for all communities around the world works as the core infrastructure for coordinating community activities. Access links can be found on websites like *remoteworkcommunity.com*, *nomads.com*, or simply by typing into a search

engine '[place] digital nomad community slack'. You normally need to register and wait for the approval of the community managers.

Slack's key features include:

- channels (e.g., #events, #announcements, #questions) to keep conversations organised (as shown in Figure 33);
- direct messages between members;
- file sharing and integration with tools like Google Drive, Notion, Zoom, etc.;
- searchable messages, so past information is easy to retrieve.

It works as a central hub where members can coordinate events, share updates, ask questions, and interact daily. For digital nomad communities, it becomes the virtual common room where newcomers integrate and ongoing activities are organised.

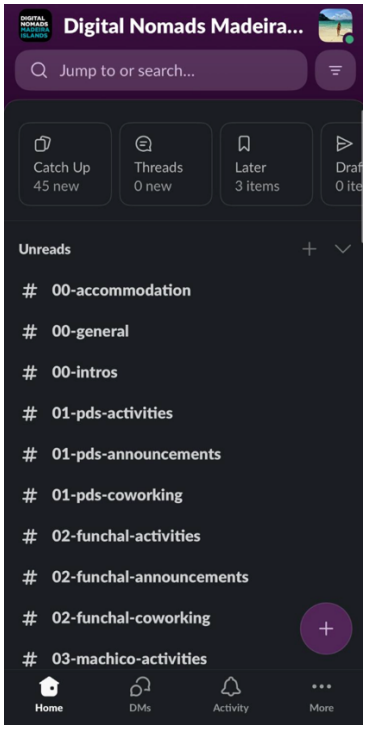


Figure 33. Organisation of the Slack channel used by the digital nomad community in Madeira.

Besides Slack, another platform born for remote work coordination which digital nomads took up for community organisation and knowledge sharing is Notion. It is an all-in-one digital workspace used for organising information, managing projects, and collaborating with others.

It allows one to create pages, databases, calendars, notes, task lists, and shared documents. For these chacarteristics, it is used by some digital nomad communities as a knowledge base where information is stored, updated and m made accessible to everyone. In Madeira, for example, it is used to list shared houses and flats for mid-term rent – this is how I found my first accommodation on the island.

Digital nomads often adopt the same digital platforms they rely on for remote work – such as Slack and Notion – to coordinate community activities, since these tools are already embedded in their everyday professional practices. This aligns with recent research showing that mobile knowledge workers augment their daily life with digital infrastructures, using familiar collaborative tools to sustain both work and social organisation, in a lifestyle where boundaries between work and other spheres of life becomes blurred (Thompson 2018; Nash et al. 2021, see Chapter 2).

For example, the Slack workspace of Digital Nomads Madeira shown in Figure 34, accessible only to registered members, functions as a semi-public intranet for event organisation, housing offers, and professional networking. FLO's observation log from December 2023 notes the coexistence of institutional voice (community managers announcing official partnerships) and peer-to-peer interaction (members arranging meetups, asking language questions). The #announcements channel displayed an institutional voice: community managers posting weekly schedules, updates on partnerships, or safety alerts (e.g. 'road closure on ER222 – avoid driving to Ponta do Sol this afternoon"). Beneath these messages, emoji reactions (👍, 🌞, 🙏) serve as micro-acknowledgements that help maintain the channel's visibility without overloading it with replies.

The #weekly-programme channel act as a ritualised temporal anchor: every Sunday a community manager posts a block-formatted schedule (e.g. Monday yoga, Tuesday coworking lunch, Wednesday hike, Friday Purple Fridays), and members responded by tagging friends, asking logistical questions, or signalling participation ('I'm in for the levada! Anyone driving from Funchal?'). This creates a stable rhythm of interaction that structured both online and offline sociality.

Peer-to-peer channels such as *#skills-exchange* and *#meetups* host spontaneous sociability: offers of workshops (see Figure 34), requests for language help and micro-collaborations (in Figure 35, for example, a community member is asking others to join a social initiative to produce positive impact on the local socio-cultural network), promotion of social events (see Figure 36) etc.. The Slack workspace therefore blurs the line between institutional coordination and bottom-up community-making, enabling a multilayered ecology of participation.

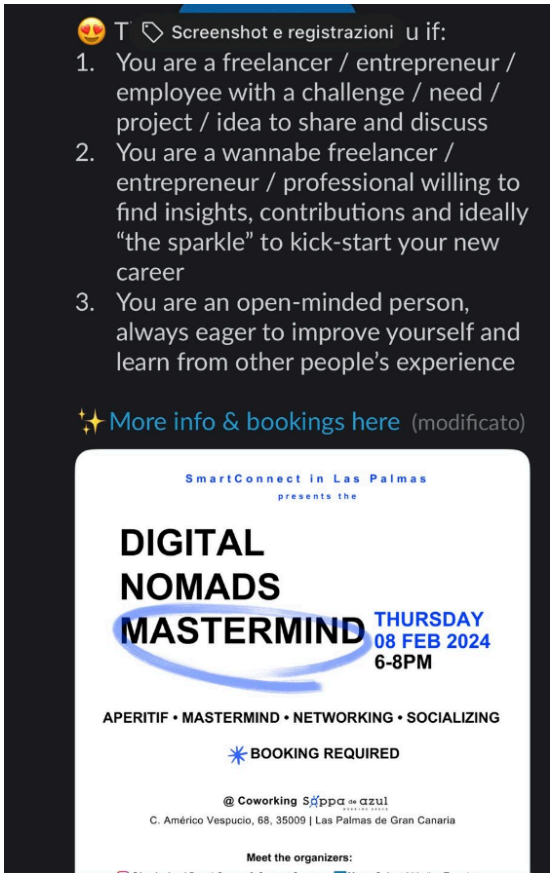


Figure 34. Promotion of a professional workshop and networking event on the Las Palmas de Gran Canaria Slack channel.

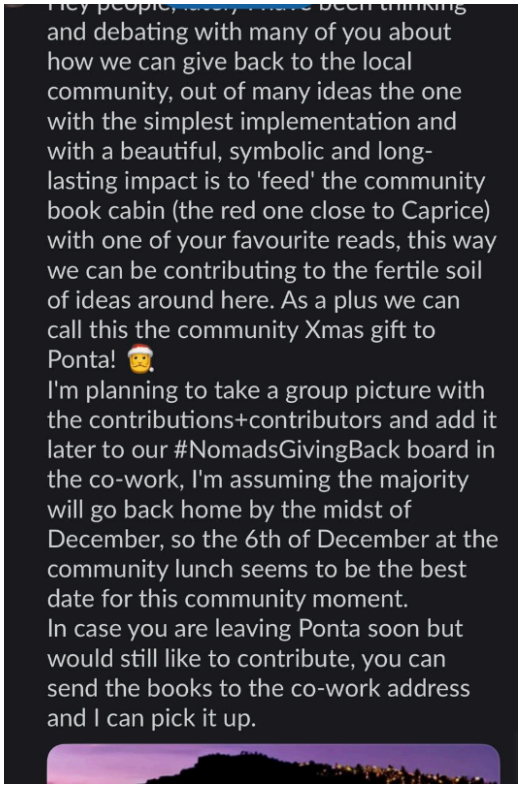


Figure 35. A post on the Madeira Digitla Nomad Slack Channel.

Figure 36. A social event promoted on the Las Palmas de Gran Canaria Slack Channel.

Linguistically, messages on Slack displayed a specific genre of 'digital professionalism' (Tagg 2023): concise, upbeat, and inclusive.

Notion, on the other hand, served as a collaborative workspace where documents like guides, maps, coworking recommendations were shared. These artefacts, while not conversational, provide evidence of knowledge co-production: community knowledge continuously updated and collectively owned.

At a lower level of formality, WhatsApp functions as the high-tempo, low-threshold layer of communication. If Slack is the structured backbone, WhatsApp is the *nervous system* through which rapid coordination occurred. The groups observed during the fieldwork were considerably more fluid, informal, and ephemeral than Slack channels. Their names often indexed specific sub-communities or temporalities. In Madeira, for example, the group 'Madeira Friends', coordinated by the managers of the non-profit organisation Madeira Friends, is divided into subgroups like 'Madeira Hikes', 'Purple Fridays Early Birds'. In Fuerteventura, the group Fuerteventura Remote Workers is the general one, but you also have Fuerteventura Nomad Surfers, and a closed group containing over 100 members named 'Cenas y Cavernicolos'. An example of the type of messages posted in the groups is provided in Figure 37.

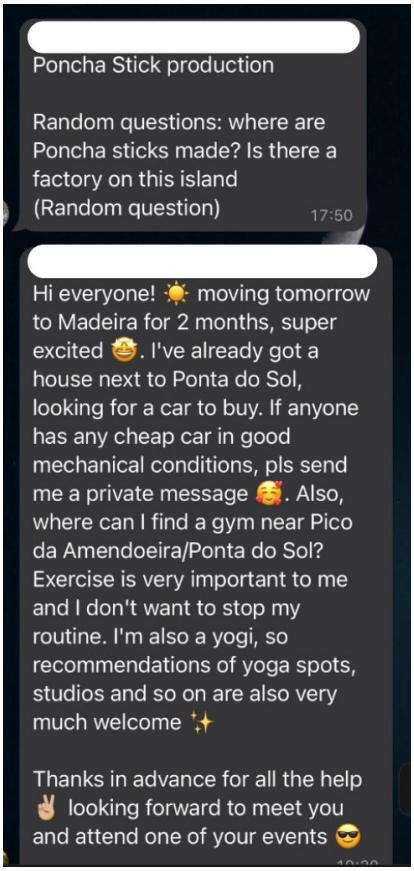


Figure 37. A screenshot of messages posted on the Madeira Friends Whatsapp chat in the sub-group related to the community based in Ponta do Sol.

WhatsApp has also been observed to facilitate micro-temporal coordination around mobility: in the Car Share Funchal group, messages like 'Leaving PDS for Funchal in 10 minutes, 3 spots available' appeared frequently. These micro-negotiations of transport illustrate how linguistic and spatial mobility intertwine: everyday communication practices directly shaped patterns of movement across the island.

In the *Hikes* group, interactions relied heavily on multimodal resources: GPS routes, photographs of weather conditions, screenshots, and quick safety checks (e.g.,'is the PR1 open today?'). Here, the interplay of images, brief text, and emoji (🔥, ⛰️, ⚠️) created a compact multimodal register suited to outdoor coordination.

The WhatsApp environment also displayed code-switching more visibly than Slack. Members often shifted between English, German, Spanish and Portuguese within the same thread, depending on situational cues (e.g., when responding directly to someone of the same background). This created an informal plurilingual repertoire that reflected the superdiverse composition of the nomad groups (an example is provided in Figure 38).

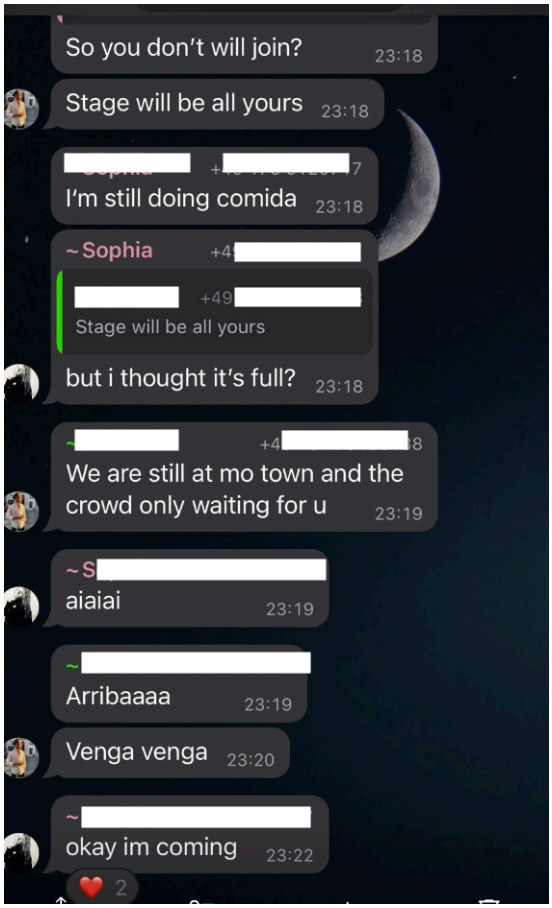


Figure 38. A screenshot of a conversation on the Whatsapp group of the Surf&Kite Nomad House.

In this conversation, Spanish is used even if none of the speakers are Spanish, mixed with non-standard informal utterances in English.

While Slack maintains an intentional boundary (registration-gated, moderated, semi-public), WhatsApp groups were porous and dynamic: members joined or left depending on location, mood, or temporary interests. This ephemerality was reflected in the frequent creation of short-term micro-groups ('Tonight's dinner', 'Working from Estalagem – join?'), which became inactive once the activity ended.

Slack and WhatsApp, thus, illustrate a stratified model of digital communication which augments social interactions at digital nomad communities. Both platforms rely on English as the primary lingua franca but were open to translingual practices, demonstrating how linguistic fluidity is embedded across layers of digital interaction. Their combined use produces a communicative infrastructure that sustain local belonging while connecting members to wider transnational networks, reflecting the dual nature of digital nomad communities as both rooted in certain locations and fluid, as members' mobility across different locations is conceived as a normal practice.

Instagram also plays a central role in the social and communicative ecology of digital nomad communities, functioning simultaneously as a personal storytelling platform and an organisational tool. Through private profiles, nomads document mobility trajectories, professional identities, and everyday experiences, creating highly curated narratives that facilitate self-presentation and community visibility. At the same time, organisational profiles, run by coworking spaces, event organisers, and community managers, serve as public-facing hubs for announcing activities, promoting local initiatives, and attracting newcomers. Instagram thus operates as both a personal and collective interface: it supports network-building, disseminates community information, and reinforces a sense of belonging across dispersed and constantly shifting groups. In this respect, the Instagram page the.nomad.linguist created for this research project positions itself at the intersection of these two uses (see Section 4.2).

In general, Instagram visual immediacy and algorithmic reach make it one of the main channels through which digital nomad communities maintain continuity despite high levels of mobility.

Posts from Madeira's Friends, Digital Nomads Madeira, Gran Canaria's Live it up Las Palmas, Corralejo Remote Work Community frequently combine English captions with visual imagery of sunsets, coffee cups, laptops, and coworking terraces (Figure 39). Linguistically, they deploy recurring lexical items – i.e., freedom, flow, connection, tribe, balance – which together form a recognisable discourse of 'lifestyle entrepreneurship' (Duffy 2017). Seemingly, in personal profiles nomads represent their lifestyle at the intesection of work and philosophy of living (Figure 40). This discourse naturalises a neoliberal narrative of work and self-realisation while masking inequalities within nomad mobility (Thompson 2019).

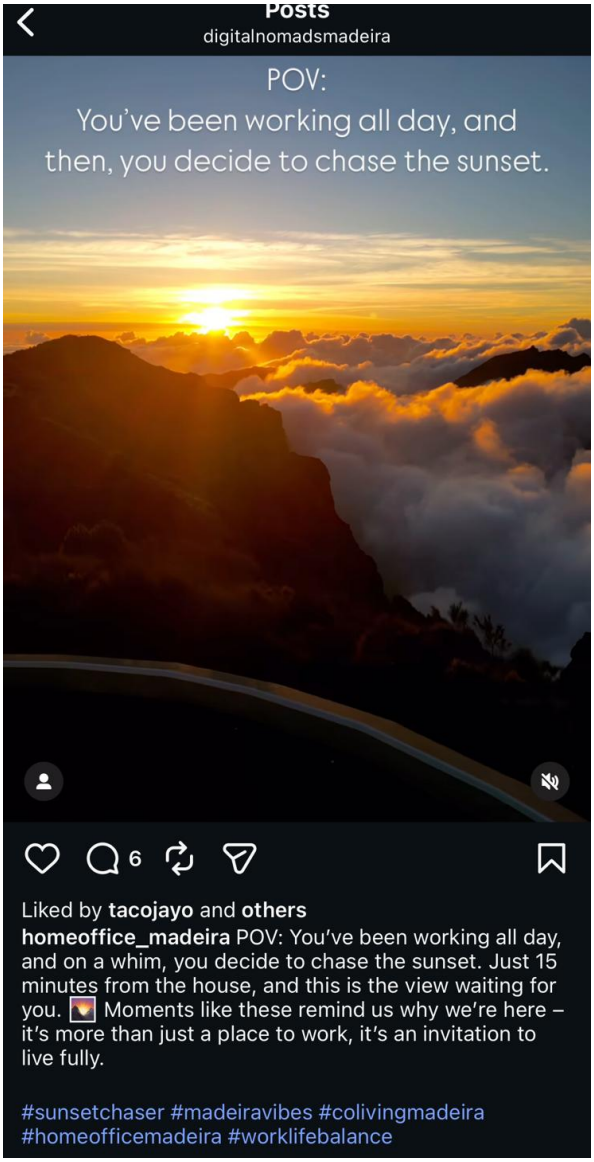


Figure 39. A screenshot of an Instagram post published on Digital Nomads Madeira.

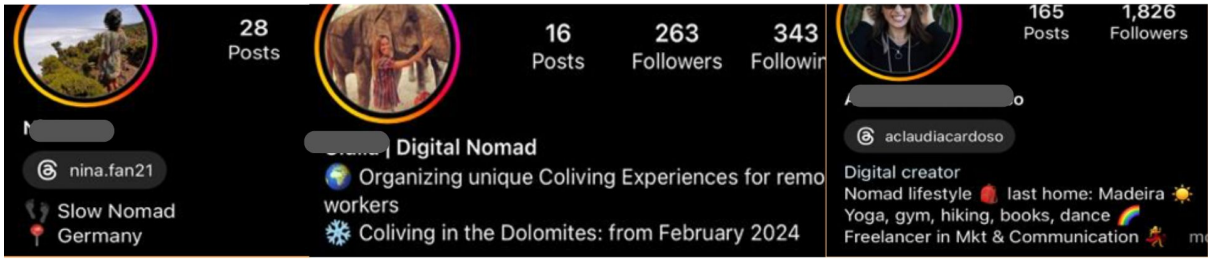


Figure 40. Screenshots of digital nomads Instagram bios.

The ethnographic exploration of hashtags revealed their dual role as linguistic signifiers and social connectors. Hashtags such as #digitalnomadmadeira, #remoteworklife, #coworkingspace, and #fuerteventuranomads did not simply categorise content; they acted as

infrastructure for visibility and affiliation. As Zappavigna (2018) notes, hashtags create 'ambient affiliation' allowing users to align affectively and ideologically with others they may never meet offline.

A qualitative exploration of ashtags referring to digital nomadism was also conducted. The following ashtags were considered on Instagram:

- #digitalnomad;
- #digitalnomadism;
- #digitalnomadlife;
- #nomadlife;
- #workfromanywhere;
- #laptoplifestyle;
- #remotework;
- #workandtravel;
- #digitalnomadlifestyle;
- #digitalnomadgirls / #digitalnomadwomen.

These hashtags provided useful information to better understand the identities and communities under investigation. The Instagram posts in many different languages and related to locations all around the world confirmed the transcultural nature of digital nomadism.

From a more strictly linguistic perspective, netnographic observations of Instagram posts align with observations of other online social media: English functions as the default linguistic resource in Instagram captions and posts associated with digital nomadism, reflecting its role as a global lingua franca within mobility-based lifestyles. Nonetheless, it is common for users to embed words, expressions, or short phrases from the local languages of the places they inhabit, for example, using emblematic terms associated with affect, place attachment, or local cultural practices. These linguistic insertions, as the use of 'tapas' in the post shown in Figure 41, seem to serve as indexical anchors, signalling emotional proximity to place and aligning one's mobility narrative with the cultural environment.

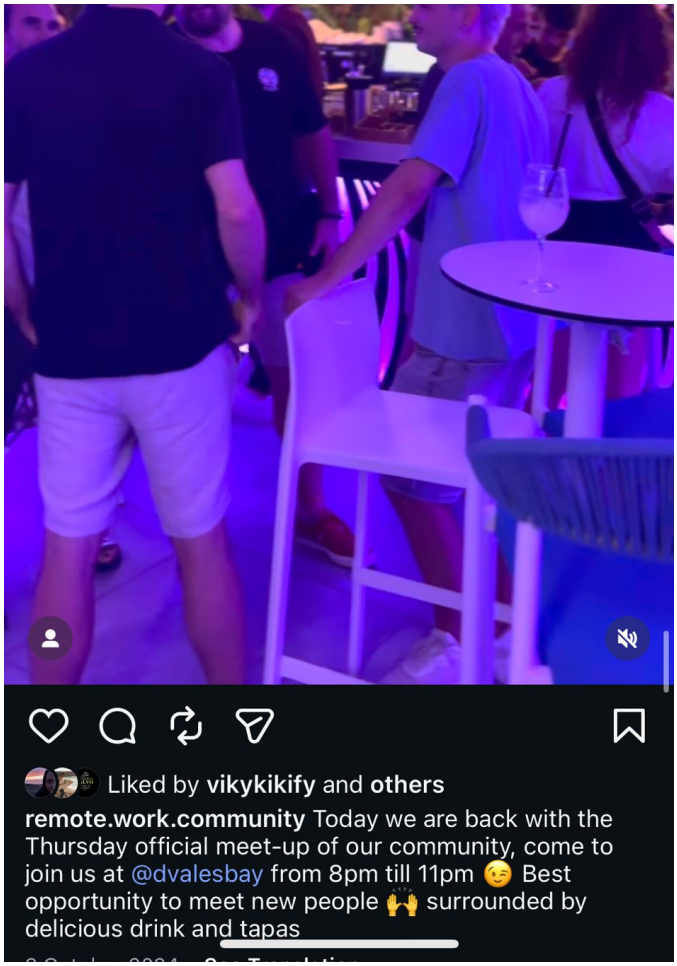


Figure 41. Screenshot of an Instagram post by one of Corralejo's community managers.

This phenomenon can be understood through the concept of *metrolingualism* (Pennycook 2010), according to which the fluid use of semiotic resources responds not only to the micro-communicative environment (speaker, interlocutor[s], place, medium etc.), but also to the larger cultural and geographical environment which the used signs are linked to. In this perspective, digital nomads' social media practices do not merely reflect bilingualism or code-mixing; rather, they enact a translocal identity performance**,** combining globalised communicative norms with locally meaningful linguistic traces. Such micro-local markers function as semiotic resources through which users negotiate belonging, authenticity, and cultural sensitivity while simultaneously reaffirming their participation in a global, mobile lifestyle.

As can be noticed in the figures presented in this section, in digital contexts translanguaging extends beyond spoken code-switching to include all semiotic resources – typography, colour, symbols, and images. Posts collected under #digitalnomadmadeira often combined short English texts with emojis representing sound, movement, or emotion, creating a multimodal syntax (Kress 2009).

An important insight from FLO's digital observations concerns the affective dimension of online interaction. Posts about loneliness, friendship, or 'finding one's tribe' were frequently observed. In this sense, visual cues, heart emojis, group photos, videos of sunsets, act as affective signifiers that support belonging across distances. Drawing from Papacharissi's concept of 'affective publics' (2015), these practices have been interpreted as collective emotional labour performed through language and image. For example, when a member announced leaving Madeira, farewell messages often combined linguistic warmth ('Safe travel brother!') with symbolic reactions (🔥,💙,🌍). These small gestures create networked intimacy (Baym 2015), a sense of closeness maintained through continuous, low-intensity interaction.

Many interviewees confirmed that digital platforms were not just practical tools but social and emotional anchors. As one participant explained during an interview in Gran Canaria: 'When I move, I already have my next Slack or WhatsApp group waiting.' This portability of sociality illustrates the merging of online and offline ethnography: digital communication does not replace presence but extends it, allowing mobility to coexist with continuity.

As the digital fieldwork progressed, participation evolved from observation to collaboration. The Instagram profile the.nomad.linguist, which I developed at the beginning of the ethnographic field research in Madeira was fundamental to keep collaboration with participants (as has been discussed in the first section of the present chapter).

Beyond its empirical contribution to the present research, engaging with netnography has offered theoretical insights into linguistic ethnography at large. It suggests that the study of language in mobile, superdiverse contexts must move beyond the physical field to embrace the whole networked field – a space constituted by digital traces, mediated communication, and algorithmic visibility. In this environment, ethnographic sensitivity remains crucial but must be complemented by technical literacy and awareness of platform dynamics. Thereby, integrating netnographic observation in the FLO project also aimed at demostrating that linguistic ethnography in the age of mobility requires methodological pluralism: immersion in both physical and digital contexts; attention to multimodal and translingual repertoires; reflexivity on the part of the researcher as both participant and algorithmically mediated observer. Digital nomad communities, with their translocal networks and continuous connectivity, epitomise the need for such hybrid approaches.

Community-based knowledge co-creation

A central principle of the FLO project is that knowledge about digital nomad communication should not be extracted from participants, but co-constructed with them. This orientation aligns with traditions of participatory action research (Lewin 1946; Reason & Bradbury 2008), community-based participatory research (CBPR) (Israel et al. 2013), and collaborative ethnography (Lassiter 2005), all of which position community members not as passive research subjects but as active contributors to the generation, interpretation, and circulation of knowledge.

Digital nomad communities are especially suitable for collaborative methodologies. As noted earlier, their organisational structures often resemble communities of practice (Wenger 1998), comprising a core group (community managers, coordinators, long-term residents), an inner circle of more engaged members, and a peripheral layer of newcomers and occasional participants. For example, in a post on the Remote Work Community Instagram page (Figure 42), community managers announce that a meeting had been held in the Surfing Colors coworking in Corralejo, Fuerteventura, to take 'decisions on the future of the community'. Such a stratified but flexible organisation facilitated FLO's efforts to involve participants at different degrees of intensity and responsibility throughout the research cycle.

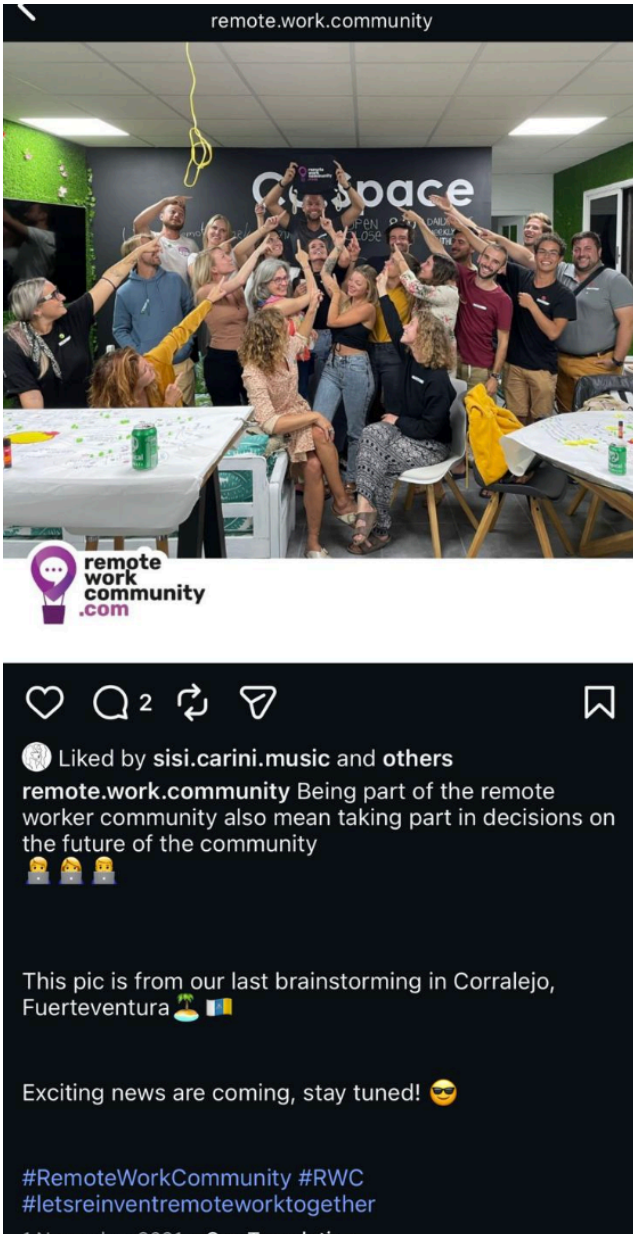


Figure 42. A screenshot of a post published on Remote Work Community Instagram profile.

More precisely, community-based knowledge co-creation goes beyond consultation: it implies that participants contribute to defining relevant questions, validating interpretations, and

shaping outputs – something which was achieved also through the informal interviews above described. During fieldwork in Madeira, Porto Santo, Gran Canaria, and Fuerteventura, community members participated directly in several phases of the research. For example, they helped identifying relevant sociolinguistic issues. Many participants during the foregrounding research phase voiced concerns about communication, accent negotiation, misunderstandings, adaptation to different varieties of English, code-mixing practices, or the pragmatic implications of using English as a lingua franca. These discussions influenced the design of methodological tools for data collection as well as the formulation of the research questions which are explored in Chapter 6. Moreover, participants had the role of validating interpretations. On several occasions, preliminary interpretations of certain statements or concepts which came up in interviews were informally rediscussed with participants during community events, coworking sessions, or through Instagram posts and stories directly engaging audience with polls and open questions. Indeed, because digital nomad communities live 'onlife' (Floridi 2015) – i.e. simultaneously online and offline – part of the co-creation took place through digital platforms. These acted as participatory arenas, where research questions could be tested and refined through interactions with community members. In this way, the recurrent practice of mixing English with local lexical items (e.g., Portuguese, Spanish) was identified and validated collaboratively, long before systematic corpus analysis confirmed its relevance (see Chapter 6).

In line with CBPR principles, FLO actively pursued forms of shared dissemination:

- blog articles deriving from interviews were co-reviewed with some participants before being published on *The Nomad Linguist* website;
- participants contributed to the transcription and interpretation of selected excerpts of conversations, sometimes helping identify humour, cultural references, implicit alignments, or stance-taking markers that would be invisible to an outsider;
- community managers in Madeira and the Canary Islands were regularly updated about preliminary findings and provided feedback on the accuracy of contextual descriptions.

But co-creation also shaped the data collection and management processes:

- consent models were adapted based on community feedback (a lawyer member of the Fuerteventura community openly validated the consent model, reassuring participants in the corpus data collection);
- metadata categories (e.g., context of interaction, mobility background, language attitudes) were refined through conversations with participants who identified relevant dimensions for their communicative behaviours;
- some participants explicitly asked for access to anonymised survey data for their own purposes, reinforcing the relevance of open data.

Throughout the project, participants actively helped data and information gathering through spontaneous feedback, comments on social media posts, suggestions about which events or communities to observe; self-initiated reflections about their own linguistic experiences; in-

person and via-chat sharing of screenshots, Instagram reels, personal anecdotes that helped contextualise corpus data.

These interactions created reflexive loops, where community interpretations enriched the analytical process and the researcher's presence stimulated further metalinguistic reflection among participants, in a form of 'reflexive amplification' (Pink et al. 2016): the more participants recognise the researcher as someone studying communication, the more they articulate communication as an object of attention.

By grounding the project in community participation, several goals could be achieved. First, analytical depth, because community interpretations reveal nuances that a purely external analysis could not identify. Secondly, contextual validity, meaning that findings do actually reflect participants' lived realities, not just theoretical assumptions. Moreover, community participation ensure research ethical accountability; it enhance methodological innovation, as ethnographic approaches are increasingly recognised in corpus linguistics studies (as discussed during sessions and keynote speeches at some of the most important Conferences on the subject, like TALC2024 in Manchester and CL2025 in Birmingham). Finally, knowledge co-creation has lead to the development of a sustainable and impactful research project, enhancing long-term engagement and facilitating dissemination.

Thereby, the ethnographic, autoethnographic, and netnographic components of the FLO project demonstrate that meaningful research on mobile, translocal communities must be grounded in collaborative, situated, and reflexive knowledge practices. Community-based knowledge co-creation was not an ancillary dimension of the project, but rather the condition of possibility for understanding how digital nomads communicate, organise social life, and give meaning to mobility, for collecting and processing data.

Undertaking a participatory perspective has thus epistemological, methodological, and ethical consequences. Epistemologically, it allows linguistic ethnography to draw on the experiential expertise of participants, aligning with the sociolinguistics of mobility (Blommaert 2010; Blommaert & Rampton 2011) and with theories of superdiversity and transcultural communication. Methodologically, it distributes analytical responsibility across a network of actors, strengthening contextual validity and reflexivity. Ethically, it responds to recent calls in mobility studies for research that is transparent, reciprocal, and accountable to those whose lives it represents (Nowicka & Ryan 2015).

At the same time, co-creation highlighted the limits of ethnography alone. The richness and variability of linguistic practices observed in digital nomad communities made evident the need for additional layers of systematicity and breadth. For these reasons, the ethnographic phase naturally led to – and directly informed – the development of the Where's Your Language At? survey (which has been partly discussed in Section 4.3, and topic of the next sub-section). The survey operationalises insights that emerged from participatory observation and daily interaction, transforming them into variables suitable for quantitative analysis. It expands the research beyond the immediate ethnographic sites, allowing digital nomads living in other

regions to contribute data on language repertoires, communicative preferences, and mobility patterns. In doing so, it completes the first axis of FLO's mixed-methods design by broadening the empirical base established ethnographically and by preparing the ground for triangulation with corpus analysis.

The next Sub-section (4.4.3) therefore turns to survey research, examining details and presenting results of the questionnaire final version.

4.4.3 Survey research

Survey research has a long-established role in sociolinguistics and applied linguistics, where it is widely used to investigate language attitudes, self-reported linguistic practices, biographical trajectories, and ideological orientations that cannot be fully captured through interactional data alone (Garrett 2010; Dörnyei & Taguchi 2010; Schleef 2014). Although more recent methodological traditions in linguistic ethnography and discourse analysis approaches have privileged the analysis of naturally occurring interaction, questionnaires remain a central instrument when the research goal involves describing broader sociolinguistic tendencies across a population. Their strength lies in the ability to collect standardised, comparable answers from multiple participants, providing a level of systematicity that complements and extends observational findings (Schilling 2013).

Within the FLO project, survey research complements ethnographic observation and corpus analysis by providing structured, comparable evidence on participants' linguistic repertoires, mobility trajectories, attitudes towards English, and communicative orientations in transnational settings. As already introduced in Section 4.3, the questionnaire *Where's Your Language At?* was therefore conceived not as an isolated instrument but as a tool embedded within a broader mixed-methods architecture.

As discussed in Section 4.3, the study of digital nomadism poses considerable methodological challenges. Digital nomads constitute a highly mobile, geographically dispersed population without a stable sampling frame; their identity boundaries are fluid, and self-definition varies considerably according to personal, professional and ideological factors. Because of these characteristics, traditional fieldwork approaches alone may fail to capture the breadth of linguistic backgrounds and communicative orientations that characterise these communities. Hence, an online survey seemed to offer a pragmatic solution to the challenge of *following* participants across multiple locations, as it could be administered remotely, completed asynchronously, and distributed through digital channels that are already used by the communities under study.

Recent developments in sociolinguistics and mobility studies have increasingly emphasised the value of mixed-methods approaches in the study of migration and transnational lifestyles (Blommaert 2013; Canagarajah 2017). In particular, when investigating mobile or translocal communities, surveys provide a necessary complement to ethnography: they can broaden the empirical focus, generate comparative evidence across participants, and offer quantitative

grounding for qualitative interpretations. In FLO, the survey does not aim at making probabilistic claims about a bounded population – something unfeasible given the absence of a sample frame – but rather at producing theoretically meaningful insights within an exploratory design.

In practice, survey methodology appeared as particularly suitable in this context for at least three reasons. First, it facilitates the documentation of participants' socio-biographical profiles (profession, level of education, mobility patterns, linguistic biography). Second, it allows for the operationalise of complex constructs, such as plurilingualism, translanguaging, linguistic fluidity, or perceived linguistic needs, into measurable indicators and scales (Dörnyei & Csizér 2012). Indeed, not all of these concepts are not easily observable, thus require carefully designed items to ensure validity and reliability. Third, surveys enable the analysis of ideological and affective dimensions of communication, such as attitudes toward English as a lingua franca, adherence and/or resistence to monolingual norms. These dimensions often remain implicit in interactional data and may only surface when participants are explicitly prompted to reflect on their experiences.

From a methodological standpoint, adopting survey research requires careful consideration of the types of error that may arise in the design, administration and analysis of questionnaires. The Total Survey Error (TSE) framework (Biemer 2010) distinguishes between representation error, stemming from non-random sampling or nonresponse, and measurement error**,** which includes issues such as ambiguous wording, poor layout, sequence effects, or cognitive overload. Both types of error are particularly relevant when targeting populations like digital nomads, who are difficult to sample representatively and often complete surveys on mobile devices or in unstable connectivity conditions.

Convenience sampling emerged as the most viable strategy for this type of population. While probabilistic sampling remains the gold standard for inferential statistics, it is rarely attainable in sociolinguistic research involving open groups (Teddlie & Yu 2007). Convenience, time-place, and snowball sampling – which were all integrated to reach participants for the survey under discussion – are recognised alternatives in exploratory or community-based studies (Bornstein et al. 2013). Previous surveys on digital nomads, such as the Flatio Reports (2023, 2024), have used similar strategies and emphasise the impossibility of producing statistically representative samples in this field. Within this methodological frame, the purpose shifts from estimating population parameters to generating theoretically informative datasets and mapping tendencies. In practice, this implies acknowledging from the outset the epistemological status of the data: the results of *Where's Your Language At?* do not generalise to the total digital nomad population in a statistical sense, but they contribute to building and refining theoretical constructs about mobility, plurilingualism and transcultural communication. Such an approach aligns with contemporary sociolinguistic research on superdiverse populations (Vertovec 2007; Blommaert 2013), where the goal is not representativeness but the documentation of diversity, trajectories and emergent practices.

The adoption of an online survey also reflects a broader trend in sociolinguistic and applied linguistics research. As Schilling (2012) notes, web-based questionnaires can reduce interviewer effects, facilitate honest self-reporting and allow respondents greater flexibility.

At the same time, digital survey modes introduce specific methodological considerations. Dillman's Tailored Design Method (2014) stresses the need for coherent visual design, consistent formatting of scales, and minimisation of cognitive load in order to increase both response rate and data quality. In the pilot phase of *Where's Your Language At?*, several issues emerged related to layout, length and the visual presentation of Likert scales on LimeSurvey. These insights directly informed the redesign of the final questionnaire (see Section 4.3), illustrating the importance of iterative testing in mixed-methods research.

In short, the questionnaire *Where's Your Language At?* was conceived for two interconnected purposes. First, it seeks to document participants' linguistic trajectories and communicative strategies. The survey is aimed at collecting information on the languages participants speak, how they learned them, how they use them in different contexts, and how they perceive multilingual communication in transnational environments. Questions about mixing languages, translanguaging practices, technological mediation, and experiences of (mis)understanding provide a first layer of empirical evidence that can be compared with corpus data. Second, it allows for the creation of a structured dataset that can be triangulated with ethnographic and corpus components. The survey data allow for linking socio-biographical factors with features observed in interactional recordings. Reported use of ELF, perceived linguistic insecurity, or attitudes towards multilingual practices can help contextualise interactional patterns documented in the NomadLingo corpus. This cross-component triangulation is central to the analytical strategy of FLO, as it combines emic perspectives (participants' reported experiences), situated observation, and linguistic evidence.

The present section thus provides an account of how the questionnaire was designed, piloted, revised, and ultimately administered in its final version. Particular attention is given to the transition from the extended pilot study (which has been discussed in Section 4.3, Tedesco 2024b) to the full survey, highlighting the methodological choices that guided the simplification, restructuring and operational refinement of the instrument. The resulting anonymised dataset, accompanied by full question texts in .pdf format and documentation on administration procedures, has been deposited on Zenodo (Tedesco 2025b), in line with open science and FAIR practices.

Designing the questionnaire

The questionnaire *Where's Your Language At?* was designed following a rigorous and incremental process informed by sociolinguistic theory, survey methodology, and preliminary ethnographic insights gathered during the early phases of the FLO project. What follows is a discussion of the main conceptual, operational and technical considerations that guided the development of the instrument.

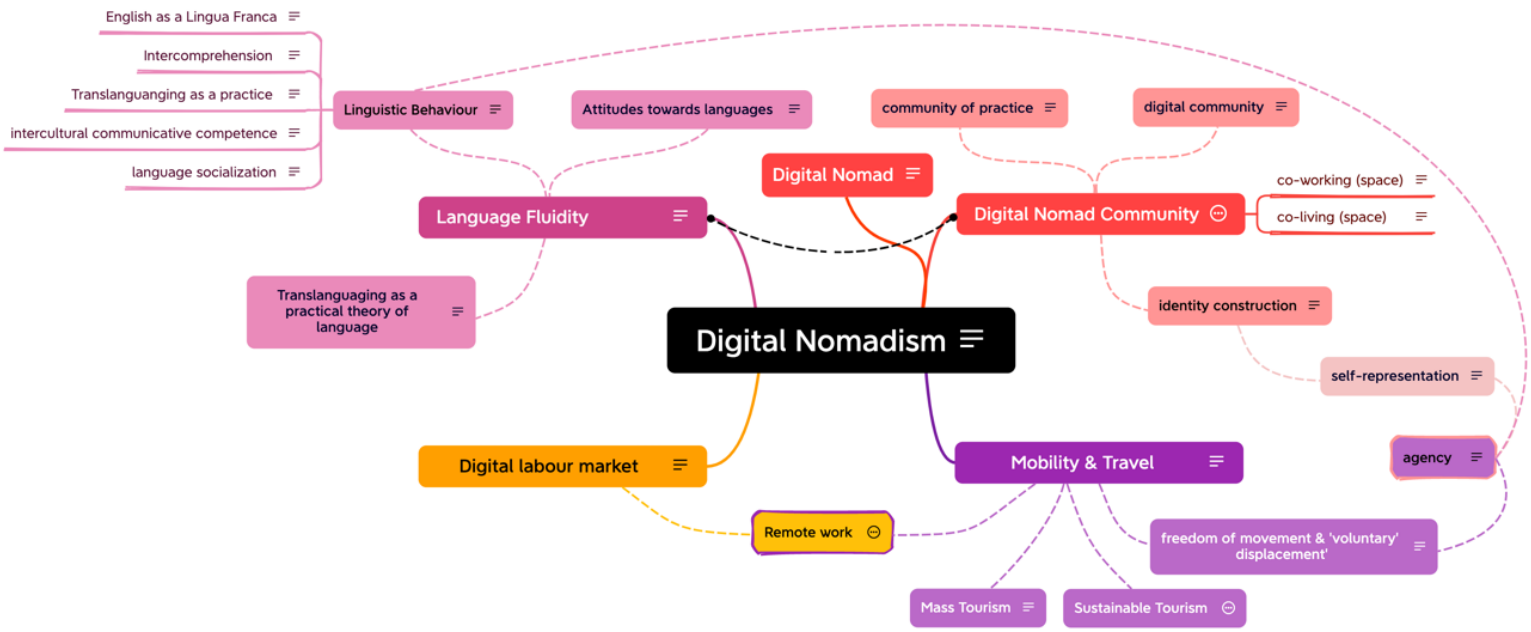


Figure 43. Conceptual mapping for survey design.

As shown in Figure 43, the conceptual starting point is digital nomadism. The questionnaire avoided rigid classificatory approaches and instead sought to capture behavioural, attitudinal and socio-biographical indicators that reveal proximity to nomadic lifestyles, also considering the blurred links between digital nomadism and socio-economical phenomena like mass tourism and digital labour (see Chapter 2 and Section 4.3 of the present chapter).

A second key conceptual area concerns linguistic behaviours, encompassing all observable aspects of language use, from the physical act of speaking (like pronunciation and intonation) to the complex social and cognitive processes behind communication, understood here as dynamic, practice-based collections of linguistic and semiotic resources (Bennett 1976).

Third, linguistic attitudes are conceptualised as feelings, beliefs, and perceptions toward a language or language variety and its speakers (Bourdieu 1991; Baker 1992; Garrett 2010). Methodologically, attitudes can be investigated implicitely (e.g., association tasks) or explicitly (Heller et al. 2015). The questionnaire includes several items explicitly designed to explore attitudes, perceptions, and idelogies towards plurilingualism, English predominance and translanguaging. Explicit questions were chosen over implicit tasks for two main reasons: first, in terms of survey readibility, they fit better with the general design of the questtionaire; secondly, in a view of research as knowledge co-construction and given the participatory nature of the project, all the items in the questionnaire were conceived to prompt participants' metalinguistic reflection.

Summing up, the process of designing *Where's Your Language At?* required translating the theoretical constructs discussed in Chapters 2 and 3, into measurable variables suitable for survey research. This translation, central in general survey methodology (Babbie 2016;

Dörnyei & Taguchi 2010), is referred to as operationalisation, i.e. the process of defining empirical indicators for abstract concepts. Because sociolinguistic constructs are often multidimensional and context-dependent, operationalisation in this project demanded a careful balance between theoretical accuracy and pragmatic measurability. Figure 44 shows how the concept of translanguaging was operationalised in one of the questions in the survey.

Look at these examples taken from social networks:

'Vamos to the beach!'

'We propose you a real Italian aperitivo'

'Estão readys?!'

How often do you mix different languages like in the examples above... ?

	Never	Seldom	Sometimes	Often	All the time
... while socialising	○	○	○	○	○
... at work	○	○	○	○	○
... when writing	○	○	○	○	○
... when speaking	○	○	○	○	○

Figure 44. A question on the practice of tanslanguaging in interaction in *Where's your language at?*.

A crucial component of this process was the extended pilot study (Tedesco 2024b, see Section 4.3) Practically, in line with best practices in survey design (Schleef 2014; Dillman 2014), operationalisation followed five principles:

I. Ensuring continuity with the theoretical framework;

Each block of the questionnaire corresponds to one or more conceptual areas elaborated in Chapter 2 and 3. For example, the conceptualisation of plurilingual repertoires as dynamic and practice-based motivated a distinction between languages that 'can be spoken', 'used daily', and 'learned' (and the different contexts of learning), as shown in Figure 45. The treatment of translanguaging and semiotic fluidity led to the introduction of open questions which could directly address linguistic creativity and metareflection, as shown in Figure 46. The analysis of ELF underpinned attitudinal items on ease of expression, perceived difficulties, and language preference in specific social settings. Thus, operationalisation is not merely procedural but embedded in the epistemological and theoretical commitments articulated earlier in the thesis.

Do you consider yourself bilingual (or plurilingual)?

*By 'bilingual' or 'plurilingual' we mean that you currently speak two or more languages, you can use them effectively and without particular effort to communicate in your daily life. Note that this definition is very broad: **feel free to give the answer you feel most appropriate** to describe your language situation.*

Yes No

How many languages can you speak?

By the expression 'can speak a language' we mean that you are generally able to communicate using this language. You should include in your count both the languages you use on a daily basis and the languages you know but don't use so often.

Which language(s) did you learn at school?

Select all that apply

German

Have you learned any other languages outside school (e.g. private language courses, online, socializing with your friends, travelling)?

Yes No

How strongly do you agree or disagree with the following statements?

*(1 – strongly disagree, 2 – disagree, 3 – **neither agree nor disagree,** 4 – agree, 5 – strongly agree)*

	1	2	3	4	5
I am satisfied with the education I got at school in English					
I am satisfied with the education I got at school in languages other than English.					
I am satisfied with the amount of intercultural activities I did at school (e.g. visits in museums, travels, cultural exchanges).					

Figure 45. Questions about linguistic repertoire in *Where's your language at?*.

Think about mixing languages, inventing words or using language in an unconventional way like in the examples above.

Are there any *strange* words or expressions that you use or have heard and would like to share with us?

Figure 46. An open question about translanguaging in *Where's your language at?*.

II. Using multiple indicators for multidimensional constructs;

As discussed in Chapter 2 and 3, constructs such as digital nomad identity, plurilingualism, linguistic openness, or transcultural communicative behaviour are inherently multidimensional. For this reason, the survey does not rely on single items but on clusters of questions capturing: self-perceived behaviours (e.g. frequency of English use across contexts), attitudinal orientations (Likert-scale evaluations on language preferences), experiential dimensions (perceived ease, misunderstanding), ideological positions (orientation toward monolingual vs. plurilingual practices). The pilot played a critical role here: it revealed which dimensions were clearer to respondents and which required simplification or reframing. Prioritising behavioural indicators over abstract categorisations;

One major insight emerging from both literature review and questionnaire piloting phase is that digital nomadism cannot be operationalised as a binary variable ('digital nomad / not digital nomad'). Identity labels are unstable and influenced by social desirability, personal ideology, and local community norms. The pilot showed that many respondents refrained from

self-identifying as digital nomads despite meeting behavioural criteria; conversely, some claimed the label while engaging in relatively sedentary or non-mobile lifestyles.

This highlighted the need for behavioural indicators (remote work percentage, reasons for travel, self-identification) rather than identity-based classification. Using behaviourals questions as they were in the pilot to identify respondents has proved complex. As a result, the final questionnaire includes three core behavioural variables that replace the original seven-criteria classification system used in the pilot, as shown in Figure 47.

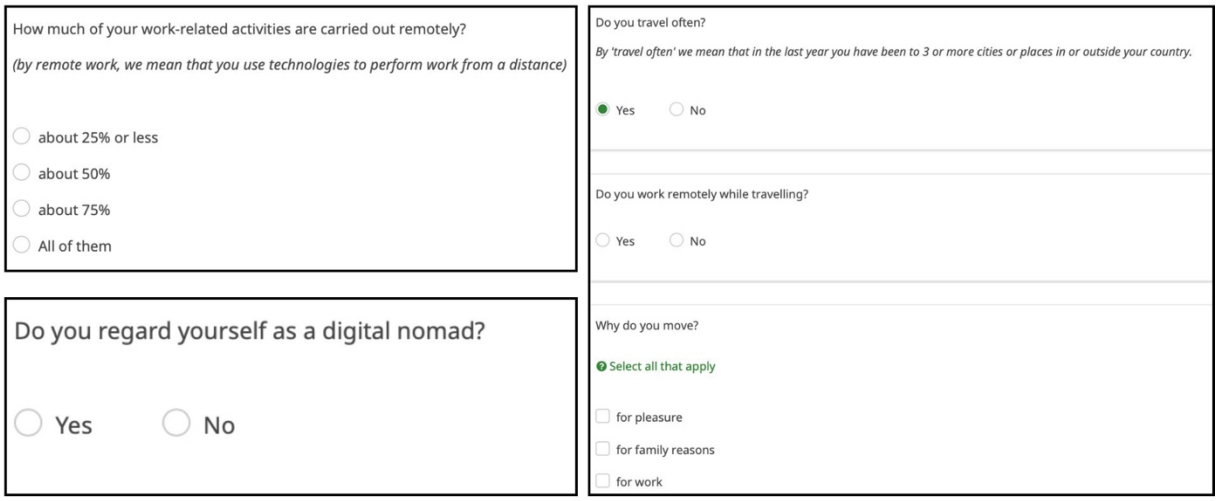


Figure 47. Behavioural questions to identify the target population in *Where's your language at?*.

III. Differentiating between behaviours, attitudes, ideologies, and resources;

Drawing on theoretical distinctions in Chapter 3 (e.g. attitudes vs ideologies; practices vs repertoires), the questionnaire separates:

- Behaviours: frequency-based items ('How often do you…?');
- Attitudes and ideologies: Likert items (evaluations, preferences; as shown in Figure 48);
- Repertoires: open or list-based questions on languages known and used (as shown in Figure 45).

IV. Allowing empirical testing of the theoretical relationships outlined in Chapter 3;

Operationalisation was designed to enable the investigation of theoretical hypotheses emerging from the literature review, including the relation between mobility intensity and linguistic fluidity (frequency of mixing or translanguaging), the association between mobility intesity and attitudes towards plurilingualism, the potential links between transnational friendship networks, cultural openess and plurilingual communicative practices. The pilot dataset provided preliminary evidence that these relationships could indeed be explored

quantitatively, but also revealed where the instrument required simplification to enhance analytical clarity.

V. Using the pilot to refine operational definitions.

The extended pilot was instrumental in showing which conceptual distinctions were meaningful to respondents, which items elicited high-quality, interpretable data, which scales produced measurement error, where cognitive burden hindered comprehension. For instance, items asking for 'months spent in the country of residence' were often skipped or misinterpreted, hence they were replaced with simpler items aligned with mobility definitions discussed in Chapter 3. Long Likert grids produced fatigue; they were reduced or restructured. Items on language ideologies were made more explicit or split into clearer sub-questions. Thus, the pilot served not only as a technical test of the platform but as an epistemic test, allowing theoretical constructs to be recalibrated through empirical response patterns. For example, Figure 48 shows how a question aimed at investigating attitudes towards ELF and plurilingualism was reshaped in the two versions of the questionnaire. Items that produced inconsistent or uninterpretable patterns during piloting were either reworded, relocated, or eliminated in the final survey.

How strongly do you agree or disagree with the following statements?

*(1 – strongly disagree, 2 – disagree, 3 – **neither agree nor disagree,** 4 – agree, 5 – strongly agree)*

	1	2	3	4	5
I like engaging with new languages, learning new words and ways of saying from my friends.	○	○	○	○	○
In my social life, I would prefer to speak more English than I do now .	○	○	○	○	○
In my social life, I would prefer to speak less English and more frequently other languages.	○	○	○	○	○
I would like to know English better than I do now.	○	○	○	○	○
In my social life, sometimes I feel stressed because I have to speak (a) language(s) that I don't feel totally comfortable with.	○	○	○	○	○
I would like to know the language of the place where I live better than I do now.	○	○	○	○	○

How strongly do you agree or disagree with the following statements?

*(1 – strongly disagree, 2 – disagree, 3 – **neither agree nor disagree,** 4 – agree, 5 – strongly agree)*

	1	2	3	4	5
I like engaging with new languages, learning new words and ways of saying from my friends.	○	○	○	○	○
I like learning the language of the places where I travel.	○	○	○	○	○
In my social life, I would prefer to speak less English and more frequently other languages.	○	○	○	○	○
I would like to know English better than I do now.	○	○	○	○	○

Figure 48. A comparison between a question in *Where's your language at? Extended Pilot* and *Where's your language at?*.

Before constructing the questionnaire, an analytical review of existing instruments in similar domains was carried out. This included surveys on digital nomadism (Flatio Reports 2023, 2024), transnational mobility surveys (Eurobarometer), ELF-related language surveys (including the ongoing speakENG project), multilingualism & education surveys (e.g. Erasmus impact studies). None of these instruments fully addressed the sociolinguistic dimensions relevant to FLO. Existing digital nomad surveys tended to focus on lifestyle, satisfaction, or economic data rather than language. By contrast, language-background surveys often lacked items addressing mobility, identity representation, transculturality, and technological mediation. This gap justified the development of a dedicated sociolinguistic instrument tailored to mobile remote workers.

The questionnaire comprises three blocks of questions, namely: a) personal information (14 questions); b) language background and education (11 questions); c) linguistic behaviours and attitudes (11 questions). The complete list of questions are presented in a .pdf file and an .lss file allows for consultation and reuse on LimeSurvey software; these two files can be downloaded from Zenodo (Tedesco 2025b), but the .pdf file is also reported here in Appendix 3.

Regarding technical implementation, as discussed in FLO's DMP (see Sub-section 4.2.), LimeSurvey was selected because it is GDPR-compliant, open-source, and hosted on the University of Bologna server.

Piloting the questionnaire

The extended pilot study represented a crucial methodological phase in the development of *Where's Your Language At?*. As discussed by van Teijlingen and Hundley (2002), piloting is an essential step in survey research, as it exposes issues of design, sequencing, conceptual consistency and data quality before full-scale implementation. While the results of the extended pilot study have already been discussed as part of the foregrounding research, here a general note on the different phases of testing and piloting is provided.

Concretely, the pilot followed a multi-step testing process consistent with recommendations by Dörnyei and Taguchi (2010) and Dillman (2014):

a) Expert review: Survey specialists and colleagues (researchers from the University of Bologna) were invited to evaluate the initial version. Their feedback focused on wording, alignment between theoretical constructs and operational indicators, redundancies or conceptual overlaps between items, appropriateness of scale formats, and potential sources of priming or sequence effects. This stage resulted in adjustments to definitions, simplification of item phrasing, and the addition of clearer instructions for multi-part questions.

b) Structured field pre-piloting: A small group of 8 participants belonging to the community under study completed the survey under observation and then participated in feedback interviews and a focus group. This phase revealed real-time issues regarding: survey length and cognitive fatigue, low engagement with long Likert grids, difficulty interpreting certain

behavioural indicators (e.g., "months spent in country of residence"), confusion caused by excessive branching, and problematic categories for nationality, languages learned at school, and mobility motivations.

Importantly, the pre-pilot confirmed that several constructs, especially those relating to language ideologies and plurilingual attitudes, required simpler operationalisation to avoid creating measurement error.

Then, the extended pilot was distributed and analysed as discussed in Section 4.3. Data quality checks – specifically, click count analysis, mono-answer detection, and scale distribution patterns – revealed several weaknesses. Items with extremely low selection rates indicated unclear or redundant response categories; long Likert grids encouraged patterned responding, increasing measurement error (Krosnick 1991). These findings were foundational in restructuring the final questionnaire.

Overall, the pilot survey provided first descriptive insights, with respondents showing high levels of mobility, advanced education, strong exposure to English, and active plurilingual practices. These patterns confirmed the relevance of the survey focus and the adequacy of the constructs derived from theory. It also provided methodological and technical feedback by highlighting several structural issues, like questionnaire's excessive length and several items cognitively demanding or poorly aligned with respondents' interpretive frames. As above discussed, the pilot also resulted in epistemic clarification of operationalisation, as it helped refine the connection between theory and measurement.

Survey administration

The final version of *Where's Your Language At?* was administered between 7 February and 5 June 2024 through LimeSurvey. Here the administration procedures, sampling strategy, and data-cleaning process that produced the final dataset are discussed, as well as the implications of such procedures for the validity, reliability and interpretability of the data.

As anticipated, the final survey adopted a convenience-based, theoretically oriented sampling strategy. Given the characteristics of the communities under investigation, probabilistic sampling was neither feasible nor desirable. Instead, the sampling strategy combined: convenience sampling, through open participation; time–place sampling, by promoting the survey in spaces where digital nomads tend to cluster (coworking spaces, colivings, community meet-ups); snowball sampling, as respondents shared the link through personal networks. The underlying rationale was not to approximate a statistically representative sample of *all digital nomads*, but to collect diverse and information-rich profiles of mobile remote workers. This aligns with the theoretical sampling logic often adopted in exploratory sociolinguistic studies (Schilling 2013), where the goal is analytical depth and construct exploration rather than population estimation.

The survey was promoted through:

- the project webpage *The Nomad Linguist* with a dedicated page[22];
- Instagram posts and stories targeted at digital nomads;
- dedicated WhatsApp and Slack communities of several European digital nomad hotspots;
- on-site dissemination through the use of flyers and posters in coworking and community hubs (as shown in Figure 49).

Figure 49. Flyers advertising *Where's your language at?* at a coworking space in Los Lajares and a meet-up event in Corralejo (Fuerteventura).

During the administration period, the survey received 111 total entries, and 76 usable responses, after removal of blanks and technically incomplete submissions. This represents a substantial improvement compared with the pilot (58 entries, 24 usable), confirming that the structural and technical revisions implemented in the final instrument increased both response quality and completion. Web-survey literature confirms that format, layout, and item clarity significantly shape completion rates (Couper 2008; Dillman 2014), particularly in populations completing surveys on mobile devices. Indeed, completion improved as a result of shorter survey length, improved mobile compatibility and better structured question grouping.

Once the survey was closed, all data were exported in both .csv and .xlsx formats, using semicolon (";") as breaking character and quotation marks ("") as text delimiters. The cleaning

[22] https://www.thenomadlinguist.eu/wheres-your-language-at/.

process followed standard web-survey processing procedures: a) Removal of empty responses; b) Removal of technical duplicates; c) Exclusion of submissions with critical missing values – responses missing all key behavioural indicators or only containing very initial respones were omitted.

The .xlsx version includes a yellow highlight for responses not representative of the population under study. Unlike the pilot, representation was not determined through a seven-criteria operationalisation, but was based on three streamlined behavioural indicators: travelling frequently, living in multiple places within a year, and working remotely. This classification simplifies respondent profiling without reintroducing interpretative subjectivity. The final dataset therefore reflects both quality control and theoretical refinement**.**

All procedures complied with GDPR and University of Bologna ethical standards. Respondents were informed about the voluntary nature of participation, the possibility to skip questions, the anonymisation of data before storage and publication, and the intended scientific use of the dataset. No direct personal identifiers (name, email) are included in the dataset. The dataset, together with the questionnaire (PDF and .lss formats) and metadata, is stored in the Zenodo repository (Tedesco 2025b) for open-access consultation and potential replication.

Description of responses

What follows is a general description of the dataset deriving from the survey *Where's your language at?*. These results are then discussed with relation to specific research questions in Chapter 6.

First, an overview of the respondents' socio-demographic characteristics is provided in Table 7, based on the 76 usable responses collected in the final survey (February – June 2024).

Table 7. Demographic information on respondents to the survey *Where's your language at?*.

Gender	
Female	51%
Male	45%
Non-binary	3%
Prefer not to say	1%
Country of birth (macro-area)	
Europe	84%
North America	7%
South America	4%
Asia	1%
Africa	1%
Central America	1%
Level of education	
Master's degree	49%
Bachelor's degree	40%

Ph.D.	8%
High school	4%
Field of education	
Business, administration and law	20%
ICT & Engineering	17%
Arts & Humaniteis (incl. languages)	17%
Natural sciences / Statistics	7%
Health & Welfare	5%
Services (Tourism, Transport)	4%
Agriculture / Veterinary	1%
Other	29%
Field of work	
Marketing	20%
ICT Services	17%
Education	15%
Finance & Insurance (incl. Crypto)	8%
Accommodation & Food	7%
Arts & Creative Services	4%
Sale & Retail	4%
Health & Social care	4%
Professional services (non-marketing)	3%
Public sector	1%
Other	18%
Remote work	
100% remote	51%
About 75%	16%
About 50%	17%
About 25% or less	16%
Intense travel	
Yes	74%
No	14%
N/A	12%
Working while travelling	
Yes	74%
No	14%
N/A	12%

Respondents span a wide age range, though the majority fall between 25 and 40 years old, consistent with international reports on digital nomad demographics (Flatio 2023, 2024). The median age is expected to be in the early 30s, aligning with the pilot findings (median = 33). This concentration reflects both the socio-economic characteristics of mobile remote work and the high mobility typical of early-to-mid career professionals. From a sociolinguistic perspective, this age profile is relevant because individuals in this bracket typically display

advanced educational attainment, high exposure to English in study/work trajectories, established international mobility networks, and strong adaptability to plurilingual environments. Indeed, these characteristics may entertain a relationship with linguistic behaviours, particularly the frequent use of English as a lingua franca and the adoption of flexible multilingual strategies (Kinginger 2013; Murphy-Lejeune 2002; Doiz et al. 2014).

Gender distribution is balanced overall, with a slight prevalence of female respondents – similar to patterns observed in mobility-focused educational surveys (Erasmus Impact Study 2014). The inclusion of a 'non-binary' option reflects the inclusive orientation of this project and aligns with recommendations in contemporary sociolinguistic survey literature (Garrett 2010). Although gender is not used analytically as an explanatory variable in this thesis, documenting gender composition helps to describe the population and understand social positioning.

One of the most distinctive features of the dataset is its high level of transnational diversity. Respondents come from a broad set of nationalities, largely – but not exclusively – European. Countries such as Italy, Germany, France, Portugal, Spain, the Netherlands and the UK appear prominently, alongside participants from Eastern Europe and few non-European regions.

Moreover, location where responses were collected, shown in Figure 50, is consistent with digital nomad mobility trends represented in the literature (see Chapter 2) and in the results from the piloting phase borders.

Figure 50. IP location of responses in *Where's your language at?*.

Consistent with results from the pilot and with literature on digital nomad profiles (Cook 2023), respondents overwhelmingly display high educational attainment, with most holding a Bachelor's or Master's degree, and several holding a PhD. Fields of education mirror labour-market segments commonly associated with remote work and transnational careers, such as ICT and digital services, Marketing and communication. A smaller number of respondents come from STEM or Health & Welfare backgrounds. The educational distribution is significant because it may reflect higher exposure to English during study, and shape attitudes towards mobility and transcultural collaboration. Seemingly, the occupational profiles of respondents

strongly align with sectors known to sustain remote work. The most represented categories correspond to ICT services, marketing, digital professional services, (online) education, and creative industries. These findings resonate with the main industries reported in digital nomad surveys (Flatio 2023; 2024) and justify the strong presence of English in both professional and social settings, which is analysed in Section 6.2. A crucial behavioural indicator concerns remote work frequency. In the final dataset, a large proportion of respondents report working 75% or more of the time remotely, with many indicating 100% remote work.

Respondents display high levels of mobility and tend to mix travel and work. These indicators align with the theoretical definition of mobility and digital nomadism elaborated in Chapters 2 and 3 and provide a robust empirical foundation for analysing communicative practices.

Overall, the demographic profiles emerging from the final survey confirm high educational and linguistic capital and reveal substantial geographical mobility. Hence, results align with known socio-economic profiles of digital nomads, thus validating the sampling strategy outlined in the above paragraphs. These demographic data provide the contextual basis for interpreting linguistic behaviours and attitudes.

The second section of the questionnaire was aimed at exploring respondents' linguistic backgrounds, self-reported repertoires, language learning trajectories, and the role of English in their education and work practices. The questionnaire includes eleven items dedicated to linguistic background (Section B in Where's Your Language At?), covering: (i) self-assessed bilingualism/plurilingualism, (ii) languages used in daily life, (iii) contexts of acquisition, and (iv) the role of English in study and work. Together, these variables operationalise core constructs discussed in Chapter 3 – particularly plurilingual repertoires, mobility-driven language socialisation, and linguistic trajectories – within the constraints of a large-scale survey.

82% of the participants self-identify as bilingual or plurilingual. This aligns with the pilot survey and with ethnographic observations within the FLO project, where respondents regularly draw on two or more languages in social, professional, and everyday communicative practices. Self-assessment, while subjective, remains a meaningful proxy for perceived linguistic agency and mobility (Dörnyei & Taguchi 2010; Busch 2017). The dataset includes a numerical self-report (B4): 'How many languages can you speak?', with 41% of bilingual respondents, 59% of respondents stating to be able to speak 3 or more languages. Moreover, participants were asked to list the languages they use in everyday life. The distribution (as reported in Figure 51) shows a clear dominance of English as a lingua franca, followed by other European languages.

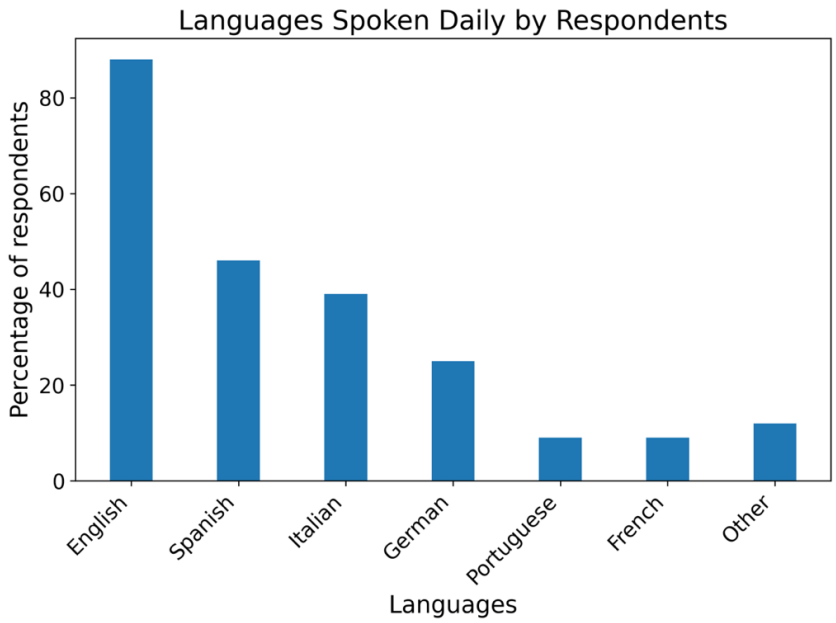


Figure 51. Languages spoken daily by respondents in *Where's your language at?*.

English emerges overwhelmingly as the primary communicative resource across contexts while local European languages reflect participants' mobility trajectories.

Participants were asked where they learned the languages they speak (B5–B6). The distinction between formal (school/university) and informal (travel, self-study, immersion in communities) learning reveals hybrid learning paths, as shown in Figure 52.

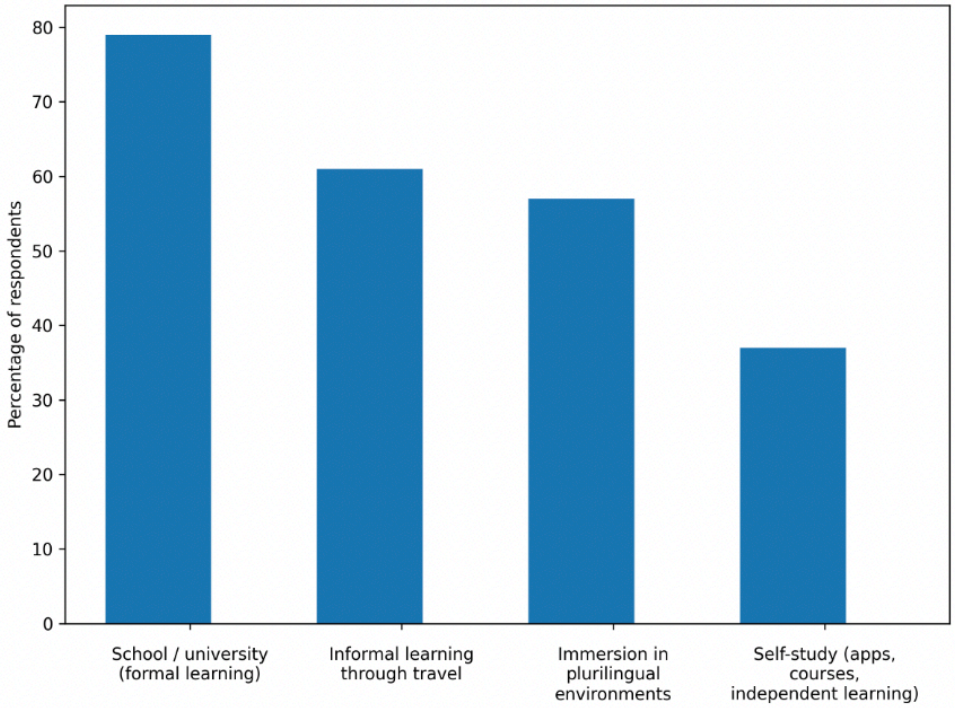


Figure 52. Language learning modalities in *Where's your language at?*.

The percentages reported in the Figure refer to respondents selecting the option, thus are non-exclusive. In general, these results show a steady interest towards language learning and reflect positive attitudes towards plurilingualism as shown by results from the third section of the questionnaire. Moreover, the overlap of formal and informal acquisition is characteristic of *plurilingual development* as conceptualised in the Council of Europe (2018), and aligns with theories of language socialisation in mobility (Wei 2018).

Results from three questions investigating the role of English in participants' life show that English is central in both educational and professional domains, with 46% participants stating to have received their full education in English, 71% to have worked in English, and 64% to currently use English in their professional life.

The survey also asked about the nationality of close friends (B10), as a proxy for diversity of social networks, with the majority of respondents stating to mostly have friends of different nationalities from their own. This confirms the translocal and transcultural nature of respondents' social environments.

Finally, participants were asked whether they were satisfied with the language education received during school/university. 39% of respondents declared they are dissatisfied with their language education. Compared to the pilot study, a smoothing effect is observed: in the pilot, participants expressed a clear contrast between relative satisfaction with English-language education and dissatisfaction with education in other languages. In the expanded sample, this contrast becomes less polarised, a pattern that is typical when increasing the number and heterogeneity of participants, as individual extremes tend to converge toward more moderate aggregate tendencies (De Vaus 2013; Bryman 2016). Yet, despite this attenuation, a substantial level of dissatisfaction persists. Linked to findings from the third section of the questionnaire, this result suggests that linguistic offer in formal educational settings may not entirely meet the needs of mobile subjects like those represented in the survey.

The final section of the survey examines how respondents orient themselves towards English and other languages in their daily practices, and how they position themselves ideologically within translingual environments. These variables operationalise core constructs discussed in Chapter 3, particularly language ideologies, English as a lingua franca (ELF), translanguaging, and plurilingualism. Overall, the data portray a highly mobile and linguistically flexible population, where English functions as a shared communicative resource but not as a normative or identity-imposing force. Instead, participants show strong acceptance of multilingual diversity, and report communication styles characterised by mixing, adaptability, and collaborative linguistic negotiation.

Consistent with results from the second section, respondents report using English across several domains, although the intensity varies. English is used most extensively online and at work, with a slightly lower (yet still considerable) presence in daily-life and social interactions. Figure 53 summarises these patterns.

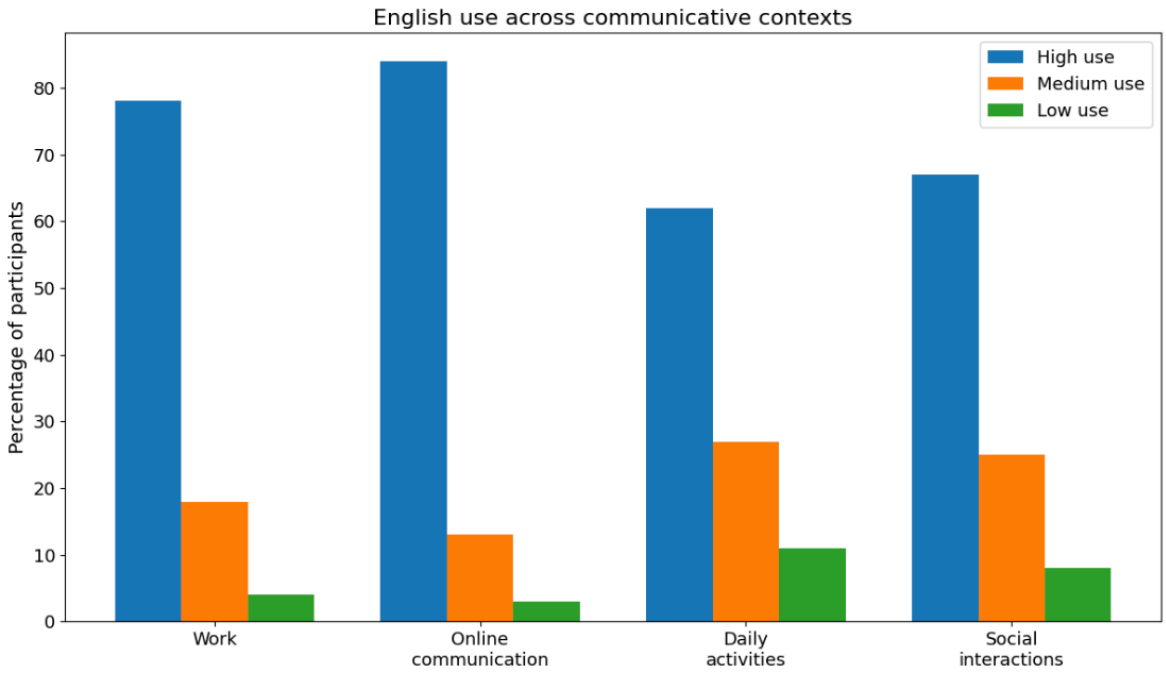


Figure 53. Frequency of English use across domains as represented in *Where's your language at?*.

These figures show that English is predominant in all aspects of life, including social interactions. Thus, they confirm what emerges from observation data, that English generally serves as the connecting language in digital nomad communities. However, most respondents state that they mix languages spontaneously (71%) and that doing so feels natural (76%). The survey responses reflect what the corpus reveals in practice: translanguaging appears not (only) as a problem-solving strategy but as a daily communicative choice shaped by mobility, identity, and habit (as will be discussed in Chapter 5 and 6).

While respondents generally feel comfortable using English, a more complex picture emerges when considering linguistic insecurity. Nearly half of the respondents expresses occasional doubts about the adequacy of their English, even though they perceive themselves as effective communicators. Attitudes towards language norms reveal widespread openness to linguistic diversity.

As shown in Figure 54, over half of the respondents reject the idea that English must be spoken 'properly'. Further survey results show that an overwhelming majority of respondents feels comfortable (94%) views knowing multiple languages as an advantage. Opinions differ, however, regarding adaptation to the local language: while many respondents see this as desirable, others emphasise practicality or the transient nature of their stays. This ideological complexity mirrors ethnographic evidence: digital nomads value multilingualism and mobility, yet navigate contexts where integration is uneven or short-term.

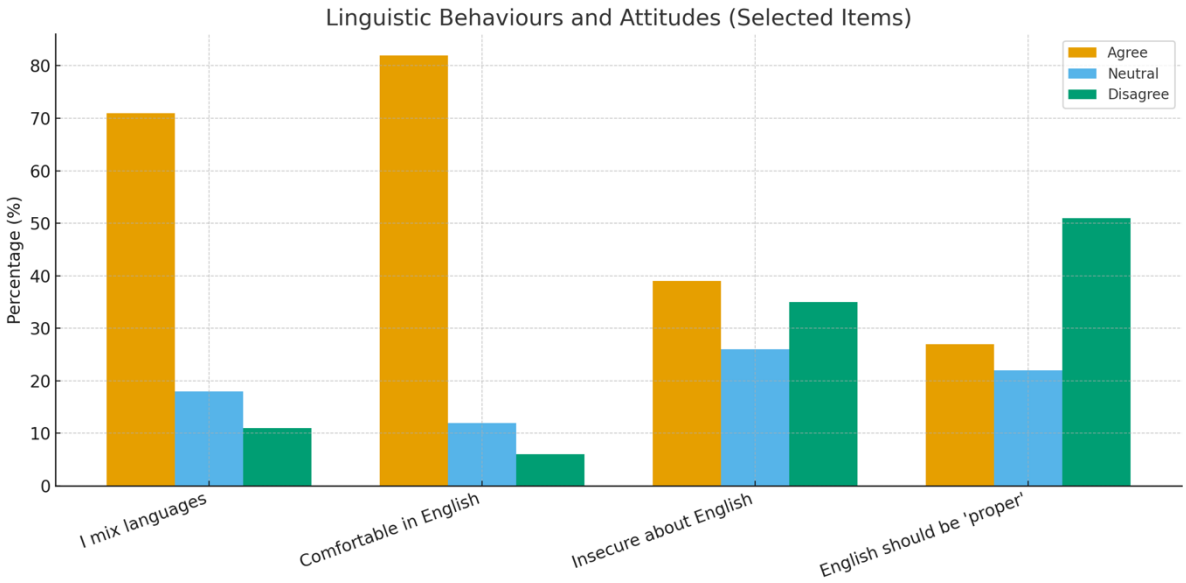


Figure 54. Selected linguistic behaviours and attitudes of participants in *Where's your language at?*.

Taken together, these findings depict respondents as linguistically agile individuals who navigate English and other languages with confidence, flexibility and creativity. Their attitudes align with plurilingual frameworks (Council of Europe 2018), and their perceived communicative practices meet the representation of interactions in the NomadLingo corpus (which will be described in greater detail in Chapter 5). Yet, the presence of linguistic insecurity – despite high proficiency and extensive use – highlights the influence of *pure* English ideologies (Milroy & Milroy 1985), which shape how speakers perceive their own communicative competence.

4.4.4 Data-intensive linguistics: Corpus methods and spoken corpora

Besides survey and ethnographic immersion, the core dataset of this doctoral research is represented by NomadLingo, a corpus of transcripts of translingual spoken interactions recorded at the digital nomad communities based in Madeira, Porto Santo, Fuerteventura and Gran Canaria. Before delving into the discussion of specific procedures for data collection and analysis (in Chapter 5), this section provides a general overview on corpus linguistics methodologies, particularly focused on spoken corpora.

Over the past three decades, linguistics has undergone a profound methodological and epistemological shift, often captured under the umbrella term of *data-intensive linguistics* (Abney 2011; Borin 2011; Hammarström 2016). This orientation is not merely about scaling up the amount of data available: it reconfigures what counts as evidence, how linguistic knowledge is generated, and how researchers position themselves within the communities they study. Corpus linguistics, in particular, has been central to this transformation. Once tied primarily to lexicography and grammar description, it has expanded into a comprehensive methodological landscape that integrates multimodal analysis, interactional data,

computational tools, and increasingly, ethnographically grounded approaches (McEnery & Hardie 2011).

Corpus linguistics today is characterised by its dual nature as both a methodology and a theoretical framework. While early manuals described corpus linguistics as a set of computational techniques for processing large datasets, subsequent works have emphasised its epistemological stance, namely that linguistic structures emerge from recurrent patterns in usage (Sinclair 1991; Bybee 2010; Goldberg 2006). This is particularly evident in usage-based and constructionist approaches which argue that grammatical knowledge is shaped through frequency, distribution, and entrenchment.

A key distinction in corpus linguistics is that between *corpus-based* and *corpus-driven* research (Tognini-Bonelli 2001). Corpus-based studies typically begin with theoretical assumptions and use corpora to test hypotheses, whereas corpus-driven research seeks to derive categories inductively from the data. Contemporary studies often blend the two, balancing inductive pattern identification with theoretically informed interpretation (Bybee 2006; Huston & Francis 2000).

Representativeness, once conceived in strictly statistical terms, is now understood more flexibly: especially in specialised or small corpora, transparency of sampling, consistency of annotation, and alignment with research aims are more important than numerical balance (Flowerdew 2004). This perspective is particularly relevant in sociolinguistics and applied linguistics, where the focus often lies on specific communities or communicative practices (Koester 2010).

Within corpus linguistics studies, spoken corpora allow researchers to investigate features of spoken discourse that written corpora cannot capture. Spoken language is inherently ephemeral, interactional, and multimodal. It is full of hesitations, repairs, fillers, discourse markers, phonological reduction, overlapping talk, and indexical expressions (Carter & McCarthy 2006; Adolphs 2008). Spoken language is not neat; it is emergent, embodied and contingent upon context. Building spoken corpora is challenging. Early projects such as the Santa Barbara Corpus (Du Bois et al. 1993) and the British National Corpus spoken component confronted the issue that transcription is always a representation of an event rather than its reconstruction (Ochs 1979).

Recording conditions also matter. Adolphs (2008) emphasises that spoken corpus builders must navigate environmental noise, microphone placement, equipment portability, and participants' awareness of being recorded. These challenges increase when research is carried out in mobile, transnational environments such as those inhabited by digital nomads.

Modern tools make the building and processing of spoken corpora possible. ELAN, developed at the Max Planck Institute, supports multi-tier, time-aligned annotation of speech, gesture, and multimodal phenomena (Wittenburg et al. 2006). EXMARaLDA excels at segmenting

complex interactions and enabling searchable formats (Schmidt 2014). PRAAT allows detailed phonetic and prosodic analysis (Boersma & Weenink 2024).

Automatic Speech Recognition (ASR) tools are improving, but multilingual, noisy interactions remain difficult for them (Xiong et al. 2016; Liang et al. 2023). Whisper (Radford et al. 2022) performs impressively in clean conditions but cannot handle overlaps, translanguaging, or nuanced prosodic phenomena without extensive correction.

Recent methodological discussions at major Corpus Linguistics Conferences like TALC 2024 and CL 2025 include attention to the ethnographic and qualitative dimension in corpus linguistics. Gavin Brookes, for example, emphasised that corpora must be 'put in context' (2025), echoing earlier work with Baker (Brookes & Baker 2021) and the linguistic ethnography tradition (Rampton et al. 2015; Copland & Creese 2015). In this perspective, corpora are not neutral linguistic objects but socially situated artefacts whose interpretation requires understanding the contexts and practices from which they emerge.

Ethnographic perspectives contribute to corpus linguistics in several ways. First, ethnography contextualises linguistic patterns. Frequency alone cannot explain why particular linguistic choices emerge; understanding requires knowledge of social relations, identities, practices, registers, and ideologies (Blommaert 2010; Bucholtz & Hall 2005). Second, ethnography informs sampling decisions. Field immersion helps researchers identify relevant communicative ecologies and situational variables (Androutsopoulos 2008). Third, ethnography enriches metadata. While traditional corpora include speaker age, gender, and nationality, ethnographically informed corpora incorporate relational roles, setting descriptions, emotional tone, mobility trajectories, and repertoire information aligned with superdiversity studies (Baker & Egbert 2019). Fourth, researcher positionality matters. As alreeady discussed my position within the community and my identity branded as *The Nomad Linguist* facilitated rapport-building, access to communities, and recruitment. Fifth, ethnography is essential for annotation. As Tusting has argued, interpretative categories – stance, irony, affect, translanguaging – cannot be reliably annotated without ethnographic insight (2019). Tools alone cannot identify such phenomena; researchers must understand community norms and interactional histories.

The integration of corpus linguistics, ethnography, and interactional approaches reflect a broader shift toward triangulated methodologies. As Pennycook (2010) and Canagarajah (2018) argue, mobility and plurilingualism require methods that capture dynamic, situated practices rather than static categories. For digital nomads, this triangulation is essential. Their communicative practices are shaped by mobility, fluid repertoires, and shifting social networks, and spoken interaction is where these dynamics are negotiated in real time

The methodological principles discussed here have been foundational for the creation of the NomadLingo corpus. The next chapter details its construction: data collection procedures, recording environments, annotation schemes, XML encoding, metadata design, and analytical possibilities.

Chapter 5

Building the NomadLingo corpus

5.1 Introduction to Chapter 5

Understanding spoken interaction requires an approach that goes beyond treating language as an object that can be simply represented in one dimension, extracted from its context. Corpus construction, in this sense, is not a mechanical step preceding analysis, but rather part of the interpretive work. The chapter presents the methodological steps required to transform naturally occurring, socially embedded interactions into a structured resource while preserving as much as possible the ecological complexity of the original communicative events.

It is organised in four main sections. Section 5.2 describes the approach and methods for data collection. Section 5.3 turns to metadata and contextual annotation, showing how each recording was documented, and how contextual information was organised. Section 5.4 focuses on processing the raw data, detailing the semi-automatic transcription pipeline, the iterative revision workflow, and the reasoning behind the tailored XML architecture. Section 5.5 introduces the linguistic annotation scheme, centred on three domains: translanguaging phenomena, misunderstanding sequences, and repair strategies. Each annotation layer captures how participants negotiate meaning, manage problems of understanding, and mobilise their repertoires dynamically across languages. Finally, Section 5.6 describes corpus formats and the web-based implementation on NoSketchEngine.

The chapter foregrounds methodological transparency and reflexivity on the assumption that the way we capture discourse conditions how it can be interpreted. NomadLingo is thus presented not only as a dataset but as a methodological framework for studying linguistic fluidity in informal, transcultural interaction.

5.2 Approach and methods for natural spoken data collection

The present section outlines the methodological approach adopted for collecting the naturally occurring spoken data for the NomadLingo corpus. Building on the principles of linguistic ethnography and spoken corpus linguistics introduced in Chapter 4 (Rampton 2007; Tusting 2020), 5.1 first clarifies the theoretical and ethical foundations underlying the collection of spontaneous interactions among digital nomads at the Madeira and Canary Islands communities. It then describes the tools and procedures adopted, followed by a detailed description of data collection contexts. Finally, an overview of the raw dataset is provided, introducing its scope, composition and main methodological challenges.

5.2.1 Naturally occurring data and consequences of ethnographic participation

As argued in s 4.2 and 4.3, understanding linguistic fluidity in contexts shaped by mobility, plurilingualism and digital mediation requires one to capture the situated, emergent and interactionally negotiated nature of communication (Blommaert 2010; Pennycook 2010; Blommaert & Rampton 2011). Spoken data are therefore indispensable, not only because they reveal translingual practices as they unfold in real time, but also because they provide access to the interactional, affective and socio-cultural dimensions that accompany linguistic action.

In this project, spoken data were collected as naturally occurring data, understood as interactions that take place independently of any experimental elicitation (Biber 1993a; Adolphs & Knight 2020). This approach has been widely adopted in variationist sociolinguistics and corpus research, where naturalistic recordings are considered crucial for analysing everyday communicative behaviour. As outlined in Chapters 2 and 3, digital nomad communities are characterised by superdiversity (Vertovec 2007), high mobility, and diverse individual as well as collective repertoires shaped by previous migration trajectories, online communicative practices and transcultural experiences (Blommaert 2010; Androutsopoulos 2015). These features make them an ideal laboratory for examining the dynamics of linguistic fluidity, translanguaging, interactional negotiation, and informal mediation, all central to the analytical aims of the FLO project. In digital nomad communities, where participation is a prerequisite for access, the data capture naturally occurring interactions within authentic social relationships, shaped by informal and contextually rich communicative settings.[23]

The researcher's position played a fundamental role. As discussed in Chapter 1 and 4, this project involved extended insider participation in community life. Because I was already part of the community when the project started, the classical observer's paradox (Labov 1972) was partially inverted: my presence did not inhibit natural behaviour but was perceived as routine and socially unmarked. Moreover, this insider stance facilitated access to private and semi-private events (coworking sessions, hikes, community dinners, bar meetups, hostel gatherings), which are often inaccessible to external researchers. It also meant that I was not merely a silent recorder but an active participant; as a result, my voice appears in several interactions. This presence is not seen as diminishing the authenticity or analytical value of the data. On the contrary, it reflects the co-constructed nature of the dataset and is consistent with participatory ethnography's views on the researcher as an active participant rather than a neutral observer (Copland & Creese 2015; Rampton et al. 2015).

As already extensively remarked in Chapter 4, ethical considerations were central throughout.

5.2.2 General principles and tools for data collection

Data collection followed an emergent design (Hammersley & Atkinson 2019), in line with grounded ethnographic practices (Tusting 2020). This meant that data collection sessions were embedded within pre-existing community events, such as dinners, hikes, coworking sessions

[23] As will be explained in Sub-sections 5.1.3, 5.1.4 and 5.1.5, events where the NomadLingo data were collected were characterised by different degree of openness and in a few cases, some setting features, like participants disposition and room setting, were arranged by the researcher. However, the interactions taking place in these environments were by no means controlled by the researcher.

or informal meetups, or co-organised with community members and managers when opportunities arose. As is typical of digital nomad gatherings, the degree of acquaintance among participants varied considerably: some were close friends, others had met only minutes before the recording began. Crucially, none of these events was organised for the purpose of recording; the primary objective was always the social activity itself, hence participants' enjoyment and the smooth unfolding of the event were prioritised. As a result, audio quality could not be treated as the number one priority during data collection, and limitations in sound capture – related to noise, movement, or environmental conditions (as shown in Table 11) – were accepted as intrinsic to the ecology of the setting. These limitations, as observed during the first sessions of data collection in Madeira, later informed iterative refinement of research tools and protocols.

My voice is present across all recordings, although to varying degrees depending on the interactional configuration. In some sessions, my participation was peripheral – limited to brief turns at the beginning or end – while in others, especially dinners hosted in the house where I lived, taking part in the conversation was unavoidable due to continuous engagement from other participants. Moreover, it should be noticed that one participant in particular, IT_MC_01, travelled with me during part of the fieldwork and was both a member of the community and my partner at the time. His presence is therefore recurrent across the dataset.

Because digital nomad interactions often took place outdoors, in bars, in coworking spaces, on beaches, or along hiking trails, the project relied on lightweight, unobtrusive recording tools (i.e., smartphones equipped with high-quality apps such as MotivAudio, which ensures reliable audio even in acoustic environments characterised by background noise . In several cases, participants contributed additional recordings from their own devices.

In Table 8, a summary of recording procedures is provided.

Table 8. Recording procedures in data collection sessions for the NomadLingo corpus.

ID	Date	Location	Location type	Event type	No. of devices	No. of groups	No. of ppl.	Audio challenges	Audio quality
M1	23/10/23	Funchal, Madeira	Outdoor (park)	Picnic	2	1	7	Participants moving around, overlapping, background noise from the port, weather conditions (wind and rain)	Medium to high
M2	01/11/23	Camara do Lobos, Madeira	Indoor (private house)	Birthday party	3	1	8	Participants moving around, overlapping conversations, fireworks in the background.	Medium to poor
M3	10/11/23	Porto Santo, Madeira	Indoor (restaurant) + outdoor (street)	Coworking lunch + walk around the old town	1	1	7	Background noise at the restaurant, device not always	Medium

								close enough to speakers.	
M4	12/11/23	Porto Santo, Madeira	Outdoor (hike)	Community hike	1	1	14	The device was held by one participant in his pocket during activity: there are many overlapping conversations, background noise and long windows with no speech.	Poor
M5	12/11/23	Porto Santo, Madeira	Outdoor (restaurant terrace)	Community lunch	1	1	15	Background noise due to weather conditions and the proximity to the sea; many overlapping voices.	Poor
M6	21/11/23	Funchal, Madeira	Indoor (restaurant)	Community lunch	1	1	5	The restaurant was very crowded and the community lunch itself counted with 20 participants: the recording features intense background noise and continuous voice overlaps.	Poor
M7	24/11/23	Ponta do Sol, Madeira	Outdoor (coworking terrace)	Coworking happy-hour	3	3	12	Background noise caused by church bells ringing and traffic; participants moving around.	Medium
M8	03/12/23	Ponta do Sol, Madeira	Outdoor (coworkign terrace)	Coworking happy-hour	2	2	11	People moving around.	Medium
GC 1	09/02/24	Las Palmas de Gran Canaria, Gran Canaria	Outdoor (coliving terrace)	Coliving dinner	2	2	13	People moving around and doing a variety of activities (like playing pool, preparing drinks).	Medium to high
GC 2	15/02/24	Las Palmas de Gran Canaria, Gran Canaria	Outdoor (beach)	Community happy-hour	2	2	7	Background noise caused by people passing by, weather conditions and the proximity to the sea.	Medium
GC 3	23/02/24	Las Palmas de Gran Canaria, Gran Canaria	Indoor (coliving)	Community dinner	2	1	8	Voices overlapping; participants arriving and leaving at different times.	Medium
F1	05/04/24	Corralejo, Fuerteventura	Indoor (private house)	Private Dinner	2	1	8	Voices overlapping; participants arriving and leaving at different times.	Poor

F2	08/04/24	Corralejo, Fuerteventura	Indoor (coworking kitchen)	Community event	2	1	5	Voices overlapping and background noise from the coworking.	Medium

As will be explained in 5.3, not all recorded data were used for the construction of the first version of the corpus, for which recordings characterised by a higher quality and more stable conversation groups were preferred. The decision on the recordings to include in the first version of the corpus depended also on other factors, such as representativeness and balance (in terms of locations, type of events and participants' demographics). The data which to date are not included in the corpus are still part of the raw dataset and can be added to future versions of the corpus. In the analyses presented in Chapter 6, some of the raw data were used to back up findings from the corpus.

During each of the sessions listed in the table contextual information was noted in a file which included the following sections:

- Name of the event;
- Type of event (e.g., happy-hour at the coworking, a social event organised by the community manager and the researcher);
- Description (including information about how participants were reached, how they were located in the space, how groups were created, where recording devices were placed etc.; pictures were included when available);
- Number of researchers who took part in the event;
- Number of participants;
- Number of recordings;
- Information about speakers for each recording (including names, nationalities and languages);
- A table containing speakers' pseudonyms and codes (see Sub-section 5.3.2).

An example of this annotation document is provided in Appendix 4.

At the end of each data collection session, a folder was created containing audio recordings in .wav format, the annotation document, signed informed consent and privacy documentation for each participant.

5.2.3. Data collection in Madeira Islands

Madeira Islands represented the first region where data were collected between October and December 2023 After an initial period of observation, a general protocol was developed for data collection.

Step 1: Organising a data collection session

This step followed three main pathways: (1) participating in an existing event and asking participants whether they would take part in the recording (as in the case of M3, M4, M5, and M6); (2) co-organising an event with community managers (as in the case of M1, M7, and M8); or (3) asking community members in advance whether it would be possible to record interaction at an already planned event (as in the case of M2).

Documentation was then prepared: the annotation document and the paper documentation for each participant, including consent forms, privacy statement and a questionnaire concerning participant cultural and linguistic background. Various strategies were used to attract participants (see x.x below).

Step 2: Collecting informed consent and information about participants

At the beginning of each session, paper documentation was distributed along with a background information questionnaire, and participants were informed about the study. In Figure 55, participants completing the booklet during the second data collection event in Madeira are shown.

Figure 55. A picture from data collection in Madeira: participants completing the booklet with consent and privacy forms, and background questionnaire.

Step 3: Setting up the recording

Participants were asked to self-record when possible; MotivAudio was set up on each device, and recording devices were placed around the interactional setting. For example, Figure 56 shows how tables were located during the Happy Hour event at the Coworking in Ponta do Sol.

Figure 56. A picture from a data collection session at the Coworking in Ponta do Sol, Madeira.

Step 4: Storing recordings

Recordings were deleted from participants' devices and stored first on the researcher's device and then, as soon as possible, on a laptop in an encrypted folder and on the Bologna University cloud.

Step 5: Finalising the Session Folder

Privacy documentation was scanned, and the circumstantial annotation document was completed with additional information (ideally immediately after the event, with notes taken during the session). To do this, it was necessary to listen to the recordings in order to assign pseudonyms and participant codes accurately.

The procedure here outlined resulted in some drawbacks: spontaneous recordings like those produced in Porto Santo were of lower quality (see Table 10); collecting consent and background information was time-consuming and challenging. For these reasons, the protocol was updated for the Canary Islands recordings (see Sub-section 5.1.4).

Table 9 provides general information about the data collected in Madeira[24].

[24] The sessions in Porto Santo are not considered in this table due to audio quality. For this reason, they were disregarded for the creation of the corpus; however, they still provide valuable qualitative insights into sociolinguistic practices of digital nomad communities and a description of the data collection sessions in Porto Santo is provided later on in this section.

Table 9. General information about data collected in Madeira between October and December 2023.

No. of sessions	5
Total length of the recorded audio	11h47min
Average length of the recorded audio x session	1h41min
Total No. of participants	37
Total No. of nationalities represented	18
Median no. of participants x session	8 (St. Dev. 2.88)
Median no. of languages spoken by participants	3 (St. Dev. 1.04)
Five most spoken languages by participants	English (37); Spanish (16); Portuguese (14); German (12); Italian (8).

For the purpose of calculating the overall duration of recorded interactions, only one .wav file was retained in cases where two devices captured substantially identical audio. By contrast, when multiple devices captured distinct conversational events or separate conversational groups – as indicated in Table 10 under *Number of Groups* – the durations of the corresponding recordings were aggregated. Further detail about the way overlapping recordings were used is provided in 5.3, where the transcription pipeline is presented.

It is also worth noting that the relative stability of the Madeira community enabled extended observation cycles and repeated interactions among the same speakers across different events, offering valuable opportunities to trace how communicative norms and linguistic repertoires evolve through sustained contact and how they are negotiated depending on the context (i.e., type of event and other speakers participating in the conversation). A similar pattern was observed in the Fuerteventura community. In contrast, the Gran Canaria community appeared – based on the available recordings – to exhibit higher participant turnover. This insight should, however, be treated with caution, as it may reflect differences in the types of events recorded across locations rather than characteristics of the communities themselves.

In Table 10, a summary of the setting, participants and context of each data recording session in Madeira, their inclusion or exclusion from the first version of the NomadLingo corpus is presented. Detailed information and pictures about data collection are included in Appendix 5 (more information about participants can be found in Appendix 7).

Table 10. Madeiran data collection sessions: An overview.

Session ID	**Location**	**Context type**	**No. participants**	**Participants nationalities**	**Included in NomadLingo (v1)?**
M1	Funchal	Public	6 + researcher	American, Italian, Romanian, German, Canadian, Iranian	YES

M2	Câmara de Lobos	Private	7 + researcher	Italian, Romanian, Portuguese, British, Taiwanese	YES
M3	Porto Santo	Public	6 + researcher	British, Portuguese, Lebanese, Polish, Italian, Romanian	NO
M4	Porto Santo	Public	14 + researcher	British, Portuguese, Lebanese, Polish, Romanian, Estonian, French, Italian	NO
M5	Porto Santo	Public	15 + researcher	British, Portuguese, Lebanese, Polish, Romanian, Estonian, French, Italian	NO
M6	Funchal	Public	5 + researcher	Russian, Dutch-American, German, Russian, Italian	NO
M7	Ponta do Sol	Controlled environment	12 + researcher	Portuguese, Italian, German, Norwegian, Lebanese, Serbian, Polish, Estonian, Brazilian	YES
M8	Ponta do Sol	Controlled environment	11 + researcher	Italian, Estonian, Portuguese, Romanian, German, and Hungarian	NO

5.2.4. Data collection in Canary Islands

A second major field site was Las Palmas de Gran Canaria, where recordings were collected in February 2024. Here, the community was observed to be more fluid than in Madeira, with frequent arrivals and departures and strong day-to-day variation in group composition. Recordings took place in colivings and at social open-air events.

A third one was Fuerteventura, where the researcher moved in March 2024. Fuerteventura features a long-standing tradition as a digital nomad hub. The most popular locations for digital nomads on the island are La Pared in the south, Los Lajares, El Cotillo and Corralejo in the north. Corralejo was chosen for data collection because it seemed to be the most active digital nomad hub at the time, with social meet-ups, coworking events, dinners, outdoor activities organised weekly. Moreover, Corralejo is the place where the project Remote Work Community was born, one of the first and most popular networks of digital nomads (see Chapters 2 and 4).

Following considerations about issues related to the general protocol for data collection in Madeira, in Canary Islands the protocol was modified in two crucial respects:

- Consent and privacy documents were provided in digital format;
- Background information about spoken languages, nationality and travel trajectories of speakers was collected through peer interviews at the beginning of each data collection session. A prompt flyer in both English and Spanish was given to the group of participants, as shown in Figure 57.

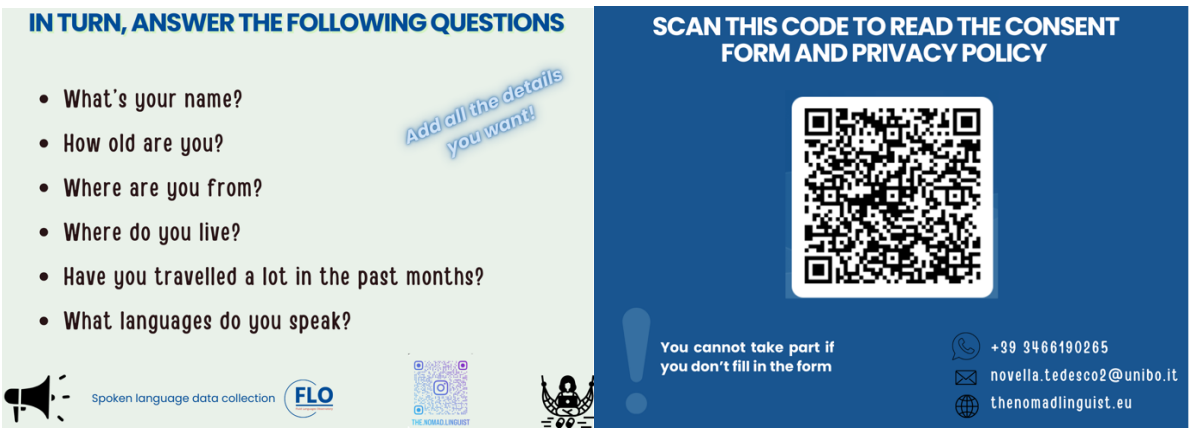


Figure 57. Flyer containing link to privacy and consent forms and prompts for NomadLingo metadata collection through audio-recording (English Version).

Table 11 provides general information about the data collected in Canary Islands, while Table 12 provides information about each data collection session, a detailed description of which can be found in Appendix 6.

Table 11. General information about data collected in Canary Islands between February and April 2024.

No. of sessions	5
Total length of the recorded audio	15h11min
Avarage length of the recorded audio x session	2h10min
Total No. of participants	37
Total No. of nationalities represented	14
Median no. of participants x session	8 (St. Dev. 2.04)
Median no. of languages spoken by participants	3 (St. Dev. 1.44)
Five most spoken languages by participants	English (35), Spanish (26), French (15), Italian (12), German (11)

Table 12. Canary Islands data collection sessions: An overview.

Session ID	**Location**	**Context type**	**No. participants**	**Participants nationalities**	**Included in NomadLingo (v1)?**
GC1	Las Palmas de	Controlled environment	13 + researcher	Argentinian, Swedish, Czech,	YES

	Gran Canaria			German, Italian, Spanish, Finnish	
GC2	Las Palmas de Gran Canaria	Public	7 + researcher	Norwegian, Australian, Czech, Italian, British, Polish, Jamaican	NO
GC3	Las Palmas de Gran Canaria	Private	8 + researcher	Finnish, Swedish, English, German, Croatian, Italian	NO
F1	Corralejo	Private	8 + researcher	Argentinian, Polish, Spanish, British/French, French, Italian	YES
F2	Corralejo	Controlled environment	5 + researcher	Estonian, Spanish, Italian, French, German	YES

5.2.5. The raw data

The raw dataset resulting from the data collection sessions (excluding the ones in Porto Santo) comprises approximately 27 hours of audio, covering 10 events and involving 74 unique speakers all together. An overview of the raw dataset is provided in Table 13.

Table 13. NomadLingo: Raw Data.

Total time recorded	26h52min
Number of sessions	13
Number of speakers	74
Number of speakers' nationalities	29
Number of languages spoken by speakers	27
Median of languages spoken by each speaker	3 (St. dev. 1.33)

From these 27 hours, 11h38m were transcribed for the first version of the NomadLingo corpus (NomadLingo1.0 open, Tedesco et al. 2025; and NomadLingo1.1, Tedesco et al. 2026), consisting of 6 sessions recorded at Madeira and Fuerteventura, with GC1 and M7 leeading to two and three separate conversation transcripts, respectively.

Challenges, such as multiple overlapping conversations and presence of languages not known to the researcher, will be addressed in 5.4, since they affected the transcription, segmentation and annotation procedures.

5.3 Metadata and contextual annotation

Building on the mixed-methods and ethnographic foundations of the project, this section explains how contextual notes were kept during and after each recording session, how metadata were structured, standardised, and shared as corpus accompanying documentation. Figure 58 shows the general workflow concerning contextual metatada.

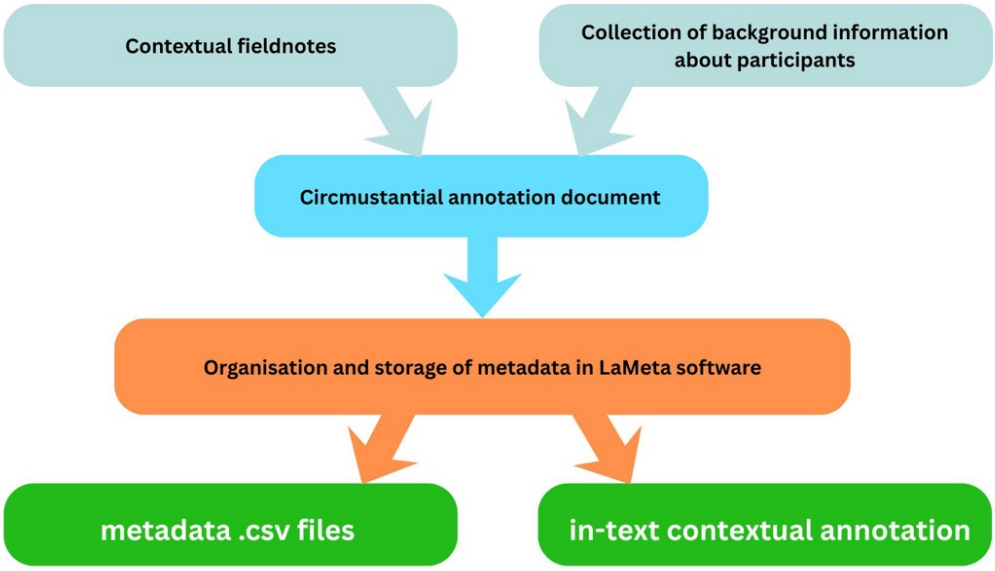


Figure 58. NomadLingo contextual metadata workflow.

The section has three parts. 5.2.1 describes the tools and procedures used for recording contextual information, including the use of structured templates and the LaMeta metadata model (Hatton et al. 2021); 5.2.2 introduces the principles and formats used for in-text annotation of contextual variables within the transcripts themselves; finally 5.2.3 provides an overview of the corpus accompanying documentation (NomadLingo1.1 open, Tedesco et al. 2026).

5.3.1 Keeping contextual notes and LaMeta for metadata storage

The systematic collection of contextual notes served as the interpretive bridge between the raw audio recordings and the goal of building metadata-rich resources. In spoken and ethnographic corpora, contextual documentation is essential for capturing the social, environmental and interactional conditions under which the data were produced – conditions that are absent from the audio signal but crucial for any subsequent linguistic, sociolinguistic or interactional analysis (Adolphs & Knight 2020; Tusting 2020).

In line with the broader triangulation framework outlined in 4.2 and 4.3, contextual notes provide information about the setting, participants, and interaction modes. They also record

situational contingencies, such as noise, lighting, interruptions, physical arrangement of participants, etc., that affect the interaction.

Contextual notes were drafted both during and immediately after each recording session through a combination of quick fieldnotesand expanded post-event descriptions resulting in the annotation documents (Appendix 4).

To guarantee interoperability and future reuse, contextual notes were transferred to LaMeta, an application specifically developed for metadata management of linguistic resources (Hatton et al. 2021). LaMeta provides a structured framework for representing session-level and participant-level information. It also allows one to link consent and privacy forms to participants (see Figure 58), a feature which, given the large number of participants in the project, turned out to be particularly useful for the management of ethical and legal documentation. The software allows for several exporting formats, including .csv.

Figure 59 shows the main metadata fields in the LaMeta *people* page.

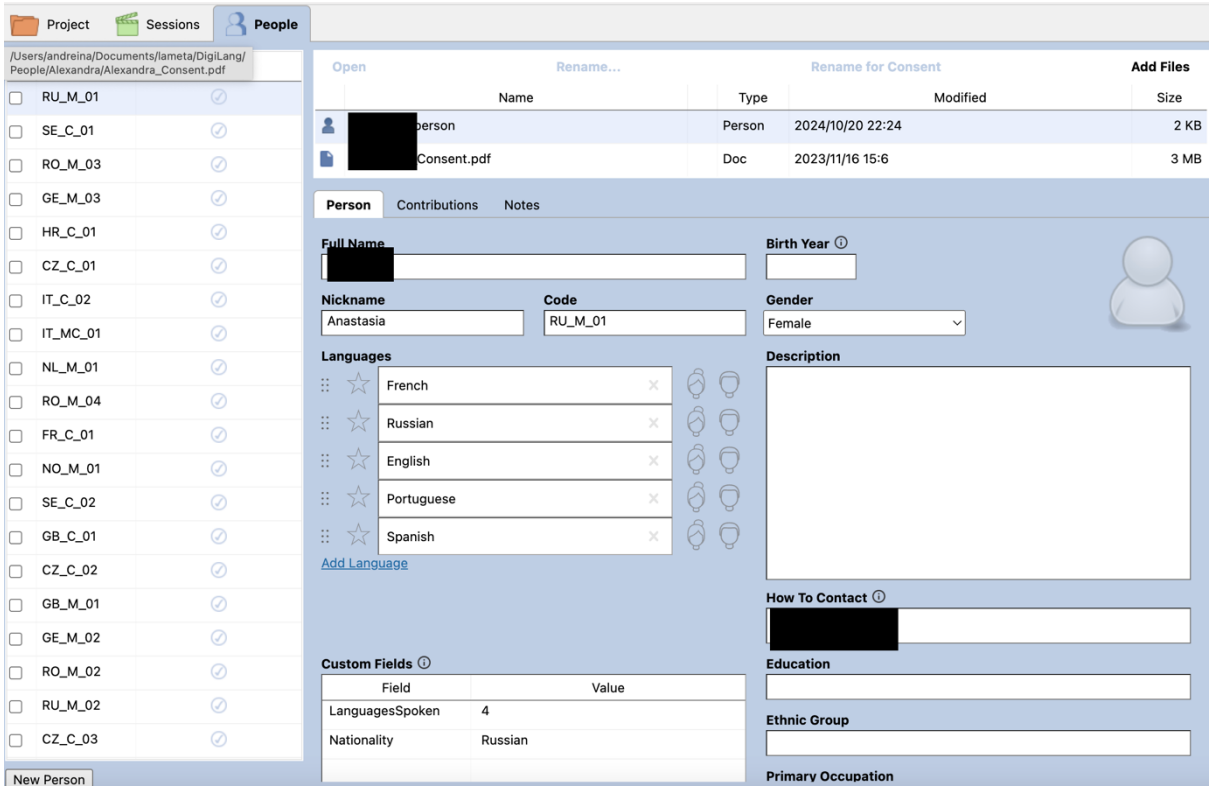


Figure 59. LaMeta *people* metadata input form.

All participants were assigned an identification code and a nickname (more details about this procedure are given in 5.3). The pieces of information which for participants were: a) full name; b) nickname; c) code; d) languages; e) number of languages spoken; f) nationality; g) gender; h) contact information. However,if more was known about participants, like their educational history or mobility trajectory, additional information was entered in the *Notes* field.

The *Sessions* panelcontains information about the data collection sessions . Specifically, the following were considered as mandatory fields for the NomadLingo project: a) ID; b) date; c) title; d) description; e) genre (conversation); f) subject language (translanguaging); g) working language (English) h) Town i) audio quality; j) number of recordings; k) overlapping; l) researcher involvement (silent observer, in cases where researcher was only observer / participant, in cases where the researcher's voice was part of the recordings); m) region (Canary Islands / Madeira Islands) n) country (Spain / Portugual); o) continent (Europe) p) planning (spontaneous, when the data collection session had not been organised by the researcher, like M6 / semi-spontaneous, when the data collection session had been planned for an event not organised by the researcher, like M2 / planned, when the event for the data collection session had been organised by the researcher, like M1); q) social context (private, when data were collected in a private house like F1 / controlled environment, when data were collected in a space that is not private but with restricted access, like a coworking as in the cases of M7 and M8 sessions / public, when data were recorded in a public space like a restaurant, as M6); r) interaction type (free / task-based, like in F2, where speakers were doing a conversation activity); s) speakers introduction (yes, for audios recorded in Canary Islands / no, for audios recorded in Madeira Islands).

In Appendix 7, metadata for participants and sessions in the FLO project are shown. The metadata participants and sessions included in the current version of the corpus are instead provided in Appendix 8.

5.3.2 In-text metadata, contextual and structural annotation

In addition to session-level and speaker-level metadata, stored in .csv files (see Appendix 8), contextual information is included in the NomadLingo corpus through in-text metadata annotation, that is, information embedded directly within the transcript. In-text annotation constitutes a key component of spoken corpus design because it enables researchers to link linguistic patterns to contextual variables when using corpus analysis tools (Adolphs & Knight 2020). For example, in the case of NomadLingo, the inclusion of in-text annotation allows analysts to track how interactional dynamics, speaker identities, contextual cues can shape and constrain the production of translingual speech.

In the NomadLingo corpus, in-text annotation follows principles derived from the TEI Guidelines for spoken data (Burnard 2004; Windhouwer & Wright 2012). While the XML element *session* provides an overview of contextual variables at session level and the <speaker> tag includes information about speakers' nationality and linguistic repertoire for each turn, other annotation layers captures the micro-level dynamics of the recording: the beginning and end of speaker turns, with the <turn> tag, overlaps, signaled by the transcription convention [], non-verbal events, noise, disruptions and insertions, enclosed in the tag <sound>.

The XML structure for the files contained in the NomadLingo corpus is shown in Figure 60.

```
<session>
    <conversation>
        <turn>
            <speaker>
            </speaker>
        </turn>
        <turn>
            <intro>
                <speaker>
                </speaker>
            </intro>
        </turn>
        <sound>
        </sound>
    </conversation>
</session>
```

Figure 60. NomadLingo1.1 XML basic structure.

The first element, *session*, is present at the beginning of each file. This element, which encloses all the others, has the following attributes: code (corresponding to Session ID in the metadata file), context, interaction (corresponding to interaction type), languages (the list of all the languages known by speakers participating in each session), location, nationalities (a numeric value representing the different nationalities of speakers in each session), planning, region, speakers (a numeric value representing the number of speakers in each session), speakers_intro, and title.

The transcribed recorded is enclosed in the tag <conversation>. Within the conversation, two main elements can be found at the same level, namely *turn*, corresponding to one conversation turn, and *sound,* containing annotation about non-verbal sounds, noise, or inaudible conversation portions. Each transcribed turn is then enclosed in a <speaker> tag, which carries the attributes ID, nationality, and languages (the languages that speakers declared to know). The tag <intro> signals the utterances in which speakers introduce themselves (the place they come from, their languages, their travel trajectories). These are only present when the attribute *speaker_intro* in the *session* element takes the 'yes' value.

The whole scheme encompassing contextual and structural annotation is presented in Appendix 9.

This annotation system serves three main purposes:

1. documentation, by allowing researchers to reconstruct the interactional ecology of the event;
2. interpretation, by providing analytical cues for understanding specific linguistic phenomena (e.g. misunderstandings triggered by noise);
3. reusability, by supporting transparent analysis and facilitating future work on the corpus.

The in-text metadata and structural annotation layers provide the foundation for the linguistic annotation described in 5.5. Indeed, tagging of translanguaging episodes, repairs, and misunderstandings requires stable turn segmentation, clearly identifiable speaker roles, and contextual cues embedded in the transcript. By linking structural, contextual and linguistic tags within a unified XML architecture, NomadLingo allows for complex search operations, such as extracting translanguaging events co-occurring with misunderstandings, or repair sequences triggered by noise (see 5.6 and Chapter 6).

5.3.3 Issues with existing TEI guidelines

It should be noticed that although the TEI Guidelines offer well-established models for representing spoken language (TEI Consortium 2025), the NomadLingo corpus does not adopt a full TEI-compliant transcription and annotation scheme. This decision reflects both the methodological aims of the project and the empirical specificities of the data. TEI schemas for spoken corpora, especially those based on the TEI <transcription> module, or the conventions used in projects like the Santa Barbara Corpus (Du Bois et al. 2000, 2005), or the British National Corpus Spoken (Burnard 2004), presuppose relatively controlled recording conditions, stable participation frameworks, clear turn-taking boundaries, and are often aimed at representing fairly homogeneous linguistic and cultural groups. None of these conditions holds in the highly fluid, transcultural – hence translingual – and acoustically variable environments in which NomadLingo data were recorded.

In particular, TEI's mechanisms for marking utterance units (<u>), annotations (<annotation>), and alignment with timeline-based segmentation (<timeline>, <anchor>) are extremely fine-grained, and would impose a structural rigidity incompatible with the data at stake, characterised by overlapping conversations, fragmented turn boundaries, shifting participation, and the presence of inaudible portions. Fully implementing TEI guidelines would therefore risk misrepresenting the interactional reality of the recordings or forcing artificial segmentations inconsistent with actual participation dynamics.

For these reasons, the NomadLingo schema follows a TEI-inspired but simplified architecture: it retains TEI principles of hierarchical organisation and XML-based turn encoding but applies these elements in a way that prioritises ecological validity and analytical usability. Among TEI's existing modules, NomadLingo most closely approximates TEI Spoken (chapter 8 of the TEI Guidelines, TEI Consortium 2025) and, to some extent, the ISO/TEI minimal spoken transcription profile (ISO/TC 37/SC 4 2014; Schmidt 2016), though both are adapted to suit the specific needs of the project. This hybrid approach makes the corpus interoperable with major spoken corpus practices while representing complex translingual interactional phenomena without distorting their natural characteristics.

A summary of the main differences between TEI Spoken Guidelines and the NomadLingo schema is provided in Table 14.

Table 14. A comparison between TEI Spoken Guidelines and NomadLingo contextual and structural annotation schema.

Feature	**TEI Spoken Guidelines**	**NomadLingo Schema**	**Rationale**
Turn segmentation	<u> (utterance units), often tied to syntactic or prosodic boundaries	<turn> units based on interactional boundaries (ecological segmentation)	Avoids imposing artificial utterance boundaries in fluid talk
Timeline alignment	<timeline>, <anchor> for precise time-coded utterances	Not used; timestamps encoded only when relevant	Timeline alignment too time-consuming for informal, noisy multi-party settings
Speaker identification	Structured <person> entries in <teiHeader>	Pseudoanonymised IDs (Nationality_Location_ Number)	Preserves anonymity; describe sociolinguistic information
Overlap encoding	<overlap> with multi-point alignment	Simplified overlap markers [] (see 5.3.1 for the full list of transcription conventions)	Reduces annotation complexity in multi-thread conversation
Environmental events	<incident>, <vocal>, <sound>	Simplified <sound> tags	Ensures structural clarity and allows to add richer contextual information within the tag (e.g. ((speakers laughing)); ((inaudible portion in Portuguese)); ((bells ringing)) – see 5.3.1 for the full list of transcription conventions)

5.3.4 Corpus accompanying documentation

The first version of the corpus, NomadLingo1.0 open, deposited in June 2025, contains metadata .csv files for both participants and data collection sessions of the whole project (hence related to the raw data), and participant and session metadata specifically linked to the transcribed parts. In the latest version of the corpus, NomadLingo1.1 open (Tedesco et al. 2026), the accompanying documentation and metadata files were revised.and the annotation schema in .csv format was added (Appendices 8 and 9). However, the metadata .csv files related to the raw data were not deposited with the corpus. These files are part of the OSF and

GitHub repositories for the FLO project[25], while the *NomadLingo1.1 open* corpus folder, that can be downloaded from the ILC4CLARIN repository, only contains metadata files related to participants and sessions included in the corpus. Specific characteristics of these files are detailed in the readme.txt file accompanying the corpus (see Appendix 10).

5.4 Processing translingual spoken data

Processing the raw data for the NomadLingo corpus required a workflow capable of dealing with the specific challenges presented by the recordings. A semi-automatic transcription pipeline was implemented, combining automatic tools (Whisper, WhisperX, CapCut) with careful human revision and ethnographic contextualisation. Automatic transcription was used as a preliminary tool; each file then underwent extensive manual revision to ensure accuracy. The performance of such tools for transcribing non-standard speech was also tested. Moreover, the section provides insights into the advantages and disadvantages of using ELAN (Wittenburg et al. 2006), at date one of the most popular tools for managing spoken language data, with a dataset made of conversations among more than 4 speakers, often fragmented and with recurrent breaks and overlaps. Sub-section 5.4.1 outlines the rationale and implementation of this semi-automatic pipeline.

Sub-section 5.4.2 introduces the pseudoanonymisation protocol developed specifically for NomadLingo, aimed at ensuring both ethical compliance and analytical usability by encoding speakers through composite identifiers that preserve relevant sociolinguistic information without enabling re-identification, and assigning nicknames to each speaker which could, at least partially, retain the cultural information provided in the real names.

The next sub-sections outline the core processes through which raw audio was transformed into corpus files.

5.4.1 Transcription pipeline and tools evaluation

The nature of the NomadLingo dataset made it necessary to adopt a semi-automatic transcription pipeline. Fully manual transcription would have been prohibitively time-consuming, while fully automatic transcription would have failed to produce accurate, reliable or ethically sound representations of translingual speech.

The recordings feature:

- large groups (6 – 14 participants);
- overlapping talk;
- noise (wind, music, waves, traffic);
- translanguaging;

[25] OSF: DOI 10.17605/OSF.IO/UK9NY; GitHub: https://github.com/novella-tedesco/FLO.

- variable microphone distances;
- temporary disappearance or arrival of speakers;
- informal speech phenomena (disfluencies, partial turns, collaborative completions).

These are known points of failure for automatic speech recognition systems (diarisation errors, false merging of speakers, smoothing of disfluencies, misidentification of languages), as also documented in automatic speech recognition (ASR) literature (e.g., Iwamoto et al. 2022; Raja & Giri 2025).

Before outlining the transcription pipeline, a comparison of the performances of the three ASR tools tested for automatic transcription is provided, namely Whisper, WhisperX and Capcut.

Whisper is an end-to-end ASR system developed by OpenAI and trained on 680,000 hours of multilingual and multitask supervised data (Radford et al. 2022). Its architecture – based on a transformer encoder–decoder model – makes it comparatively robust to noisy conditions and capable of handling a wide range of languages, accents, and speech styles. Whisper is increasingly adopted in research settings due to its multilingual coverage and open-source availability, which facilitate transparent and reproducible workflows (Kharitonov et al. 2023). In this project, Whisper was used through the WhisperGUI interface developed at the University of Bologna (Department of Interpreting and Translation), which allowed fully offline processing – an essential requirement for GDPR-compliant handling of data.

The main advantages of using Whisper have proven to be:

I. robustness to background noise: Whisper maintained relatively good performance in outdoor and semi-public environments, characterised by wind noise, overlapping voices, and environmental sounds;
II. fast processing of long recordings: the tool enabled transcription of multi-hour files in approximately one hour on a standard GPU-enabled machine, which significantly accelerated the initial transcription workflow;
III. data security: because WhisperGUI operates offline, audio never leaves the local machine, an important advantage for working with ethically-sensitive ethnographic audio;
IV. open-sourcing and reproducibility: Whisper's open-source license facilitates transparent processing pipelines, aligning with the FAIR principles discussed in Chapter 4 (Wilkinson et al. 2016). Its reproducibility also supports long-term corpus maintenance and versioning.

Analysis of Whisper output for the construction of the NomadLingo corpus revealed several systematic error types consistent with those described in ASR literature (Errattahi et al. 2018; Kharitonov et al. 2023; Iwamoto et al. 2022), namely:

I. over-standardisation of speech: Whisper often restructures natural speech into well-formed sentences, introducing punctuation and clause boundaries that were never present. This is a known effect of decoder-level language modelling (Jain et al. 2023);
II. flattening of disfluencies**:** Hesitations, repairs, false starts, and cut-offs were systematically removed or normalised. This false-fluency effect is well documented in ASR applied to conversational, multi-party speech (Shriberg 2001; Koenecke et al. 2020);
III. language misidentification in translingual segments: Mixed-language utterances were frequently classified as English only, or automatically translated, despite translingual input. ASR misclassification of code-switching is a well-established limitation in multilingual models (Yılmaz et al. 2018; Winata et al. 2022);
IV. speaker merging: Whisper does not perform diarisation; as a result, overlapping or adjacent voices werecollapsed into a single speaker stream.

WhisperX (Bain et al. 2023) extends Whisper by adding forced alignment, speaker diarisation, and more precise segment-level timing using wav2vec2-based alignment models and PyAnnote diarisation. Some of the advantages of using WhisperX included:

- more accurate timing and segmentation than Whisper alone;
- automatic diarisation, providing an initial – although imperfect – basis for assigning speaker identity;
- provision of chunking before manual revision, reducing transcription load.

However, consistent with known WhisperX errors (Bain et al. 2023), substantial failures in diarisation, over-segmentation and over-standardisation were witnessed to an extent that WhisperX did not seem to provide a better output in comparison to Whisper, while being characterised by a less user-friendly implementation. For these reasons, after the testing phase, it was excluded from the transcription pipeline.

The third piece of software tested during the project was CapCut (ByteDance Ltd. 2023). Although CapCut is designed as a video-editing tool, its transcription outputs proved to be sometimes better than those provided by Whipser or WhisperX. The software recognises laughter, deals in a satisfactory manner with background noise, and does not systematically standardise language.

Since performance varied across files, in the end, Whisper was the main tool used for transcription for practical reasons (it provides outputs in different formats and is compliant with data security requirements).

Besides ASR software, a variety of tools are traditionally used in linguistic research to manage, annotate and analyse spoken language. One of the most popular ones is ELAN (Wittenburg et al. 2006). It is the reference tool for multimodal linguistic annotation and widely used in linguistic anthropology, sociolinguistics, and interaction studies. Its advantages lie in a multi-tier annotation system, tools that allow precise time alignment, output compatibility with TEI,

multimodal workflows, and a highly structured data representation. ELAN can also read .srt files containing time stamps, like the ones created by Whisper. For these reasons, ELAN was used to process the first audio file to be contained in the NomadLingo corpus. However, ELAN is fairly heavy to run on a normal laptop and many bugs were reported while using it. Moreover, the output provided by Whisper seemed to be *not good enough* in terms of transcript-audio alignment, and segmentation. Thereby, after this first try, ELAN was excluded from data processing, for two main sets of reasons:

I. tier management problems: a multi-party interaction with more than 4 speakers – as it is often the case with NomadLingo data – would require a large number of tiers per file, making the interface difficult to consult;
II. recurring bugs which caused data loss, slowed down or impeded revision processes.

The option which seemed to guarantee a fairly fast revision process and full customizability seemed to be that of working with .txt files. Indeed, this option allowed for rapid revisions, by more or less proficient linguists (like students or proficient speakers with no linguistic research background). This approach finds support in the ethnographic literature that encourages lightweight, iterative annotation for corpora (Bucholtz & Hall 2016; Snell 2018).

Hence, the final transcription and revision pipeline was developed as follows:

1. Initial transcription produced through an ASR tool (in most cases, Whisper through the WhisperGUI interface);
2. First manual revision, focused on turn segmentation and assignment;
3. Conversion of the .txt file into an .xml file and addition of the first layer of metadata and contextual annotation, using the Bbedit text editor (Bare Bones Software Inc. 2023);
4. Second revision round and application of transcription conventions;
5. Third revision round with the identification of translanguaging phenomena;
6. Revision of non-English bits by competent speakers;
7. Fourth revision, carried out on the text file only (without the audio) to add and/or standardise structural and linguistic annotation tags through the use of regular expressions;
8. XML structure validity check through the use of the Visual Studio Code app (Microsoft Corporation 2023);
9. Final revision by external revisors (university students);
10. Implementation of revision suggestions.

For the compilation of *NomadLingo1.1*, the updated version of *NomadLingo1.0*, a further revision round of transcripts against the audio files was carried out.

The following transcription conventions were adopted (adapted from Jefferson (2004)):

- (1) for pauses longer than a second;

- - for interrupted speech;
- [ ] for overlapping speech, that is, clipped words;
- ((laughs)), ((screams)) to annotate sounds assigned to a particular speaker;
- ((laughing)), ((screaming)) to annotate sounds not assigned to a particular speaker;
- : for longer sounds;
- CAPITAL LETTERS for words pronounced significantly more slowly or with a significantly higher level of voice;
- punctuation kept to a minimum; no use of exclamation markers (!) and use of question marks (?) only when an interrogative intonation could be clearly heard;
- phonological variation standardised, morphological variation transcribed as it is heard in the audio file (this choice was made since phonetic transcription was not carried out);
- 'ehm' for any filled pauses;
- X to transcribe inaudible syllables;
- #hi# to signal that a transcription may be incorrect.

The main segmentation criterion was based on discourse units (Mahlberg 2014), that is the longest possible units that have independent meaning and are understandable even when taken out of context. Specifically, the following criteria were applied flexibly depending on the case:

- generally, one conversational turn corresponds to a segment;
- a segment can be interrupted by brief segments spoken by other speakers (in such cases, the segment is not split, and overlapping parts are indicated in square brackets in a subsequent turn attributed to the other speaker).

As soon as the first version of the transcripts were produced, the workflow moved into a second phase: collaborative revision. Given the nature of the data, it became evident from the beginning of the transcription work, that no single annotator could adequately capture the full linguistic, pragmatic and cultural complexity of the interactions.

Collaborative revision was needed for three main reasons:

a) linguistic diversity beyond individual competence: with more than six languages appearing in the corpus – often in rapid alternation, hybrid forms, or with partial borrowings – it was essential to rely on speakers who possessed a strong competence in the languages at stake. Their contributions were crucial for resolving ambiguous sequences, preserving non-standard forms, distinguishing pragmatically marked items from ASR errors, and identifying culture-specific references not easily detectable through surface cues;
b) preventing over-standardisation: individual annotators – even highly trained ones - tend to inadvertently normalise speech when transcribing. Collaborative revision, supported by detailed revision guidelines, functioned as a safeguard against this risk, ensuring that disfluencies, accented forms, partial utterances, and code-mixed segments were retained rather than corrected. Moreover, the revision process did not only concern transcription but also annotations. The

linguistic annotation system applied on NomadLingo data, as will be discussed in Section 5.5, is not exempt from a certain degree of subjectivity – something that could be argued for any annotation system (Ide & Pustejovsky 2017). Hence, external revision was needed to guarantee that categorisation of linguistic phenomena was applied in the most possible objective way.

c) embedding community expertise within the corpus and enhancing the educational value of the research project: involving digital nomads themselves brought an additional layer of contextual and sociocultural knowledge to the transcripts. Their participation strengthened the alignment between the recorded events and their subsequent interpretation, while also contributing to the project's broader aim of producing FAIR and socially grounded research outputs. On the other hand, including translation students from the University of Bologna in the group of revisors guaranteed revisors with linguistic expertises, while providing purposeful learning activities for them.

The revision unfolded across multiple stages, each of which contributed a distinct layer of expertise and refinement. For the final revision (point 9 in the transcription pipeline), a protocol was developed. XML files were shared with revisors, together with contextual notes, audio files and a guidelines file. The guidelines asked revisors to focus on translingual segments, correct application of transcription conventions, speaker attribution consistency, conversational markers such as laughter, overlaps and cut-offs, pauses, annotated segments. They explicitly reminded revisors not to standardise grammar, not to interpret unclear stretches, and avoid punctuation except question marks.

By structuring revision as a multi-party, iterative and reflexive process, the project achieved greater accuracy, higher ecological validity, reduced transcription bias, improved reliability of coding for translanguaging, repairs, and misunderstandings (see Section 5.5).

5.4.2 Applying pseudoanonymisation to the dataset

The decision to adopt a pseudoanonymisation system for NomadLingo emerged from both ethical and methodological considerations. Digital nomads constitute a population with unusually high online visibility: their mobility trajectories, professional identities, and interpersonal networks are frequently documented across Instagram, LinkedIn, Facebook groups, and Slack workspaces. This means that even seemingly innocuous linguistic details can become powerful identifiers when cross-referenced with public digital traces. At the same time, classical anonymisation techniques (e.g., assigning participants generic labels such as *S1*, *S2*, *S3*) seemed unsuitable because they erase precisely the sociolinguistic information that can be important in linguistic and interactional analysis. In the case of NomadLingo, erasing socio-cultural attributes would compromise the interpretation of sociolinguistic identity cues, which shape translanguaging, misunderstanding, humor, stance-taking, and indexicality.

Pseudoanonymisation is a widely recommended practice in qualitative and corpus-based research involving human participants, particularly when complete anonymisation would

compromise analytical value (Allen & Wiles 2016; Saunders et al. 2015). In contrast to full anonymisation, pseudoanonymisation removes identifying information while retaining structured, non-identifying attributes that are relevant for interpretation (European Data Protection Board 2019). Many corpus projects adopt controlled identity-masking systems that preserve sociolinguistic categories (Baude & Blanche-Benveniste 2006; De Decker 2020).

In NomadLingo a two-layer pseudoanonymisation system is adopted, designed to balance ethical protection with analytical utility. The first layer consists of a structured alphanumeric code assigned to each speaker. These codes are unique and follow a consistent pattern, with three components representing nationality, site of data collection, and speaker identification number (see Table 15). For example, 'IT_MC_01' is an Italian speaker whose voice was recorded in both Madeira and the Canary Islands, 'GE_M_03' is a German speaker participating in Madeira sessions, and 'GB_C_01' is a British participant in the Canary Islands.

Table 15. Speaker codes in NomadLingo.

Component	**Meaning**	**Example**
Nationality	Speaker's nationality: broad, non-identifiable category but useful for sociolinguistic analysis	IT; GE; GB
Data collection fieldsite	Location where the recording took place	M (Madeira Islands); C (Canary Islands)
Ordinal number	Internal project unique identifier	01; 02; 03

This compact coding system preserves a high level of sociolinguistic interpretability. Importantly, each element of the code is non-traceable, avoiding surname initials, city names, or other identifiers that could be cross-searched online.

Since codes are not event-specific, they also allow longitudinal tracking across events.

A second pseudoanonymisation layer became necessary because, as earlier acknowledged, names themselves often serve interactional and cultural functions in the corpus. In transcultural interactions, names may be misheard, corrected, playfully adapted, or used as resources for teasing and stance-taking. These phenomena cannot be meaningfully represented if all names are removed.

A telling example appears in the exchange:

> RS_M_01: Bogdan
> IT_MC_01: Dan?
> RS_M_01: Bogdan not Dan
> IT_M_01: Bogdan ok where are you from?

The actual name of RS_M_01, Pedrag, is misinterpreted by the Italian interlocutor as 'Pedro'.. The misunderstanding sequence here presented revolves around the phonetic shape of the name. Replacing both forms with a neutral placeholder would erase the reason for the misunderstanding, making the rest of the sequence meaningless (clarification, alignment work and conversational entry-point for IT_MC_01 to ask about RS_M_01's nationality.

For cases like this, NomadLingo adopts a non-identifying pseudonym, chosen solely for its interactional and cultural fit and not for its resemblance to the participant's actual name. In practice, this means:

- culturally marked forms (e.g., diminutives, affectionate nicknames) may be preserved through an invented equivalent;
- names indexing cultural belonging or intimacy retain their pragmatic force;
- corrections like "X not Y", as in the example above, remaininterpretable.

For example, Marc, a typical Catalan name, was replaced with Arnau; Claire, the name of a participant with double British-French nationality, was replaced with Louise, a name that could preserve the ambiguity around her nationality.

Summing up, whereas layer 1 preserves sociolinguistic profiles, layer 2 preserves cultural meaning of names in interaction. This means that if a speaker's name is not interactionally relevant, their nickname never appears: only the structured speaker code is used.

The combined system operates as shown in Table 16.

Table 16. The two-layer pseudoanonymisation system in NomadLingo.

Location in the dataset	**Appears as…**	**Purpose**
.csv metadata file, <speaker ID> tag	IT_MC_01, <speaker ID="IT_MC_01">	analytical comparability; cross-session tracking; consistency and interoperability
Body of the transcript	Marco *(only when relevant)*	to preserve conversational meaning

The adoption of these two layers contributes directly to the analytical robustness of the corpus. The structured codes enable quantitative and qualitative cross-file analyses, while providing sociolinguistic information in a compact manner. Meanwhile, pseudonyms safeguard the ethnographic specificity of conversational events without exposing participants' identities,

5.5 Linguistic annotation

The linguistic annotation added to the corpus was designed to capture phenomena that are particularly salient in informal, translingual conversational data. While the annotation scheme does not aim to provide an exhaustive account of all dimensions involved in plurilingual communication, it focuses on a set of analytically central (see Chapter 6) and theoretically motivated aspects. Since the corpus documents spontaneous interactions where speakers mobilise highly heterogeneous linguistic repertoires and rely on a range of strategies to maintain mutual understanding, negotiate and co-construct meaning, the annotation scheme focuses on three domains that are central to such type of communication: translanguaging practices (Blackladge & Creese 2017; Wei 2018; Mazzaferro 2018), misunderstanding sequences (House 1996; Cervini & Paone 2025), and repair strategies (Schegloff et al. 1977), each representing a different dimension of how speakers manage transcultural understanding through linguistic fluidity. The following sub-sections describe these three core annotation domains.

5.5.1 Translanguaging

Translanguaging is a defining feature of the NomadLingo corpus and probably the most conceptually significant aspect of the project. In line with scholarship conceptualising translanguaging as a practical theory of language (García & Wei 2014), and with sociolinguistic approaches to linguistic fluidity in mobile, superdiverse contexts (see Chapter 3), the annotation does not treat languages as discrete, bounded systems. Instead, it identifies points in the interaction where speakers draw on, negotiate, or re-deploy semiotic resources from their repertoires in ways that become visible from a language-usage perspective.

For the purposes of annotation, translanguaging phenomena are considered as all those instances where translanguaging processes become evident in linguistic data. For operational and analytical reasons, translanguaging phenomena are categorised into a) codemixing, when a segment contains words in any language or variety other than standard English (see Example 1); b) codeswitching, when a segment is in a language or variety other than standard English (Bathia et al. 2004) (see Example 2); c) translation, when a turn includes a self-translation or peer-interpreting (Antonini et al. 2017) (see Example 3).

Example 1

```
<turn ID="M7_2_0082" Length="16">
    <speaker code="RS_M_01" nationality="Serbian"
language="English,Portuguese,Serbian,Italian">
        <translanguaging type="codemixing">do you have glasses for the vino? This is called
        the vino no?</translanguaging>
    </speaker>
</turn>
```

Example 2

```
<turn ID="GC1_2_0634" Length="2">
    <speaker code="AR_C_02" nationality="Argentinian"
language="Spanish,English,Argentinian">
```

```
        <translanguaging type="codeswitching">donde queda?</translanguaging>
    </speaker>
</turn>
<turn ID="GC1_2_0635" Length="6">
        <speaker code="ES_C_04" nationality="Spanish" language="Spanish,English,French">
                it's in front of mercadona
        </speaker>
</turn>
```

Example 3

```
<turn ID="M7_2_0254" Length="2">
        <speaker code="GE_M_04" nationality="German" language="German,English">
                plans change
        </speaker>
</turn>
<turn ID="M7_2_0255" Length="2">
        <speaker code="NO_M_01" nationality="Norwegian"
        language="Norwegian,English,Portuguese">
                <interactional_move type="asking">blends strange?</interactional_move>
        </speaker>
</turn>
<turn ID="M7_2_0256" Length="4">
        <speaker code="GE_M_04" nationality="German" language="German,English">
                <translanguaging type="translation">pläne ändern sich</translanguaging>
        </speaker>
</turn>
```

This English-centric representation has been applied for reasons of convenience, given that English is the official declared language of many digital nomad communities (Tedesco 2025a) and that the majority of the segments are in English. It should be noted that the distinction among different types of translanguaging does not represent any theoretical position but has only been applied for analytical reasons.

5.5.2 Misunderstanding

Misunderstanding sequences constitute the most pragmatically rich layer of annotation in *NomadLingo 1.1*. They are defined as stretches of talk in which a lack of understanding or misalignment can be observed, generally prompting a sequence of interactional moves aimed at resolving the issue. Misunderstandings are annotated as solved or unsolved, depending on whether mutual understanding is achieved through meaning negotiation and co-construction.

In the annotation structure adopted for the corpus (inspired by Cervini & Paone 2025), misunderstanding sequences are represented as containers of turns which, when relevant, are further annotated for their interactional role. A misunderstanding sequence can involve three types of interactional move: a) asking, when speakers ask for clarification (see Example 4); b) checking, when speakers check their interlocutor(s) understanding (see Example 5); c) confirming, when a speaker confirms that their interlocutor(s) have correctly understood (see Example 6).

Example 4

```
<misunderstanding type="solved">
        <turn ID="M7_3_0097" Length="7">
                <speaker code="PT_M_02" nationality="Portuguese"
language="Portuguese,Spanish,English,German">
                        Where are you from though in Italy?
                </speaker>
        </turn>
        <turn ID="M7_3_0098" Length="1">
                <speaker code="IT_MC_01" nationality="Italian"
language="Italian,English,Spanish">
                        <interactional_move type="asking"> what? </interactional_move>
                </speaker>
        </turn>
        <turn ID="M7_3_0099" Length="7">
                <speaker code="PT_M_02" nationality="Portuguese"
language="Portuguese,Spanish,English,German">
                        Where are you from though in Italy?
                </speaker>
        </turn>
        <turn ID="M7_3_0100" Length="3">
                <speaker code="IT_MC_01" nationality="Italian"
language="Italian,English,Spanish">
                        From the north
                </speaker>
        </turn>
</misunderstanding>
```

Example 5

```
<turn ID="M7_2_0008" Length="1">
        <speaker code="Researcher_IT" nationality="Italian"
language="Italian,English,Neapolitan,Spanish,German,French">
                <interactional_move type="checking"> ok? </interactional_move>
        </speaker>
</turn>
<turn ID="M7_2_0009" Length="2">
        <speaker code="GE_M_04" nationality="German" language="German,English">
                mh mh
        </speaker>
</turn>
```

Example 6

```
<turn ID="F2_0543" Length="2">
        <speaker code="ES_C_01" nationality="Spanish"
language="English,French,Spanish">
                your finance
        </speaker>
</turn>
<turn ID="F2_0544" Length="2">
        <speaker code="IT_MC_01" nationality="Italian"
language="Italian,English,Spanish">
```

no ehm
</speaker>
</turn>
<turn ID="F2_0545" Length="1">
<speaker code="GE_C_01" nationality="German" language="German,English,Spanish">
income
</speaker>
</turn>
<turn ID="F2_0546" Length="29">
<speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
<interactional_move type="confirming"> **income exactly**</interactional_move> **depend of your income it's like if you grow like 30000 you pay seventy percent if you're like fifty of your income is like twenty-one**
</speaker>
</turn>

It should be noted that there are few cases where interactional_move(s) are not included within a misunderstanding sequence, like in Example 5. Here the action of checking participants' understanding was not prompted by any real misunderstanding signal.

This structured annotation captures not only where misunderstandings occur but also how participants collaboratively resolve them.

5.5.3 Repair

Repair annotation draws on classic conversation-analytic distinctions (Schegloff et al. 1977). Spontaneous spoken interaction routinely features hesitations, self-interruptions, retracing and reformulations; these phenomena are particularly common in transcultural communication, where speakers ned to monitor accuracy, word retrieval, or interlocutor alignment. The annotation captures two broad categories, i.e., reformulation and restart, which both embody some kind of interruption in speech flow. Restart normally occurs at the beginning of a clause (as in Example 7), whereas through reformulation speakers modify part of their speech, rephrasing it (see Example 8). The boundary between these two phenomena is not sharp, though, and some cases may fall into both categories (Schegloff et al. 1977).

Example 7

<speaker code="GE_M_03" nationality="German" language="German,English,Spanish,Russian,Portuguese">
I was born in Russia but <repair type="restart">I was I descended</repair> **of Germans and then I went back to Germany and so: so it's it's weird to say half German half Russian because-**
</speaker>

Example 8

<speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">

you need to ask for it like in thailand it's like <repair type="reformulation">
everyone every place </repair>
</speaker>

The annotation explicitly focuses on self-initiated, self-completed repair**,** as this is the most frequent pattern in the corpus and the one most directly associated with fluency in language production (analysed in Chapter 6). Collaborative repair (i.e., other-initiated repair) was not systematically annotated for this first version of the corpus.

Repairs were treated as structural elements of spoken interaction, not as competence deficits. Their distribution across speakers and contexts provides valuable insights into linguistic processing, monitoring practices, and collaborative meaning-making among digital nomads, whose repertoires differ widely in structure, proficiency, and domain-specific expertise.

Taken together, these three annotation categories provide a multifaceted view of how communication unfolds in mobile, superdiverse communities, while leaving scope for future annotation of additional interactional dimensions afforded by the open nature of the data.

5.6 Using the corpus: Available formats, download and access

This section provides an overview of the formats in which the NomadLingo corpus is currently distributed and describes how users can access, download and work with it. At the time of writing, two versions of NomadLingo open are publicly available: NomaLingo1.0 and NomadLingo1.1. As was pointed out throughout the present chapter, the main differences between the two versions are:

- the addition of two annotation layers, namely misunderstanding and interactional_move;
- revision of transcripts;
- revision of the documentation (as discussed in Sub-section 5.3.4).

Both versions are distributed in multiple formats to support different analytical needs. Plain text (.txt) files are provided to maximise usability and interoperability, while comma-separated values (.csv) files are made available for corpus analytics and statistical modelling. In addition, separate auxiliary tables in .csv format document session-level and speaker-level metadata, as well as linguistic annotation. Users can download the dataset through the ILC4CLARIN open repository (https://dspace-clarin-it.ilc.cnr.it/home).

5.6.1 NomadLingo1.1 open

NomadLingo1.1 constitutes the enhanced, linguistically annotated version of the corpus (see readme file in Appendix 10).

Summing up, NomadLingo1.1 open includes:

- .txt files annotated in XML format, containing:
    - session attributes
    - turn-based structure
    - speaker attributes
    - introductory turns tagging
    - annotation of translanguaging phenomena (codemixing, codeswitching, translation)
    - annotation of repair mechanisms (restart, reformulation)
    - annotation of misunderstanding sequences and interactional moves (asking, checking, confirming)
    - sound events and multimodal cues
- The .csv file generated from the XML through a Python script, available on the project GitHub repository[26];
- Speaker and session metadata tables in .csv and .xls formats, aligned with the XML attributes;
- Annotation scheme in .csv and .xls formats.

NomadLingo1.1 was published with the hope that the scientific community could regard it as a valuable resource for:

- qualitative linguistic ethnography
- discourse and conversation analysis
- studies on ELF and mobility-based plurilingual practices
- translanguaging research
- statistical modelling of speech, interaction and translanguaging,
- sociolinguistic analysis (i.e., how social and cultural variables affect linguistic choices).

In Chapter 6, some examples are provided on possible explorations and uses of the dataset.

Besides being openly downloadable, NomadLingo1.1 is also accessible online through the free corpus query platform NoSketchEngine of the Department of Interpreting and Translation at the University of Bologna[27].

5.6.2 NomadLingo on NoSketchEngine

To facilitate access and to support users who may not be familiar with XML-based corpus processing tools, NomadLingo can be searched through NoSketchEngine, a lightweight, open-source version of the commercial SketchEngine platform. The installation hosted by the University of Bologna provides a browser-based interface that allows users to query the corpus,

[26] https://github.com/novella-tedesco/FLO/blob/NomadLingo---A-corpus-of-translingual-spoken-informal-interactions-at-European-digital-nomad-communities/Turn_txt_corpus_into_csv.ipynb.

[27] https://bellatrix.sslmit.unibo.it/noske/public/#dashboard?corpname=nomadlingo.

explore lexical patterns, and visualise concordances without downloading any files or installing additional software.

NoSketchEngine offers a set of tools that cover the most common corpus-querying needs:

- Concordance search: retrieves all occurrences of a word, phrase, or pattern displayed in KWIC (Key Word in Context) format.
- CQL search (Corpus Query Language)**:** allows structured searches using patterns such as part-of-speech tags, alternations, or multi-word expressions.
- Frequency lists: produce wordlists, lemma lists, or keyword lists.
- Distribution analysis of annotated phenomena through *text types* (an example is provided in Figure 61.
- Filtering by metadata: users may restrict searches to specific files, speakers, or session by using text types (an example is provided in Figure 62);
- Export tools: concordances and lists can be exported as csv or txt files for further analysis.

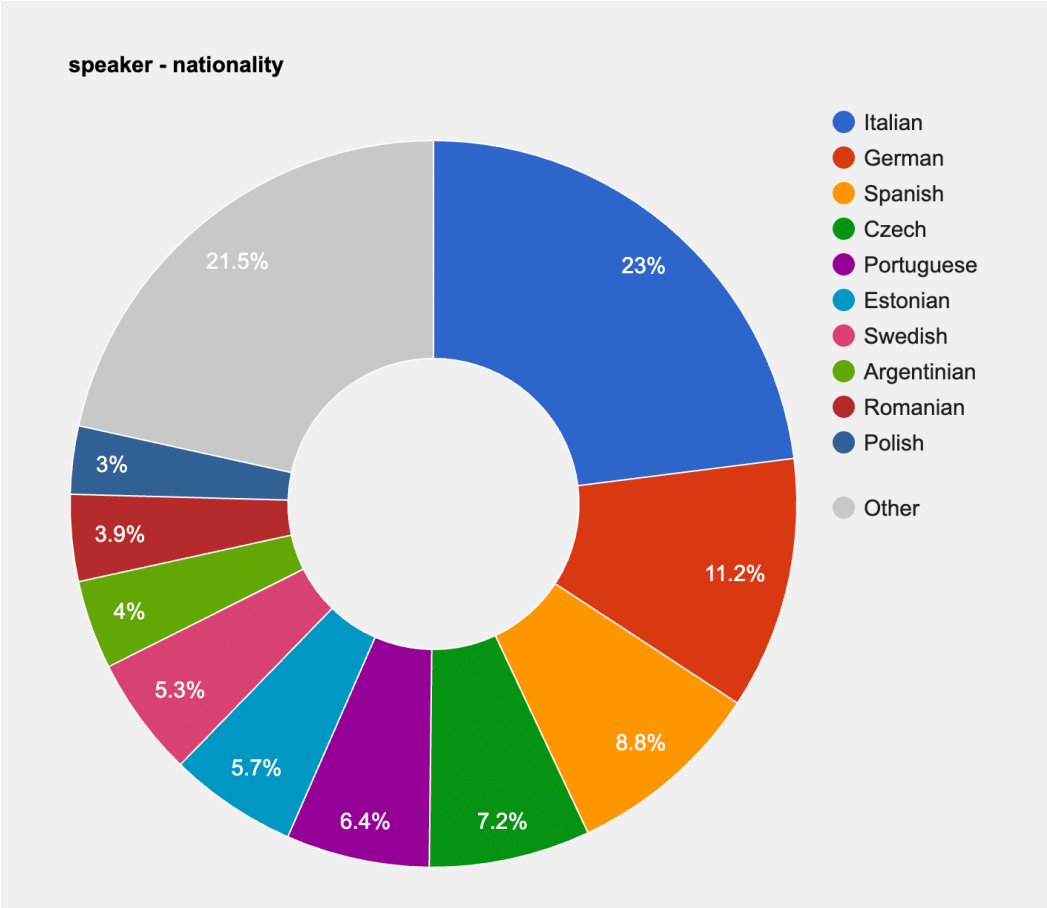


Figure 61. Speaker nationality distribution in NomadLingo1.1 produced through *Text type analysis* on NoSketchEngine.

Text types 1 (2) ••• | simple what • 32 | 366.83 per million tokens • 0.037% | KWIC

	Details	Left context	KWIC	Right context
1	#2490	ɜ of the best one of the best sessions of the season	what	? what ? today ? yes believe it or not but where ? Pı
2	#2492	best one of the best sessions of the season what ?	what	? today ? yes believe it or not but where ? Punta Ele
3	#2710	ghing)) oh ok and the local guy is the teacher ? and	what	happens with the : what happened with the : teache
4	#2715	:al guy is the teacher ? and what happens with the :	what	happened with the : teacher ? he taught me how to
5	#3273	laughs)) destroy the world ready ok did you finish ?	what	? the form ? yeah ? yeah so we start ? what 's your
6	#12774	ne here in the Canary if I have to pay less taxes but	what	the fuck I ca n't go ehm with my parents more than ɛ
7	#12953	ıll this ok and how is it working will it be recorded or	what	are you doing ? yeah it 's recording <intro> ah it 's r
8	#12969	? yeah it 's recording <intro> ah it 's recording ok ok	what	's your name so my name is Lea I 'm thirty-five year
9	#13150	lexico city and then from there to Germany </intro>	what	languages do you speak ? <intro> ehm I speak Ger
10	#17029	ask you next time question home base yes or no ?	what	? home base yes or no very complicated ((laughing)
11	#17646	dubai for example you have to spend time six days	what	? six days only no now it changed everything they c
12	#20349	ıse I have to introduce myself oh yeah I will ask her	what	is your name ((laughs)) ((laughs)) my name is ((laug
13	#20812	ıt I 'm going tomorrow are you going ? I do n't know	what	's the purpose ? I do n't know actually it 's just I do n
14	#21220	d you 're joining ? sorry ? and you 're joining ? sorry	what	? and you 're joining ? yes yeah yeah and they told
15	#24586	ou say this again ? escarbadientes escabardientes	what	's that ? escabardientes what 's that ? you do n't kno
16	#24591	entes escabardientes what 's that ? escabardientes	what	's that ? you do n't know ? esto palillo ah palillo that
17	#24767	tle stick how do you say it in german ? zahnstocher	what	? zahnstocher zahnstocher but it means the same a

Figure 62. Concordance lines for the word *what* within misunderstanding sequences.

The corpus was automatically lemmatised and POS-tagged on NoSketchEngine, setting the language to English. While this approach facilitates large-scale corpus querying through lemmatised wordlists and Corpus Query Language (CQL), it also comes at a cost. Non-standard English forms are not systematically lemmatised, and analytical procedures relying on these tools may privilege English turns, potentially obscuring the visibility of non-English items in translingual turns. At the same time, lemmatisation allows one to explore syntactic variation in English, supporting the analysis of non-standard English varieties and variation in English used as a lingua franca (see Chapter 6).

Moreover, assigning a language to the corpus on the platform allows for keyword extraction using reference corpora of that language. In this way, keywords can be used to highlight translanguaging phenomena. An example is provided in Figure 63 (reference corpus: UKWacComplete).

	Word		Word		Word		Word		Word
1	ehm	11	mh	21	nutella	31	bulga	41	napoli
2	inaudible	12	tanti	22	niente	32	levada	42	meatballs
3	ehm-	13	auguri	23	parmigiana	33	però	43	igni
4	mmhmm	14	pistacchio	24	calderao	34	cascais	44	natt
5	sì	15	teleras	25	pt_m_04	35	ponta	45	because-
6	like-	16	pistaccio	26	pt_m_05	36	funchal	46	melanzane
7	lasagna	17	yeah	27	snus	37	coliving	47	danubio
8	but-	18	zahnstocher	28	carbonara	38	instagram	48	udine
9	lorella	19	and-	29	gelatina	39	vabbè	49	tienes
10	covid	20	perché	30	pulga	40	madeirense	50	madeira

Figure 63. Translanguaging highlighted through keyword extraction using NoSketchEngine.

The advantages of using NomadLingo on noSketchEngine are numerous. The software allows for:

- immediate web-based access with no installation;
- easy-to-use for students or interdisciplinary researchers (Kilgarriff et al. 2015);
- annotation exploitation ;
- exportability, enabling hybrid quantitative/qualitative workflows.

However, it should be noticed that exploring the corpus through NoSketchEngine may reduce the potential of contextual annotation since the software was not designed for the processing of such annotation-rich corpora. Moreover, it inevitably endorses monolingual ideologies, and prevents complex, fair representation of language as *fluidly* used by plurilingual speakers (Tedesco et al. 2025a; Tedesco et al. 2025b).

Chapter 6

Analysing FLO data: Towards a better understanding of linguistic fluidity

6.1 Orientation to Chapter 6

This chapter presents four analytical case studies based on the resources collected in the FLO project, and in particular on the NomadLingo corpus. These analyses address complementary dimensions of linguistic fluidity.The chapter adopts an explicitly usage-based and interaction-oriented perspective, grounded in naturally occurring spoken data and supported by ethnographic metadata (participant and session metadata and structural and linguistic annotation). These elements enable an empirical investigation of language as social action in mobility-driven, transcultural communities, going beyond assumptions of monolingualism.

Across the four case studies, language emerges as a resource for positioning, alignment, negotiation, and community-building within the lived communicative experiences of digital nomads. The analyses foreground plural identities, transcultural repertoires, non-normative uses of English, and collaborative meaning-making. They are admittedly exploratory and small in scale, and their aim is to demonstrate how the NomadLingo corpus can be exploited together with the other datasets referred to in Chapter 4 to investigate different layers of linguistic practice, and how qualitative and quantitative approaches can interact.

Specifically, the case studies address the following dimensions:

- Case study 1 (Section 6.2) examines how speakers use linguistic resources to position themselves socially and culturally within mobility-driven communities, showing how identity is interactionally performed and co-constructed;
- Case study 2 (Section 6.3) investigates the characteristics of the English spoken by digital nomads, focusing on variation across lexical, grammatical, and pragmatic levels, and on how English operates as a permeable resource embedded in plurilingual repertoires rather than as a simplified lingua franca;
- Case study 3 (Section 6.4) analyses sequences of negotiation, misunderstanding, and irony, illustrating how understanding is interactionally achieved through repair, translanguaging, humour, and narrative re-keying;
- Case study 4 (Section 6.5) adopts a corpus-driven approach by statistically modelling translanguaging in relation to speech (dis)fluency, exploring how usage-based quantitative methods can contribute to the investigation of plurilingual competence.

The chapter progresses from more general analytical claims to increasingly specific and targeted questions. The first three case studies adopt a mostly qualitative and interaction-oriented perspective on transcultural communication The fourth one differs in scope and purpose: it illustrates how the data collected and annotated within the NomadLingo corpus can

also be utilised for statistically oriented research, thereby extending its analytical reach beyond interactional and corpus studies.

Thus, the analyses here presented allow for an empirically grounded understanding of how linguistic fluidity operates in situated interaction, and how it reflects broader social, cultural, and mobility factors. They demonstrate the analytical potential of combining ethnographic sensitivity, fine-grained annotation, and corpus methods to investigate language as a social fact in post-global communicative ecologies.

All corpus extracts are reported anonymously, with turns attributed to speaker codes and/or nicknames, in line with the ethical and methodological principles outlined in Chapter 5.

6.2 Case study 1 – Digital nomad communities: A plurality of cultures, identities, voices

6.2.1 Standpoint and aims

This first case study examines how digital nomads linguistically construct their identities and their sense of belonging to the communities in which they participate. In Chapter 2 voluntary mobility and transcultural lifestyle were presented as grounding characteristics of digital nomad identity. This analysis seeks to find out how such dynamics become *evident* in language.

The analysis draws on three complementary data sources: (1) corpus materials, specifically <intro> turns (see Sub-section 5.2.2) and selected autobiographical talk transcripts; (2) metadata about participants and sessions (Appendices 7 and 8); (3) survey results which reveal broader patterns regarding mobility, identity, linguistic background, attitudes, and practices (see Sub-section 4.4.3 and Tedesco 2025b). Taken together, these sources converge in portraying digital nomad identity as trajectory-based, plurilingual, and situated within community discourse.

6.2.2 Case study 1: Data and analytical approach

The analysis integrates the following sets of data:

- NomadLingo_rawdata_participants.csv (Appendix 7) and NomadLingo1.1 participants.csv (Appendix 8);
- Corpus files (Tedesco et al. 2026): the analysis draws on <intro> segments and unannotated utterances retrieved through a concordance-based extraction procedure. A predefined list of semantically salient lexical items (based on theoretical insights as well as keyness and frequency of words contained inside the <intro> segments) related to mobility, biography, and language experience was used to systematically query the corpus. Concordance lines were then expanded to include surrounding turns, allowing for the identification of relevant interactional sequences for analysis;
- Survey indicators of identity complexity, mobility, language use and attitudes (Tedesco 2025b).

Overall, project resources derived from ethnographic field work (i.e., corpus and interviews) contain metadata about 74 participants representing 29 different nationalities and speaking 27 languages in total, with a median repertoire of 3 languages (sd = 1.33).

The analytical framework combines linguistic ethnography, mobility-centred sociolinguistic theories (Blommaert 2010; Blommaert & Rampton 2011; García & Wei 2014; see Chapters 2 and 3), and interactional sociolinguistics (Gumperz 1982). In this study, identity is conceived not as a stable property but as discursively emergent, accomplished through stance-taking, linguistic self-positioning, and references to collective nomadic practices (Silverstein 1993; Bucholtz & Hall 2005; Du Bois 2007; De Fina 2021).

6.2.3 Identity as / through mobility

A recurrent pattern across the corpus is that identity is anchored in movement. Speakers foreground their trajectories rather than their places of origin. Mobility thus becomes a central identity resource: biographies are narrated as sequences of moves.

This is particularly evident in extracts from introductory turns annotated as <intro> in the corpus (speakers answering the questions 'where are you from?' and 'where do you live?', asked through a flashcard located at the centre of the table). For example,

- in GC1_1:

Excerpt 1[28]

Emilia: I live here no **actually I don't know where I live** because I don't have a flat anymore so ((laughs))

- in F1:

Excerpt 2[29]

Felipe: I was in February the whole month in Thailand for example but before **I 've been only this year I 've been in** Prague Gran Canaria Madrid Barcelona (1.3) yeah and **I travel a lot** because yeah I leave my apartment for a few months every year

But also outside <intro> segments, mobility emerges spontaneously as a central theme.

For example,

- in M1:

Excerpt 3[30]

[28] Turn ID: M1_0045.

[29] Turn ID: F2_0028.

[30] Turn ID: from M1_0620 to M1_0635.

Leila: so I think I always wanted to like six months **six months out of the country and six months in Sweden** and I yeah so now I 'm in the process of trying to find a second home
Khalid: a balance?
Leila: yeah balance yeah I love Sweden but not to live there full time
Ana: were you in Stockholm or ehm?
Leila: I studied in Stockholm but I was born in Malmo
Ana: I have many friends in Malmo
Leila: yeah yeah
Ana: ‘ cause there 's like a- ' cause I live in Copenhagen
[…]
Leila: did you grow up in Copenhagen?
Ana: no **I grew up in Romania and then I moved** so I lived for the last eighteen years in Danmark

- in M7_3:

Excerpt 4[31]

Paulo: So **you decided to move here without coming here**?
Aksel: To Port- ehm to Madeira **yes but I 've been in Portugal for three years** (4.5) I love it I love Portugal and so far I love Madeira
Paulo: oh you ended up in an awesome place

These brief snippets index a stance: identity is constructed through movement (e.g., ‘I travel a lot’, ‘I grew up in Romania and then I moved’) and through the capacity to relocate and adapt (e.g., Aksel deciding to move to Madeira without knowing the place at all, or Emilia joking about the fact that she doesn’t know where she lives, because she doesn’t have a flat anymore). Through the presence of irony and assertive replies it can be understood that uncertainty is not framed as instability but as a core feature of nomad life – something the group recognises and ratifies (see Sub-section 6.2.5).

As a matter of fact, humour and self-irony frequently appear in <intro> turns or early conversation sequences, softening identity claims and fostering group alignment. Consider the following examples:

- from F2:

Excerpt 5[32]

Felipe: I 'm fluent in English I hope and I 'm fluent in Spanish for sure and I understand French but only understand and **I don't like talking French to be honest ((laughing))**
Elina: ah ok
Felipe: I 'm not I don't like **no I 'm not comfortable talking French ((laughing))** It 's a bit of a stretch and it 's **ehm I 'm shy yeah I mean I 'm shy ((laughing))**

- from GC1_1:

[31] Turn ID: from M7_0147 to M7_3_0149.

[32] Turn ID: from F2_0028 to F2_0030.

Excerpt 6[33]

Emma: ehm I speak: obviously Swedish ehm English and **a little bit S-Spanish un poco**
Ignacio: un poquito
Emma: y un poquito español
Petra: **un poquito ((chuckles)) nice** […] I moved to Las Palmas a month ago : so I 'm just exploring Las Palmas for now and I speak czech obviously ehm English and I just had my first class ((laughs)) **I just started with Spanish ((laughing)) so yeah I can't show off my Spanish as Emma but maybe next week ((laughing))** after my ((laughs)) after my second class

As will be further explored in Section 6.4, humour plays a key role in transcultural communication by making difference affiliative rather than divisive, as this conversation sample about an Italian and an American speaker negotiating the pronunciation of the word 'pistachio' shows.

From M1:

Excerpt 7[34]

Marco: and every time say me yes we have pistaccio and me **what is pistaccio**? pistacchio and the people say pistaccio it 's pistacchio it 's not pistaccio it 's not like ((laughs))
Mary: **we say pistachio** we 're like **sh** pistachio
Marco: but in Chicago pistachio? wow
Mary: [yeah]
Marco: **pistachio wow**
Mary: mh mh
Marco: and all the time also in thailand pistaccio ((laughs))
Mary: ((laughs)) it was for me wow when i- I was in San Marino last month and I had that pistachio liqueur it 's a creamy : pistachio : oh my god

Utterances as in the examples above index: I) a form of linguistic egalitarianism, challenging monolingual norms; II) transcultural alignment, because humour reduces social distance; III) community orientation, in the sense that affective reference to shared experiences like continuous relocation, remote-working issues with timezones and cultural shock reinforce community belonging – as will be further discusses in Sub-section 6.2.5.

6.2.4 Plurality in language resources as an identity trait

As already introduced through the extracts above, translingual self-descriptions serve to *perform* membership. For instance, in F2, Marco replies to the question 'what languages do you speak' with the following answer:

Excerpt 8[35]

Marco: Italian a little bit of English and un poco de Español ((laughs))

[33] Turn ID: from GC_1_0221 to GC_1_0225.

[34] Turn ID: from M1_0204 to F2_0214.

[35] Turn ID: F1_0502.

Plurilingualism itself becomes an identity marker. Speakers often highlight their plurilingual repertoires, aligning with the community's plurilingual nature and projecting an identity of openness and adaptability. Moreover, story-telling about language learning and linguistic repertoires often conveys positive views on plurilingualism, as in the example below.

From F2:

Excerpt 9[36]

Louise: **I 'm half English half French so I grew up bilingual** anyway
Mateo: ok but-
Louise: so **it makes you more likely to speak other languages** I think
Mateo: that 's right studies say that
Clara: yeah yeah yeah but **how did you learn Chinese** for example?
Louise: oh **I didn't I wanted to learn Arabic** but-
Mateo: me too I wanted to learn Arabic when I was younger

As reported in 6.2.2, corpus metadata describe all participants as plurilingual, and survey results (see Chapter 4) confirm this pattern:

- the vast majority of participants use more than two languages daily;
- 57% attribute language learning to immersion in plurilingual environments;
- 61% report informal learning through *travel*.

Thus, in digital nomad communities, languages are perceived as necessary resources, a mobility asset. For instance, in a conversation extracted from M7_2, a participant (identified by the nickname Eliise in Excerpt 10) values competence in Spanish as gate-opening to Latin America travel:

Excerpt 10[37]

Samuel: what language do you want to learn?
Eliise: the next one for me improve Portuguese because I really like Portugal
Samuel: but something new?
Eliise: no
Samuel: no? so you 're ok? with those?
Eliise: **I speak German English Estonian Russian now Portuguese and Spanish not really like Spanish for sure but like Spanish is very useful that 's why I wanna travel in Latin America**
Samuel: yeah South America
Eliise: yes
Samuel: so you need to go to Brazil I 'm telling you you 're gonna have a lot of fun
Eliise: but right now everybody is talking about Brazil I don't know

[36] Turn ID: from F1_0362 to F1_0368.

[37] Turn ID: from M7_1_0167 to M7_1_0178.

Within the communities under study plurilingualism constitutes a symbolic resource enhancing mobility capital, social access and community membership. This last dimension will be now discussed.

6.2.5 Performing community membership

Speakers rarely declare to be digital nomads explicitly. Instead, they perform belonging by referring to practices, infrastructures and lifestyle routines typical of nomad hubs.

Lexical items which help speakers perform community membership through digital nomad discourse include references to shared community rituals, as exemplified by Excerpt 11.

From M1:

> *Excerpt 11*[38]
>
> Leila: yeah but I'd rather like do something else other than driving I think **the pizza thing** is a great idea

Here, the general noun *thing* and the article *the* work as deictics referring to shared knowledge about community life. As a matter of fact, through common keywords extraction (performed on SketchEngine which allows to focus the analysis on common or rare words; see Sub-section 5.6.2), the general nouns *place* and *thing*, both appear within the first 30 keywords of NomadLingo, and *the* appears within their most frequent collocates.

6.2.6 Summing up: Digital nomad identity as relational and co-constructed

Identity does not emerge solely from self-descriptions but also from how interlocutors *ratify* and categorise themselves and each other. Summing up, digital nomad identity emerges in the dataset as:

> a) mobile and trajectory-based, because speakers routinely foreground where they have been and where they are going;
>
> b) community-enacted, since belonging is performed through shared infrastructures and practices, including plurilingual/translingual speech;
>
> c) plurilingual and adaptive, because linguistic repertoires appear central to identity display and social alignment;
>
> d) relational and co-constructed, with humour, stance, and interaction shapingrecognition and community membership

Corpus, metadata, and survey evidence thus reveal a dynamic, mobility-driven identity model. The examples were here presented in a descriptive manner, with the aim of showing how

[38] Turn ID: M1_1010.

language can be investigated to highlight sociological cues. In doing so, the analysis prepares the ground for the case Studies presented in the remainder of Chapter 6.

6.3 Case study 2 – Nomad English: Beyond English as a lingua franca

6.3.1. Reconsidering ELF through mobility

This case study examines how English is used by digital nomads in ways that destabilise early conceptualisations of English as a lingua franca (ELF, see Chapter 3). Rather than functioning as a simplified medium, as sometimes assumed in early ELF scholarship (e.g., Jenkins 2007), English emerges in the NomadLingo corpus as a porous linguistic resource, part of a translanguaging repertoire which is shaped by mobility trajectories and transcultural encounters (Tedesco forth.).

The goal here is not to define a new ELF variety, but to show that the communicative practices of digital nomads challenge theoretical models that treat languages as stable, bounded systems tied to homogeneous groups. Instead, FLO data reveal English as:

- embedded in plurilingual repertoires;
- permeable to other languages;
- highly variable across contexts, speakers and activities;
- used creatively;
- indexical of belonging, mobility, and lifestyle alignment;
- constitutive of what participants' mobility capital, not simply a tool for intelligibility.

Grounded in the corpus, survey data, and ethnographic insights, and in a view of *languaging* as an active ongoing process of meaning-making (Love 2017; see Chapter 3), this case study advances an empirical view on ELF as an activity, a resource for transcultural alignment, a capital for social access, and a site for cultural brokering.

6.3.2 Analytical framework: Integrating variation, repertoires and interaction

The analytical framework used in this case study integrates five perspectives as interdependent interpretative layers through which the use of English within the communities under study becomes intelligible, namely: (i) variationist sociolinguistics (Tagliamonte 2012; Di Garbo 2023), (ii) interactional sociolinguistics (Gumperz 1982; Coupland 2007); (iii) ELF studies (Nero 2024), (iv) translingual approaches (Canagarajah 2012); (v) language ideologies, attitudes and affordances (Woolard 2020).

Following Labovian traditions, variation is understood as situated and socially meaningful, rather than as deviation from a norm (Campbell-Kibler 2010; Eckert 2016). However, given the superdiverse nature of digital nomad communities (discussed in Chapters 2 and 3), linguistic variants cannot be interpreted as a property of a stable and identifiable demographic group; on the contrary, variation is treated as an indexical practice; its analysis includes

considerations about aspects of the corpus like metalinguistic reflections, stance-taking, interactional features (e.g., alignment), as well as situational factors and speakers' repertoires. The empirical approach to variation proposed here thereby considers the communicative event, with all its variables, as the analytical unit. Through reference to various sets of ethnographic data, i.e. survey results, self-ethnography and interviews (described in Chapter 4), linguistic features observed in the corpus are also linked to underlying language ideologies, to explore the ways in which the English language becomes the setting for negotiating identity and community access.

Before presenting the results of the analysis, it should be noted that the BNC (Love et al. 2017), the COCA (Davies 2008), and the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2005) have been used as reference corpora for analyses related to variation and varieties, notwithstanding the limits of their mutual comparability[39].

6.3.3 Linguistic variation as index of mobility, transcultural repertoires and community building

Variation manifests itself at multiple levels. First, the corpus includes the use of different English varieties, reflecting speakers' educational backgrounds, previous residence, and exposure to global media. Second, both intra-speaker and inter-speaker variation are observable across phonological, lexical, grammatical, and syntactic domains, often interacting with translanguaging practices.

For instance, the corpus includes lexical items associated with specific English varieties. An example of American English is the expression *down the road* (by the speaker AM_M_01, grown-up in Chicago), used to mean 'in the future', which shows a higher normalised frequency in the COCA (10.86) than in the BNC (7.8). Similarly, the Canadian English term 'snowbird' (by the Canadian speaker CA_M_01), referring to individuals who migrate seasonally to warmer places, reflects mobility practices that resonate strongly with digital nomad lifestyles translated into a cultural-specific term. These examples illustrate how speakers mobilise variety-specific lexical resources that index both geographical trajectories and shared experiential knowledge.

Intra-speaker variation provides further evidence of this flexibility. Intra-speaker variation refers to the alternation between different linguistic forms by the same speaker across or even within the interactional sequences (Campbell-Kibler 2010; Eckert 2016). An illustrative example concerns the lexical alternation between *flat* and *apartment*. Reference corpora reveal that *apartment* is considerably more frequent in American English than in British English, while *flat* shows the opposite distribution (see Table 17). Speaker CZ_C_02, who resides in the UK and completed his entire education in England, might be expected to consistently use the British English term *flat*. Instead, after initially using *flat*, he also produces the utterance 'like having an *apartment* like somewhere warm'.

[39] https://www.english-corpora.org/coca/help/compare-bnc.asp.

Table 17. Frequency of *apartment* and *flat* in reference corpora.

Lemma	COCA norm. freq.	BNC norm. freq.
apartment	65.88	16.58
flat (N)	15.56	23.64

The alternation between *flat* and *apartment* within the same speaker's production raises questions about the factors influencing lexical choices. To explore possible factors, a closer look should be given to the conversation. Indeed, in GC1_2 we can notice that CZ_C_02, a Czech-born British-educated speaker uses the word *apartment* until the word *flat* is used by his interlocutor (ES_C_04). After ES_C_04 uses the word *flat* for the first time, CZ_C_02 alternates between the two variants. Thus, possible explanations are alignment with interlocutors' lexical choices, lexical priming effects (Pickering & Garrod 2004), or an interaction of the two.

Moreover, also inter-speaker variation emerges when comparing speakers' repertoires. An example concerns the alternation between *arrive* (Latin etymology) and *come* (Proto-Germanic etymology) among speakers with and without Romance languages in their linguistic repertoire. As shown in Table 18, speakers who know Romance languages display a higher relative use of *arrive*, while the normalised frequency of *come* is similar across both groups. The ratio between the raw frequency of *arrive* and *come* in one of the reference corpora used in this study, namely the BNC (shown in the last column of the table), seems to confirm that *come* is much more frequent than *arrive* in British English.

Table 18. ARRIVE vs. COME across speaker groups.

Group	Sub-corpus size	ARRIVE (raw)	COME (raw)	ARRIVE/COME ratio
Romance language	64,573	21	74	0.28
No Romance language	12,986	2	16	0.12
BNC	112,338,376	494	31,826	0.051

This pattern suggests that lexical preferences may be subtly influenced by cross-linguistic transfer, as well as being mediated by interactional and contextual factors.

The corpus also contains non-standard grammatical variants, for example in the domain of adjectival comparison, specifically the use of analytic comparative forms (e.g. more + adjective) in contexts where synthetic comparatives are conventionally expected in standard English (e.g. -er forms). These forms are produced by speakers with diverse linguistic backgrounds, such as SE_C_03 (Swedish), IT_MC_01 (Italian), and TW_M_01 (Taiwanese), as illustrated in Table 19.

Table 19. Examples of comparative forms extracted from NomadLingo.

Speaker	***Instance***
SE_C_03	*But it's* ***more healthy*** *for you.*
IT_MC_01	*I prefer I prefer because is* ***more fast***
TW_M_01	*Sicily is* ***bigger***

While in Italian *più* functions as a general comparative marker which may affect the preference to use *more* instead of the suffix *-er*, the presence of this non-standard structure across speakers with different L1s suggests that such forms cannot be reduced to transfer alone. Other morphosyntactic phenomena observed include prepositional variation with verbs of movement (e.g., recurrent alternation between *went + in* and *went + to*), third-person singular marker omission, and relative pronoun simplification, patterns previously described in ELF research (Mauranen 2012). For example, the relative frequency of relative pronouns (*which, who, that*) in NomadLingo is 129.56, much lower than the relative frequency of the same pronouns in the Santa Barbara Corpus of Spoken American (198.69).

A final observation concerns the relationship between linguistic repertoire size and variation. While it is difficult to quantify variation exhaustively, what can be quantified is translanguaging density in relation to the number of languages spoken by each speaker. Using the corpus in CSV format (as downloadable from the ILC4CLARIN repository), translanguaging was operationalised as the number of annotated translanguaging events (*trans_type_count*). To control for unequal participation and interaction length, translanguaging was consistently normalised by the number of tokens attributable to the same analytical unit. This yields a density-based measure of translanguaging that allows meaningful comparison across speakers, sessions, and contexts.

At the speaker level, translanguaging density was examined in relation to the number of languages known by each participant. For each speaker, translanguaging events were summed across all their turns and normalised by the total number of tokens they produced. The number of languages known was treated as a stable individual attribute and retrieved once per speaker.

Despite the intuitive expectation that broader repertoires might correlate with increased translanguaging, the analysis revealed no meaningful association between repertoire size (i.e., number of languages spoken) and translanguaging density. This suggests that translanguaging cannot be straightforwardly predicted from the size of a speaker's linguistic repertoire, indicating that natural translingual practices are not competence-driven, but rather result from situationally emergent practice. [40]

[40] By stating that 'natural translanguaging practices are not competence-driven' I do not aim to say that plurilingual comptence cannot be trained, but rather that in a context where no training regarding plurilingual communication and translanguaging was given to participant (as in the case of the NomadLingo corpus), the occurrence of translanguaging phenomena is not correlated to the size of speakers' repertoire, and other situational as well as

To further explore factors which could affect the presence of translanguaging phenomena, translanguaging density was calculated for each session by normalising the total number of translanguaging events occurring in that session by the total number of tokens produced within it. Given the categorical nature of sessions, no correlation analysis was performed. Instead, translanguaging was examined distributionally across sessions. This analysis is presented in greater detail in Sub-section 6.3.5.

However, looking at speakers' repertoires within the same session, it was noticed that in sessions with higher translanguaging density speakers often share more languages than in sessions with lower translanguaging. In the latter sessions, plurilingual repertoires, still present, tend to overlap less.

Finally, translanguaging density was analysed in relation to speaker nationality. To avoid distortions caused by unequal participation, translanguaging was first normalised at the speaker level and then aggregated by nationality.

While the resulting distribution shows heterogeneity across national groups, high within-group variance and uneven group sizes suggest that in the dataset under analysis nationality does not function as a causal or predictive variable. Instead, observed differences among speakers likely reflect the interactional roles speakers occupy, the contexts in which they participate, and the specific session ecologies in which they are embedded, besides obvious idiosyncratic differences.

6.3.4 Translinguistic variation and lexical richness

One of the advantages of a corpus like NomadLingo is that it allows for an expansion of traditional variationist perspectives by adopting a translanguaging lens. From this viewpoint, translations and translinguistic alternatives can be understood as options for expressing similar meanings through different linguistic resources. Rather than treating such alternatives as deviations, this approach conceptualises them as part of speakers' lexical capabilities across repertoires.

The negotiation of the word *pistachio* in Excerpt 7 exemplifies how phonological variation intertwines with humour and meaning negotiation. Similarly, the alternation between *filme* and *movie* presented below illustrates how speakers exploit translinguistic variants to convey meaning and align with their interlocutor(s), as shown in the extract from F2 involving Portuguese and English variants of the same concept (i.e., film).

From F2:

Excerpt 11[41]

---

biographical factors – not all of which are captured by the ethnographic data gathered for the project – may be more influential. For a deeper discussion on the differences in translingual practices between trained and non-trained participants see (Tedesco et al. forth.).

[41] Turn ID: from M2_0399 to M2_0409.

Francesca: when you watch a **movie** that is an English movie
Lorella: but say it in Portuguese come on speak Portuguese
Olly: eh yeah speak Portuguese
Francesca: yeah but-
Marco: let's try this game ((laughs))
Olly: no she can do it
Francesca: voce assistci a um **filme** eu falo un po' di mh brazileiro ((chuckles))
Amalia: yeah eu adoro português do brasil
[…]
Francesca: e: quando você assiste a um **movie** a um **filme** é dublado?

Lexical richness, in this perspective, refers not merely to the size of a monolingual lexicon, but to speakers' ability to draw on semiotic resources traditionally conceived as part of different systems to construct meaning collaboratively. Let's consider a further example from F1:

Excerpt 12[42]

Alain: how do you call this when you do this like **washing money** how do you call that?
((inaudible portion))
Louise: **blanqueamiento [de capitales]**
Mateo: [de capitales no?]
Louise: de capitales
Luigi: ma l'avete letta su internet questa cosa qua? *((laughs))*
Alain: **blanchiment blanchiment de l'argent**
Louise: c'est pas un mot que je dis beaucoup en français *((laughs))*
Alain: *((inaudible portion in French))*
((laughing))
Maja: ok so three languages for everything
Alain: ok *((laughs))*
Alain: how do you say that ehm in Spanish?
Mateo: **blanquea- blanqueo de capitales**
((Inaudible portion))
Maja: y en catalan?
((laughs))
Louise: **blanqueamen de diners** ((laughs)) no está mal
Arnaut: ((laughs))
Adele: But are you Catalan?
Louise: no I'm half English half French
Adele: ah ok and why do you know do you speak Catalan?
Louise: I lived in S- in Barcelona for eleven years-

When translinguistic variants co-occur within the same conversational unit, as in collaborative translation sequences like the ones above, meaning emerges through metalinguistic reflection. These interactions blur language boundaries and foreground speakers' reflexive engagement with language. Moreover, instances like these exemplify language socialisation practices, which can be variously observed through corpus and ethnographic data.

[42] Turn ID: from F1_0606 to F1_0625.

By extracting non-standard English terms from the corpus (through keywords, as explained in Sub-section 5.5.3), some recurrent lexical fields can be identified, as shown in Table 20.

Table 20. Lexical items and semantic fields in translanguaging.

Lexical item	English translation	No. hits in the corpus	No. of texts	Semantic field
pico	peak	12	2	sport & activities
carbonara	carbonara pasta	12	3	food & eating
levada	a walking trail alongside old water channels	11	2	sport & activities
pistacchio	pistachio	11	2	food & eating
parmigiana	eggplant lasagna	8	2	food & eating
salute	cheers	6	2	food & eating
madeirense	madeiran	6	2	places
mercadona	[the name of the local supermarket]	5	2	places
bruschetta	toasted bread	5	1	food & eating
gelato	icecream	5	2	food & eating
vino	wine	4	3	food & eating
bozen	Bolzano / Bozen (name of a city in italy)	3	1	places
surfear	to surf	1	1	sport & activities

The extracted lexical items cluster around a small number of recurrent semantic fields. Terms related to food and eating, sport and outdoor activities are particularly frequent, reflecting situated everyday practices and shared experiences within digital nomad communities. A smaller set of items referring to places functions as a localising resource, anchoring interaction to specific cultural and geographical contexts. Translingual lexical richness thus appears to be patterned rather than random, emerging especially in domains characterised by high cultural salience and experiential relevance.

6.3.5 Speakers' agency, linguistic awareness and sociolinguistic ideologies

Survey results (described in Chapter 4) and corpus data converge in revealing highly positive attitudes toward language learning and plurilingual practices. On average, participants express a desire to use less English and more local or additional languages in social interactions, and most report engaging in translanguaging practices frequently. Importantly, participants do not generally fear negative evaluation of their English, either by peers or by language teachers.

82% of survey respondents self-identify as bilingual or plurilingual, a finding consistent with ethnographic observations within the FLO project. Survey responses to questions about creative language use reveal pride in linguistic creativity, through invented words, hybrid expressions, and playful calques. For instance, these are some of the answers to a question asking participants about their own *language mixing* practices:

1) ‘I do invent words when I cannot come up with the correct expression or if it isn’t part of my vocabulary but I don’t remember any in particular now’;
2) ‘I am in a international relationship (german/italian). We mostly speak English, but we mix all the time in the same sentence all 3 languages + we created our own expressions based on sound similarities/errors one made in the other lnguage (eg. I made a bistecca = I made a mistake)’;
3) ‘Hello together! (German people when they want to say "hello everyone") We see us! (Also German people when they want to say "See you!")’.

These examples thus illustrate linguistic creativity as a valued practice (Tedesco 2025b).

Natural language data in the corpus confirm the positive attitudes towards plurilingualism revealed through survey and ethnographic participation, as shown in the samples from the corpus reported in the present section as well as in Section 6.2 (e.g., Excerpt 9 where Louise and Mateo discuss about the positive effects of bilingualism on language learning and cognition).

Qualitative corpus analysis further reveals metalinguistic reflections on pronunciation, dialects, and lexical choices, often embedded in humorous or affiliative sequences. By xtracting concordance lines of *say, language, pronunciation*, metalinguistic reflections can be retrieved from the corpus. Consider the following samples:

Excerpt 13[43]

Emma: **it 's a dialect thing** if you ask her she will say something different but from the south where I 'm from gudnatt
Lotte: gudnatt

or, from M2, a Tawainese speaker commenting about her difficulties in learning English:

Excerpt 14[44]

Olly: that's why it's really hard for us to learn English yeah because it's not our language
Amalia: **this is for your vocals**
Olly: yes

Metalinguistic reflections in the corpus often take the form of transcultural discussions, as in the interaction reported below, where a Portuguese / Brazilian couple explains the meaning of their dog’s name to an Estonian nomad friend.

Excerpt 15[45]

[43] Turn ID: GC1_0505 and GC1_0506.

[44] Turn ID: from M2_0856 to M2_0858.

[45] Turn ID: from M7_1_0295 to M7_1_0310.

Samuel: it was bella but then Matild's dad said it was a good idea to have like some reference of the place and bulga is a cute name I think it's cute it matches and also **pulga pulga** with the p it's like
Ana: mmm paras- como fue parasitas?
Samuel: parasites
Ana: **parasites** parasites they're in the dogs and they-
Samuel: fleece they fleece
Ana: **like scratching so that's called pulga in Portuguese**
Eliise: [i know yeah] it's like a pet box
Ana: yeah something like that
Eliise: yeah I know what you mean
Samuel: but **in Portuguese it's a cute word** that thing is pulga you can call your you can call your girlfriend pulga you know
Eliise: oh like my pulga like my parasite ((laughs))
Ana: just in Portugal or like you can call like a kid pulga
Eliise: yeah but like in brazil?
Ana: no in brazil you don't call anyone pulgas

As already pointed out, interactions also demonstrate language socialisation processes (Duranti et al. 2012), where speakers jointly negotiate meaning, norms, and acceptability, like in the extract below, from M2:

Excerpt 16[46]

Francesca: like for example Recife ah ah recife **I say Recife**
Amalia: we can say-
Francesca: they say in Brazil **they say Hecife**
Amalia: Recife ehm I- eu fui
Francesca: **they say Recife** they uh uh aspira eh
Amalia: ehm depende
[…]
Amalia: **depende da: da localidade**
Francesca: hm hm no la erre però
Amalia: na Europa também esiste essa diferença […] I can dizer mais forte and I can dizer mais suave
Francesca: ah ok ok ok

A theme emerging from both corpus data (through evaluation and stance-taking) and informal interviews conducted during fieldwork (see Chapter 4) is the tension between normative orientations towards English (which may signal the persistence of hidden monolingual ideological frameworks) and the positive attitudes towards plurilingualism and acceptance of translingual practices. On the one hand, through attitudes, stance-taking and language usage English consistently appears as the main and necessary language for communication within digital nomad communities; on the other, speakers display positive attitudes towards linguistic creativity and the use of multiple languages. An interview with Fabi, creator of the Remote Work Community Network[47] (Tedesco 2024a), confirms English as the key linguistic capital to access a nomadic lifestyle, while also emphasising the value of plurilingualism and reporting

[46] Turn ID: from M2_0866 to M2_0878.

[47] The interview is readable in full length at nomadlingo.eu/blog.

frequent language mixing in everyday interactions. When asked whether English is essential for digital nomads his answer is a firm yes. But then he adds, 'In the Nomad Cruise the organisers were always asking to speak English. They were afraid that we would start speaking German preventing the engagement of people from other countries', also demonstrating a view on English as a connecting resource. Moreover, Fabi underlines that 'at the same time it is not nice if you are in a place, let's say here [Fuerteventura], for six months and do not speak a word of Spanish', emphasising the value of plurilingualism and then reporting frequent language mixing in everyday interactions. This tension becomes visible in interactional practice, where BrE/AmE standards are stressed by speakers in a way that shows their consciousness about normative ideologies and their humorously challenging attitudes towards them, as in this example:

Excerpt 17[48]

Martin: where did you get **the cutlery** guys?
Eva: here
Emilia: oh here so enjoy everybody
Eva: enjoy
Carlos: it 's **so good to hear like proper English** where did you guys have the cutlery? ((laughing))
Martin: oh bang it
Carlos: **chaval**

Carlos's remark *it's so good to hear like proper English*, produced with laughter and followed by the Spanish lexical item *chaval* (an idiomatic way of addressing a friend, could be translated in English into *man*), simultaneously indexes awareness of a monolingual norm and playfully challenges it.

Thus, interactions like the one above, together with more serious conversations about positive aspects of plurilingualism present in the corpus, suggest that the coexistence of normalising and disruptive language ideologies is a defining feature of digital nomad linguistic attitudes. It is likely for this reason that translanguaging occurs more frequently in private contexts of interaction, where reduced audience design, which often co-occurs with more homogeneous groups and shared awareness of participants' linguistic repertoires, and lower risk of negative evaluation allow speakers to challenge monolingual norms more openly while maintaining social alignment (as shown in Figure 64).

[48] Turn ID: from GC1_2_0001 to GC1_2_0008.

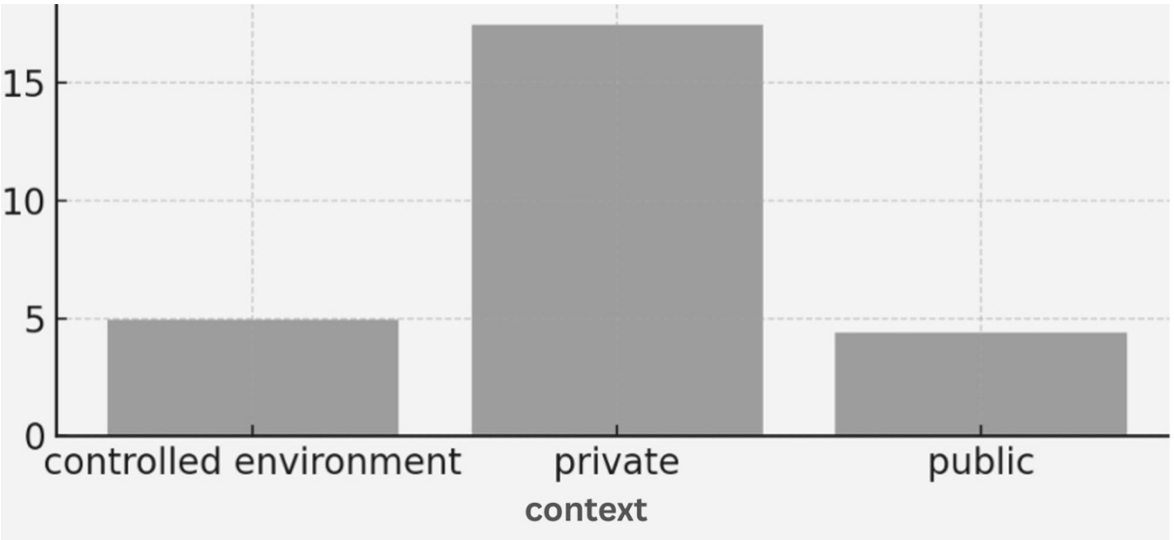


Figure 64. Translanguaging occurrences according to context type.

In this figure, translanguaging is examined across interactional contexts (private, semi-public, public), following the method outlined in Sub-section 6.3.3. For each context, translanguaging events were normalised by the total number of tokens produced within that context, yielding context-specific translanguaging densities.

Translanguaging is present in all sessions, but its density varies substantially across interactional ecologies. Private or semi-private social settings (e.g. birthday celebrations, shared meals in coliving spaces) display the highest translanguaging density, while more public or institutionally framed encounters (e.g. coworking-related events) exhibit lower density. A clear gradient emerges: private contexts exhibit the highest translanguaging density, semi-public contexts show intermediate values, and public contexts display markedly lower densities. This distribution supports the interpretation of translanguaging as a situated, context-sensitive practice.

Sessions that show higher translanguaging density relative to their token count tend to share the following features:

- informal social framing (e.g. dinners, chats, small-group encounters);
- private or semi-private settings, often characterised by speakers who are more familiar with each other than in public contexts;
- participation by all speakers and shorter turns;
- high repertoire overlap, often involving closely related languages (often Romance ones, given the recording settings).

In these sessions, translanguaging seems to function as an interactional resource embedded in humour, stance-taking, and narrative sequences. It is rarely marked or repaired. In other words, translanguaging is interactionally routinised, not exceptional.

The sessions showing less frequent translanguaging density typically involve:

- public or semi-public settings (e.g. open events, coworking-related activities);
- larger or less stable participant constellations.

Importantly, translanguaging is still present but it is interactionally constrained, often limited to brief lexical insertions and explicitly marked.

In general, as could be expected, the results show that translanguaging correlates with partial overlap of speakers' repertoires, and with typologically or historically related languages co-present in the session. Figure 65 shows translanguaging per session.

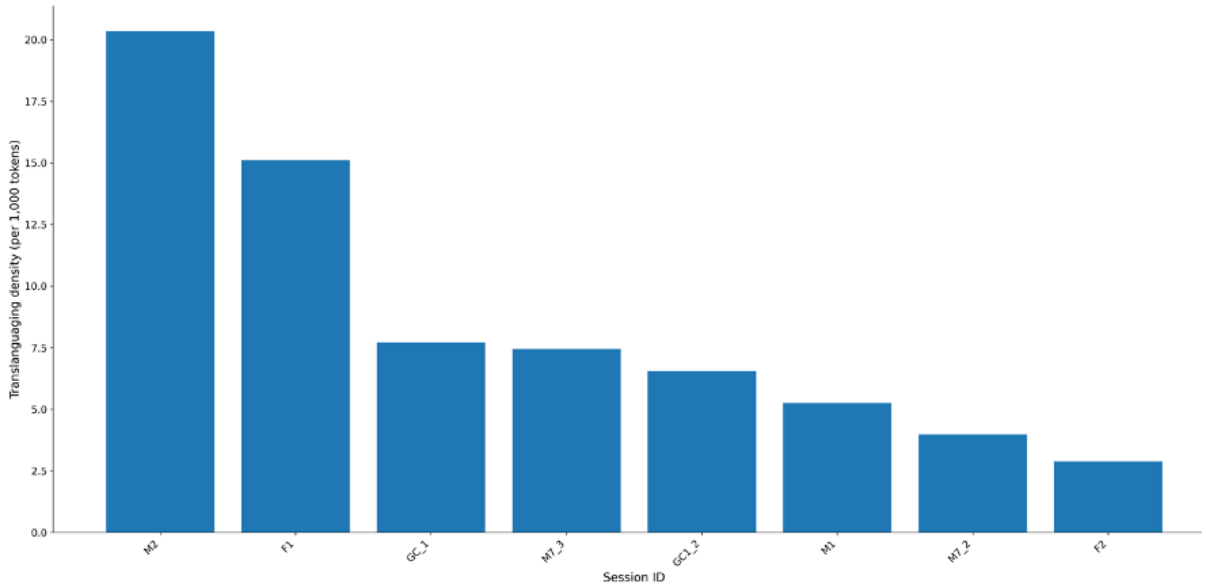


Figure 65. Distribution of translanguaging tags per session.

All sessions contain translanguaging, with a weighted average of 69.82 translanguaging tags per session. However, as already discussed and visible from the Figure, density and interactional role vary substantially. The sessions with the highest translanguaging density are the birthday party in Câmara de Lobos (M2) and the private dinner in Fuerteventura (F1), followed at some distance by the dinner at Coliving Kite and Surf (GC1_1). These sessions are characterised by a strongly informal framing, private or semi-private settings, and high interactional density, with frequent storytelling, humour, and affiliative talk. In these interactional ecologies, translanguaging operates as an unmarked and widely shared resource through which speakers routinely construct meaning, negotiate stance, and build affiliation, rather than as an exceptional communicative practice.

Sessions such as the happy hour at Coworking in Ponta do Sol (M7) and the picnic in Funchal (M1) show moderate translanguaging density. Here, translingual phenomena are consistently present but more locally contained, typically emerging in specific sequences such as jokes, cultural references, or meaning negotiations following moments of misunderstanding. These sessions occupy hybrid interactional ecologies, combining social interaction with a partially public audience design.

The lowest translanguaging density is observed in Nomad Chats (F2), the only interaction which was based around a brainstorming conversation task (more information are found in Chapter 5). In this session, translanguaging is systematically constrained, though never absent. Turns tend to be longer and interactional patterns more structured, English as a lingua franca emerges as the default communicative means, and higher interactional monitoring can be observed.

6.3.6 English in translanguaging

The distribution of translanguaging tags across sessions demonstrates that, while interactional variables such as context, participation framework, and audience design influence how and how often translanguaging occurs, no session is characterised by strictly monolingual communication. Rather than functioning as a bounded and autonomous code, English consistently emerges as a permeable and expandable resource.

The English observed in the corpus is therefore best described as an enriched form of English, shaped by the constant availability of multiple linguistic resources. Lexical insertions, like *pulga* (Excerpt 15) or *escarbadientes* (Excerpt 18), hybrid constructions like *I can dizer mas fuerte* (Excerpt 16), collaborative translations, as in the case of the *money washing* sequence (Excerpt 12), and metalinguistic reflections do not represent deviations from a norm, but constitute routine communicative practices.

These findings align with a growing body of research that challenges monolingual and code-segregated models of language use. Studies in translanguaging and translingual communication have repeatedly shown that speakers do not alternate between discrete linguistic systems, but draw on integrated repertoires to make meaning (e.g., García & Wei 2013; Canagarajah 2013; Wei 2018). From this perspective, translanguaging is not an exceptional strategy, but the default mode through which plurilingual speakers communicate. The NomadLingo data strongly support this view: translanguaging is not restricted to speakers with particular language profiles, nor to specific languages; it is rather distributed across sessions, speakers, and interactional contexts.

Crucially, FLO datasets also reveal that these practices are not merely tolerated but interactionally ratified and socially valued within digital nomad communities. As previously discussed, translanguaging is frequently associated with humour, stance-taking, affiliation, and collaborative meaning-making. Rather than being treated as a problem or repaired as a communicative failure, it is often left unmarked or used as a resource for alignment and inclusion.

In this sense, English operates simultaneously as a lingua franca and as a co-creation space: it provides a common ground for interaction, while remaining open to enrichment from other languages. The data thus challenge models of ELF that foreground simplification, reduction, or convergence, and instead point towards a view of English as an adaptive, situated and co-constructed resource. The absence of strictly monolingual communication across the corpus, together with the absence of monolingual speakers within survey participants, reinforces the

argument that, in these communities, translanguaging is not how speakers sometimes communicate, but how communication regularly takes place.

6.3.7 Linguistic capital and social affordances: A culture of acceptability, a model of accessibility

The analysis presented in this section shows that English within the NomadLingo corpus functions simultaneously as a resource for accessibility and a site of negotiated acceptability. English enables participation in highly mobile and heterogeneous communities and provides access to interactional spaces that would otherwise remain inaccessible. In this sense, it reflects what Pennycook describes as the widespread perception of English as 'a key to access modernity, progress, and social mobility' (2017: 7).

However, the empirical evidence demonstrates that accessibility does not imply monolingual communication. Across all sessions, English is systematically embedded within plurilingual and translanguaging practices. While contextual variables such as setting, interactional framing, and audience design influence translanguaging density, the corpus provides no evidence of strictly monolingual interaction. Variation appears as gradient and interactionally regulated rather than categorical.

This points to a strong culture of acceptability within digital nomad communities. Variation from standards and translanguaging phenomena are rarely treated as problematic and are instead integrated into humour, stance-taking, alignment, and meaning negotiation. In line with Canagarajah's observation that transcultural communication is 'much more multilingual, multimodal, and polysemic' than dominant ideologies suggest (2017: 13), English in the corpus emerges as permeable, flexible, and socially inclusive. Documenting such fluid uses of English can contribute to: (i) understanding language use and variation in post-global societies; (ii) challenging monolingual ideologies through empirical evidence; and (iii) informing usage-based language teaching models that reflect the actual communicative practices of post-global generations.

This perspective prepares the ground for the following case studies, which examine negotiation sequences (Case Study 3, i.e. Section 6.4) and fluency/disfluency modelling (Case Study 4, i.e. Section 6.5).

6.4 Case study 3 – Negotiation and misunderstanding in transcultural communication

6.4.1 Misunderstanding: A space for meaning co-construction in transcultural interactions

The analysis here presented focuses on sequences where participants negotiate meaning, repair misalignments, or use irony to reframe or soften emerging trouble. Based on the annotation categories described in Chapter 5, the analysis illustrates several misunderstanding sequences and how these are resolved through multimodal cues, contextual knowledge, and translingual practices. A key insight is that misunderstandings do not index communicative failure; rather,

they are productive sites for establishing epistemic access, recalibrating roles, and co-constructing membership. Irony, humour, and playful reframing emerge as resources for maintaining face and solidarity, especially when cultural knowledge is unevenly distributed.

6.4.2 Analytical framework for the analysis of misunderstanding sequences

The analysis draws on the annotated version of the NomadLingo corpus. Misunderstanding sequences were identified through the annotation scheme described in Chapter 5, which includes tags for misunderstanding, interactional moves (e.g. asking, clarification requests), repair segments, and translanguaging occurrences.

Methodologically, the analysis is informed by conversation analysis, particularly the principle of *why that, why now?,* which treats interaction as locally organised and sequentially meaningful (Schegloff & Sacks 1973; Schegloff 2007). However, unlike classical conversation analysis, this study does not abstract interaction from its sociolinguistic and experiential context. Speakers' repertoires, their mobility trajectories, and shared / differing cultural backgrounds are treated as analytically relevant resources. In this sense, the analysis aligns with interactional sociolinguistics and pragmatic approaches to naturally occurring data, where meaning is understood as emergent from the interaction between linguistic form, contextual knowledge, and participant orientations (Haugh 2013).

6.4.3 Signalling trouble and initiating negotiation

Across the corpus, interactional trouble is rarely marked by overt breakdowns. Instead, misunderstanding is typically signalled through subtle interactional cues such as non-linguistic sounds (e.g. 'I'm a gynecologist / mh? / gynecologist / oh ok), delayed responses, or short clarification requests, like 'what?' or 'sorry?'.

Given the informal nature of the data, it is not surprising that interactional moves tagged as *asking* constitute the most frequent type of interaction among the ones annotated, as informal talk often relies on implicit coordination rather than explicit meta-communication (Couper-Kuhlen & Selting 2018).

Clarification requests often take minimal and non-threatening forms, allowing speakers to flag non-understanding without interrupting the flow of interaction. Turns tagged as *interactional_move type = asking* have an average length of 3 tokens with the most frequent word being *what*, which occurs 30 times in these segments, and *mh / ehm* occurring 12 times. Non-signalled misunderstanding sequences, identified during manual annotation through attention to communication breakdowns, suggests a high tolerance for partial understanding and a shared orientation toward collaborative meaning-making, as in the examples above.

Importantly, understanding is rarely treated as the responsibility of a single speaker. Instead, misunderstandings trigger collective repair strategies, in which multiple participants contribute explanations, examples, translations, or reformulations. This distributed reformulation work reflects the communal orientation of the interaction and reinforces the idea that understanding is jointly achieved rather than individually owned. The following extract, in which speakers

are reflecting on the word *escarbadientes* (meaning *toothpick* in English), constitutes a good example of such meaning-making practices.

Excerpt 18[49]

Emma: **how do you say** this again?
Ignacio: escarbadientes
Maria: escabardientes **what's that?**
Adela: escabardientes
Ignacio: you don't know? **esto palillo**
Maria: ah palillo
Emma: that's easier
Ignacio: yeah but but ehm **escarbadientes means the: the action of like take off something of your: teeth** like escarbar dientes </translanguaging>
Emma: ah take off fish from your teeth ((laughs)) and you think he is like a farmer saying that?
Maria: maybe he is
Ignacio: es como que escarbadientes es como que **en argentina las cosas les pones un nombre segun lo que hacen so in argentinian we put the name to the things depends the action**
Emma: mh mh
Ignacio: so like the action is the object like the action of this is the object in argentina escarbadientes because escarva el diente
Lotte: **you get that in German a lot**
Adela: but actually **it says that in English as well it's a tooth picker**
Emma: tooth picker yeah
Adela: yeah it's a picker for your tooth
Lotte: but it's **the same in German**
Emma: the little stick **how do you say it in German?**
Lotte: **zahnstocher**
Ignacio: what?
Lotte: zahnstocher
Ignacio: zahnstocher
Lotte: but it means the same as in English so it's also
Emma: tooth picker
Lotte: tooth picker yeah
Ignacio: zahnstocher zahnstocher
((laughing))
Lotte: that was very good
Ignacio: thank you
((laughing))
Lotte: **so how do you say it in swedish?**
Emma: in the south dialect or in the-?
Lotte: both
Emma: **tandpetare**
Adela: which is also an explanation of what it is
Emma: yeah it's a tooth picker as well

This extract provides a clear illustration of collaborative meaning-making through translanguaging, where participants do not merely search for lexical equivalents but actively reflect on how meaning is constructed across languages. Particularly salient is Ignacio's metalinguistic explanation of *escarbadientes*, where he explicitly decomposes the word into its semantic and functional components ('the action of like take off something of your teeth').

[49] Turn ID: from GC1_1_0452 to GC1_1_0488.

This explanation shifts the interaction from translation to metalinguistic analysis, prompting other speakers to draw parallels with other languages.
Through this process, meaning is co-constructed across languages via comparison, reformulation, and shared reflection, rather than resolved through a single target-language equivalent. The sequence thus exemplifies how translanguaging practices foster metalinguistic awareness, enabling speakers to negotiate meaning by foregrounding the semantic motivation of words while simultaneously reinforcing social alignment through humour and mutual engagement.

6.4.4 Cultural assumptions, peer translation, and narrative re-keying

Many misunderstanding sequences are rooted in culturally specific knowledge, including food practices, local customs, linguistic conventions, or idiomatic expressions. Rather than being treated as problematic, these moments frequently become opportunities for cultural explanation and narrative expansion. Excerpt 15 where Samuel, Ana and Eliise discuss the meaning of the Portuguese word *pulga* and its use in the European and Brazilian varieties, triggered by a clarification request about the dog's name, is an example of such mechanisms of meaning co-construction. Narrative re-keying is particularly visible in sequences where a misunderstanding prompts a shift from linguistic talk to storytelling, as in the extract presented in Sub-section 6.3.4 where the introduction of the Catalan language leads Louise to talk about her life experience in Barcelona. What begins as a clarification request evolves into the recounting of an anecdote, thereby re-framing interactional trouble as an opportunity for community construction and emotional bonding.

Peer translation plays a crucial role in this process. Participants often translate not only lexical items but also cultural meanings, re-keying narratives so that they become accessible to interlocutors with different backgrounds. These translations and reformulations involve paraphrasing, exemplification, and analogy, often supported by gestures or laughter. Consider the collaborative translation of the Spanish word *escarbadiente* reported above, or of the term *washing money* presented in Case study 2 (Sub-section 6.3.4). In both cases, as additional languages are brought into the interaction, the sequence shifts from clarification to playful exploration, culminating in shared laughter and explicit positive evaluations.

6.4.5 Irony and humour as alignment and misalignment resources

Irony and humour are pervasive in misunderstanding sequences and serve multiple interactional functions. Indeed, in the 82 misunderstanding sequences present in the corpus, laughter annotations appear 44 times. Rather than being approached as autonomous pragmatic strategies, they are here analysed here as emergent responses to interactional trouble, following moments of partial understanding or epistemic imbalance.

Humour and irony can soften potential face threats, index stance, and negotiate alignment or mild misalignment without escalating conflict. In several corpus examples, ironic comments follow moments of partial misunderstanding, allowing speakers to acknowledge difference while maintaining social cohesion. Crucially, irony does not erase asymmetries in knowledge

but renders them interactionally manageable. Through playful reframing, speakers can display awareness of difference while reaffirming mutual recognition and belonging. Humour also enables speakers to temporarily suspend normative expectations about linguistic accuracy or cultural competence. By laughing at misunderstandings, participants frame them as shared experiences rather than individual shortcomings.

Besides misunderstanding sequences, the corpus also displays reflections about misunderstandings. The following is a transcript of a conversation extract not included in the current version of the corpus, where an Italian speaker humorously narrate his misunderstanding of the word *morador*:

Excerpt 19

Marco: After seen is free the park
Daniela: seven or eight? Seven?
Marco: mh mh
Daniela: so you're goo
Bogdan: I paid until 6.15 so I should be good
Marco: well, don't pay
Bogdan: maybe because in my country if you overstay five minutes you know that you're going to get the ticket in Belgrade-
Daniela: it's monitored
Bogdad: well it's not monitored or anything
Daniela: oh good ((chuckles)
Bogdad: it's just people checking like what do you call them-
Marco: but in Italy you pay a lot […] The first time I come to Madeira I park in Funchal center where there was a sign only for morador
Daniela: ((laughs)) with the M on the: on the:
Marco: eh yeah but I don't know what is it then ok park I parked I was like boh 10 minutes I go I park I take my things I return back to the home ehm was the police with the ehm paper-
Paulo: [fine]
Marco: I say no perdona perdona it's only dieci minutos non lo se que es morador the police start to scream **esto è solo para morador ((laughs)) e io non lo so que es morador porque in Italia morador is like muratore ((laughs))**
((laughing))

In this extract, Marco mixes Spanish, Portuguese, Italian and English to explain that he had misunderstood the Portuguese word *morador* (trnsl. resident) with the Italian *muratore* (trnsl. construction worker). This tendency to laugh about misunderstanding pairs with observations from fieldnotes and interviews where participants explicitly report positive attitudes toward linguistic experimentation and non-standard language use.

6.4.6 Misunderstanding as an interactional resource in transcultural communication

This case study points out that misunderstandings in transcultural interaction are structurally embedded moments of negotiation. They reveal how participants mobilise linguistic, pragmatic, and interpersonal resources to sustain interaction, manage diversity, co-construct meaning and emotional bonds.

The findings support a view of communication as fundamentally collaborative and adaptive. (Translingual) reformulations are central resources in misunderstanding sequences, as shown by the frequent occurrence of reformulation and translanguaging tags within misunderstanding sequences. When trouble arises, speakers mobilise a variety of semiotic resources to re-express meanings, clarify intent, or scaffold understanding. These practices blur boundaries between languages and challenge monolingual assumptions about communication.

Across the case study, irony is understood not as a communicative goal, but as a by-product of collaborative meaning-making, arising when participants successfully transform interactional trouble into shared understanding and affiliation. Irony and translanguaging thus emerge as key mechanisms through which speakers navigate cultural assumptions and maintain alignment. These practices are particularly salient in informal, low-risk interactional settings, like the ones represented in NomadLingo, where participants share a strong orientation toward inclusivity and mutual understanding (Schlegloff 2007).

Speakers draw on their repertoires to recalibrate who knows what, who has the right to explain, and how knowledge circulates within the group. In doing so, misunderstanding becomes a moment of epistemic reorganisation, where access to meaning is negotiated and redistributed.

Overall, this case study reinforces the argument advanced in previous sections of the Chapter: transcultural communication among digital nomads is not characterised by convergence toward a single norm but by flexible negotiation, relational work, and a shared acceptance of communicative fluidity.

6.5 Case study 4 – Exploring the cognitive load of translanguaging through statistical modelling of disfluencies

6.5.1 Introduction: A usage-based perspective on translanguaging and speech (dis)fluency

Although fluency is widely regarded as a prerequisite for successful communication, it remains a relatively under-theorised dimension of language research, where analytical attention has traditionally focused on linguistic competence rather than spoken performance (Hymes 1972). Within such perspectives, fluency is often treated as an unmarked, default outcome of rule-following behaviour, rather than as a dynamic achievement emerging from real-time language production (Segalowitz 2010; McCarthy 2010). By contrast, research on speech production has increasingly foregrounded disfluency, manifested through hesitations, pauses, repetitions, false starts, and self-repairs, as a central window onto the cognitive and interactional processes underlying spoken language use. However, despite growing empirical attention, both fluency and disfluency remain conceptually vague constructs, characterised by terminological ambiguity and theoretical heterogeneity (Lickley 2015).

Disfluencies are not merely errors or performance failures. Rather, they constitute meaningful linguistic phenomena that index planning difficulty, lexical access, monitoring processes, and

interactional negotiation (Clark & Wasow 1998; Fox Tree 1995). From this perspective, disfluency provides valuable insights into cognitive load during speech production, as well as into speakers' adaptive strategies for maintaining fluency under communicative pressure. Empirical work has shown that increased cognitive load is systematically associated with higher rates of disfluency, particularly in contexts involving multitasking, complex planning, or linguistic challenge (Lindström 2025).

These insights are particularly relevant in transcultural contexts. A growing body of work suggests that bilingual and plurilingual speakers exhibit higher baseline levels of disfluency, especially in non-dominant languages, reflecting increased processing demands and reduced automaticity (Eggers et al. 2017; Eggers 2021; Byers-Heinlein et al. 2022). At the same time, translanguaging practices have been shown to function as adaptive resources that mitigate foreign language anxiety, support meaning-making, and foster communicative fluency, particularly in educational and interactional settings (Mauranen 2006; Thu Ha et al. 2021). This dual positioning of translanguaging as both potentially cognitively demanding and communicatively enabling raises questions about its relationship with speech disfluency.

Against this backdrop, the present case study adopts a quantitative approach to explore how translanguaging relates to disfluency in naturally occurring spoken interaction. The analysis builds directly on the findings of Case study 2, where the distribution of translanguaging phenomena across speakers, contexts, and interactional settings was mapped, and potential sociocultural predictors of translanguaging use were explored. While Case study 2 treated the issues of *where* and *by whom* translanguaging occurs, the present study shifts the analytical focus to *how* translanguaging unfolds at the level of speech production, and whether it is associated with increased or reduced cognitive load as indexed by disfluency.

Specifically, this case study operationalises translanguaging phenomena and disfluency measures as quantitative variables and investigates their statistical relationship using mixed-effects regression modelling. As will be discussed, preliminary results indicate that translanguaging does not incur a general fluency cost once relevant covariates are taken into account. On the contrary, translanguaging appears to co-occur with differentiated disfluency profiles, suggesting a reconfiguration rather than a simple increase of production difficulty. For some speakers and contexts, translanguaging is associated with higher fluency, reinforcing theoretical accounts (ref) that conceptualise translanguaging as an enabling communicative resource rather than as a compensatory strategy triggered by linguistic deficiency.

6.5.2 Operationalisation of disfluency

Disfluency refers to any interruption or irregularity in the normal flow of speech. Such interruptions are often unintentional, but they may also fulfil communicative and interactional functions depending on contextual, pragmatic, and speaker-related factors. Far from being marginal phenomena, disfluencies are an inherent component of spoken language and have long been studied for what they reveal about cognitive processes, interactional organisation, and linguistic competence (Lickley 2015).

Within psycholinguistics and discourse-oriented research, disfluencies are understood as informative linguistic markers that provide insight into planning processes, monitoring mechanisms, and speakers' adaptive strategies in real-time language production (Clark & Wasow 1998; Fox Tree 1995). In this study, given the absence of phonetic alignment and fine-grained prosodic annotation in the NomadLingo corpus, the operationalisation relies on orthographically transcribed features that have been shown to reliably index disfluency in corpus-based research. This choice aligns with a growing body of work demonstrating that systematically annotated surface features can provide robust proxies for underlying cognitive and interactional processes when applied consistently and transparently (Götz 2013; Gilquin 2016; Ma et al. 2025).

Three main groups of disfluency phenomena are targeted in the present study: impediments, repairs, and discourse markers. These categories are motivated both theoretically and empirically and are treated as analytically distinct dimensions of disfluency in subsequent modelling (see Section 6.5.6).

Impediments capture prosodic and morphological obstacles that affect the continuity of speech planning and sentence structuring. The following phenomena are classified as impediments:

- Filled pauses, uniformly transcribed in the corpus as *ehm* (see Chapter 5);
- Sound lengthening, transcribed using the symbol ":", excluding cases occurring at the end of a turn, where lengthening may signal emphasis or turn-final prosody;
- False starts and truncated words, excluding truncations occurring at turn end, which are more likely to signal interruption by another speaker;
- Repetitions of single words, bigrams, and trigrams, occurring within a limited token span.

False starts and repetitions partially overlap with repairs both conceptually and operationally. However, while repairs were manually annotated in the corpus and typically involve larger stretches of speech, repetitions were automatically retrieved through a Python script developed for this case study and included in the project's GitHub repository[50]. The script identifies (i) immediate word repetitions, (ii) repetitions within a three-token window, and (iii) repeated bigrams and trigrams within the same span. This distinction allows impediments to capture local disruptions in planning and lexical access, while preserving repairs as a separate category linked to monitoring and correction.

Impediments are widely interpreted as indicators of increased cognitive load during speech production, particularly during lexical retrieval and syntactic planning (Goldman-Eisler 1968; Levelt 1989; Clark & Fox Tree 2002). For this reason, they are treated in this study as a primary index of disfluency caused by high level of cognitive load.

[50] https://github.com/novella-tedesco/FLO/blob/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026/Case_study4_Tedesco_2026.ipynb.

Repairs refer to self-initiated corrections, restarts, and reformulations produced after initial articulation. Following established work in conversation analysis and psycholinguistics (Schegloff et al. 1977; Levelt 1983), two macro-types of repair are distinguished in the corpus annotation (see Sub-section 5.5.2 for examples):

- Restarts, where speakers abandon an ongoing utterance and begin again, often due to hesitation, re-planning, or difficulty retrieving linguistic material;
- Reformulations, where speakers rephrase or adjust a prior utterance to improve precision, appropriateness, or intelligibility.

Repairs reflect post-articulatory monitoring processes rather than planning difficulty per se. They therefore index a different stage of speech production, operating at the interface between cognitive evaluation and interactional management.

Discourse markers are included as a third disfluency-related dimension, following research that highlights their frequent co-occurrence with hesitation and their relevance for interactional organisation (Schiffrin 1987; Blanchard & Buysse 2019). Discourse markers are defined here as pragmatic items that structure discourse, manage turn-taking, signal stance, or maintain interpersonal alignment. While they may appear in disfluent speech, they are not inherently indicative of processing difficulty. Based on Schiffrin's framework (1987, 2012) and further applications to corpus-based studies (Aijmer 2002; Buysse 2012), the following discourse markers are included: *so, well, you know, I mean, like, really, actually,* and *yeah* (the latter when used as a filler rather than as a backchannel or confirmation). Like repetitions, these occurrences are retrieved automatically using a Python script. Finally, translanguaging segments were manually analysed to look for non-English discourse markers like the Italian *cioè* or the Spanish *pues*.

In line with previous work on fluency and speech production, disfluency is thus treated here as a multidimensional phenomenon. Impediments, repairs, and discourse markers correspond to different functional stages of speech production and interaction. Impediments primarily index cognitive load during planning, repairs reflect monitoring and correction processes, and discourse markers relate to pragmatic and interactional organisation. Collapsing these phenomena into a single undifferentiated measure risks obscuring theoretically relevant patterns. For this reason, the present study adopts a stepwise analytical strategy, as discussed in Sub-section 6.5.4.

6.5.3 Case study 4: Dataset

The analyses presented in this case study are based on the .csv version of the NomadLingo corpus (Tedesco et al. 2026). The unit of analysis adopted throughout this study is the turn, defined as a continuous stretch of speech produced by a single speaker without interruption. Each row of the dataset corresponds to one turn and includes information on speaker identity, linguistic repertoire, interactional context, translanguaging phenomena, and of disfluency and linguistic complexity indicators.

The dataset used for the statistical models comprises 7,226 turns (i.e., datapoints) produced by 47 speakers. The version of the dataset employed for Case study 4, together with all scripts used to derive variables not originally present in the CSV corpus file, is available in the project's GitHub repository[51]. This includes all codes used for the computation of lexical diversity, content complexity, repetition-based impediments, and discourse marker indices, ensuring transparency and full replicability of the analytical pipeline.

The dataset is composed of two worksheets: the main one, used in the model, and the disfluencies worksheets, containing the different elements counted to calculate disfluencies phenomena. The main dataset includes the following columns:

- session_ID;
- context;
- speaker_ID;
- speaker_nationality;
- speaker_languages;
- num_langs(Plurilingualism);
- turn_ID;
- turn_length;
- lexical_diversity(STTR);
- content_complexity(mean_ortographic_word_length);
- transcript;
- translanguaging_any;
- codeswitching;
- codemixing;
- translation;
- repairs;
- discourse_markers;
- impediments;
- disfluencies.

Lexical diversity was operationalised using the Standardised Type–Token Ratio (STTR), a length-controlled measure widely adopted in studies of spoken discourse and learner language (Malvern et al. 2004; McCarthy & Jarvis 2010). STTR captures the ratio between the number of unique lexical items and the total number of tokens, while controlling for the well-known sensitivity of basic Type–Token Ratio (TTR) to text length. In the present study, lemmas were used as counting units (i.e. types), while tokens correspond to all running words. This choice allows lexical diversity to be assessed independently of inflectional variation.

[51] Python scripts can be found here: https://github.com/novella-tedesco/FLO/blob/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026/Case_study4_Tedesco_2026.ipynb. The dataset can be downloaded from: https://github.com/novella-tedesco/FLO/blob/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026/dataset_Casestudy4_tedesco_2026.xlsx.

Lexical diversity was computed separately for each turn following a conditional procedure based on turn length. For segments containing fewer than ten tokens, lexical diversity was calculated using the basic TTR:

$$\mathrm{TTR} = \frac{\text{Number of unique lemmas}}{\text{Total number of tokens}}$$

For segments containing ten tokens or more, a moving-window approach was applied, following the procedure implemented in *lexically.net* and described in Malvern et al. (2004). Each segment was divided into overlapping windows of ten consecutive tokens (tokens 1–10, 2–11, and so forth). For each window W_i, a local Type–Token Ratio was computed as:

$$\mathrm{TTR}_{W_i} = \frac{\text{Number of unique lemmas in } W_i}{10}$$

The final STTR value for each segment was obtained by calculating the arithmetic mean of the TTR values across all windows:

$$\mathrm{STTR} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{TTR}_{W_i}$$

where n corresponds to the total number of ten-token windows in the segment. No additional scaling or normalisation was applied. This procedure yields a length-controlled measure of local lexical diversity that is particularly suitable for short, turn-based units in spoken interaction and has been widely used in fluency and disfluency research (Malvern et al. 2004; McCarthy & Jarvis 2010)[52].

Lexical diversity was included as a control variable because previous research has shown that richer lexical variation is systematically associated with fluency and reduced disfluency, both in first and additional language production (McCarthy & Jarvis 2010). In spoken interaction, higher lexical diversity has been linked to more automatised access to lexical resources, while reduced diversity often co-occurs with hesitations, repetitions, and other markers of processing difficulty (Biber et al. 1999; Tavakoli & Skehan 2005; Skehan 2009).

In addition to lexical diversity, content complexity was included as a further linguistic control variable. Content complexity was operationalised as mean orthographic word length, defined as the average number of characters per word within each turn. This measure has been used as a surface-based proxy for lexical and processing complexity in spoken discourse, with longer words generally associated with increased planning demands and higher cognitive load (Biber et al. 1999; Skehan 2009). Including this variable allows the model to control for the possibility

[52] see also https://lexically.net/downloads/version7/HTML/type_token_ratio_proc.html.

that disfluency patterns are driven not by translanguaging per se, but by increased lexical or conceptual complexity within turns.

Lexical diversity (STTR) and content complexity were included in the models to control for linguistic factors that are known to influence fluency and disfluency independently of interactional or sociolinguistic practices. By accounting for these variables, the analyses aim to isolate the specific contribution of translanguaging phenomena to different dimensions of disfluency, while avoiding confounds related to lexical richness or processing demands intrinsic to the content of speech.

6.5.4 The method: A mixed-effect linear regression model to predict disfluencies based on translanguaging phenomena

The method reported in this section is based on a hierarchical, multilevel modelling framework, implemented through linear mixed-effects regression models. This approach is particularly well suited to the analysis of spoken interaction data, where observations are inherently non-independent and structured across multiple levels, such as turns nested within speakers and speakers embedded within interactional contexts. Unlike traditional linear regression, mixed-effects models allow for the simultaneous estimation of fixed effects, capturing systematic relationships between predictors and outcomes at the population level, and random effects, accounting for structured variability across units such as individual speakers.

This modelling framework has become standard in linguistics and the social sciences for the analysis of language production data, where speaker-specific baselines and repeated observations are the norm rather than the exception (Baayen 2008; Baayen et al. 2008; Gelman & Hill 2007). In interactional and sociolinguistic research, mixed-effects models are now widely recognised as a prerequisite for valid statistical inference (Barr et al. 2013), as they allow researchers to model speaker-to-speaker variability and contextual heterogeneity. The models and their implementation are adapted from Ma et al. (2025) and available from the Github and OSF repositories[53].

The analysis was conducted in three progressive stages, each addressing a distinct but complementary research question. Importantly, the models do not constitute competing analyses; rather, they form a cumulative analytical sequence, moving from global patterns to increasingly fine-grained, functionally and theoretically motivated distinctions. An overview of the analytical progression is provided in Table 21.

Table 21. Overview of analytical models for measuring effects of translanguaging on disfluencies.

[53] Link to Github file: https://github.com/novella-tedesco/FLO/blob/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026/Exploring%20the%20cognitive%20load%20of%20translanguaging%20through%20statistical%20modelling%20of%20disfluencies.R. OSF project's DOI: 10.17605/OSF.IO/UK9NY.

Model	Focus	Dependent variable(s)	Key predictor(s)	Analytical goal
Model 1	Global effect of translanguaging on speech disfluency	General disfluency index (in the dataset called 'disfluencies')	Translanguaging (any)	Establish baseline effect
Model 2	Functional differentiation of disfluency types	Impediments, Repairs, Discourse markers	Translanguaging (any)	Identify disfluency-specific effects
Model 3	Specific effects of different translanguaging practices	Impediments	Codeswitching, Codemixing, Translation	Test specific hypothesis about impediments as possible triggers of translanguaging

This staged approach minimises collinearity between predictors, maximises statistical power in the global analysis, and enables a transparent interpretation of both aggregate and type-specific effects. Consistent with the data-driven orientation adopted throughout this thesis, the formulation of Models 2 and 3 as well as the research questions they address was informed by the empirical results obtained from Model 1, rather than being specified a priori. This iterative modelling strategy allows the analysis to remain theoretically grounded while responsive to data structure and characteristics.

To examine the overall relationship between translanguaging and disfluency, Model 1 is fitted with general disfluency (i.e., the sum of impediments, discourse markers and repairs) as the dependent variable. The analysis focuses on the effect of translanguaging presence, operationalised as a binary predictor, while controlling for linguistic, speaker-related, and contextual factors known to influence fluency and disfluency in spoken interaction. All analyses included random intercepts for speakers to account for repeated observations. No random slopes are included, as the primary goal of Model 1 was to establish a robust population-level baseline rather than to model inter-speaker variability in translanguaging effects.

Building on the baseline analysis, Model 2 examines whether translanguaging is differentially associated with distinct functional dimensions of disfluency, reflecting both cognitive and interactional aspects of speech production. While Model 1 treats disfluency as an aggregated construct, Model 2 disaggregates disfluency into three theoretically motivated categories: impediments**,** repairs, and discourse markers, as discussed in 6.5.2.

Separate linear mixed-effects models were fitted for each disfluency type, retaining translanguaging presence as the main predictor of interest and including the same set of linguistic, speaker-related, and contextual control variables as in Model 1.

Building on the results of Model 2 and their statistical significance (see Sub-section 6.5.6), Model 3 focused specifically on impediments as the dependent variable. This choice is also

theoretically motivated: unlike repairs and discourse markers, which predominantly reflect post-articulatory monitoring and interactional management, impediments (e.g. hesitations, cut-offs, sound lengthening) are widely interpreted as indicators of increased cognitive load during speech planning and lexical selection. Impediments therefore constitute the most plausible candidates for triggers of translanguaging, rather than its discursive consequences. Model 3 examined whether the different translanguaging strategies (codeswitching, codemixing, and translation), are differently associated with cognitive disfluency.

Translanguaging strategies were operationalised as separate binary predictors. This feature-based representation allows for a fine-grained assessment of the differential cognitive profile of individual translanguaging phenomena without imposing mutually exclusive categorical labels. Random intercepts for speakers were retained to account for repeated observations.

6.5.5 Model 1 results: Translanguaging, plurilingualism and disfluency

As shown in Table 22, translanguaging presence does not exert a statistically significant effect on the general disfluency index ($\beta = -0.07$, SE = 0.06, $p = .245$). Although the coefficient is negative, indicating a tendency towards lower disfluency in translanguaging turns, this effect does not reach statistical significance.

Table 22. Fixed effects estimates for Model 1 (Translanguaging on general disfluency).

Predictor	Estimate (β)	Std. Error	t	p
(Intercept)	5.13	0.24	21.44	< .001
Translanguaging (any)	−0.07	0.06	−1.16	.245
Context: private	−0.03	0.07	−0.43	.668
Context: public	0.14	0.08	1.70	.088
Plurilingualism	0.07	0.03	2.89	.006
Turn length	0.11	0.001	116.26	< .001
Lexical diversity (STTR)	−5.85	0.16	−37.37	< .001
Content complexity	0.01	0.01	1.02	.308

Several control variables display robust and theoretically expected effects (see, for example, the recent study by Ma et al. 2025). Turn length emerges as a strong positive predictor of disfluency ($\beta = 0.11$, $p < .001$), indicating that longer turns are more prone to disfluent phenomena. Lexical diversity (STTR) shows a strong negative effect ($\beta = -5.85$, $p < .001$), suggesting that turns with richer lexical variation tend to be more fluent. Plurilingualism is positively associated with disfluency ($\beta = 0.07$, $p < .01$), indicating higher baseline disfluency among speakers with more complex linguistic repertoires. This effect is consistent with a substantial body of psycholinguistic research showing that bilingual and plurilingual speakers often exhibit increased rates of hesitation phenomena, particularly in spontaneous speech production (Hartsuiker & Notebaert 2010; Szmrecsanyi 2014; Ma et al. 2025).

Importantly, this pattern should not be interpreted as evidence of reduced linguistic competence. Rather, previous studies have shown that bilingual speech production involves the simultaneous activation of a wide range of resources, which increases demands on lexical selection and control mechanisms (Green 1998; Kroll 2017). As a consequence, plurilingual speakers may experience more frequent moments of lexical competition and planning difficulty, which are commonly manifested as hesitations, pauses, or other forms of disfluency (Bialystok 2009; Gollan & Ferreira 2009). In this sense, the present result can be interpreted in favour of translanguaging theories which assume linguistic repertoires as holistic and interconnected.

Furthermore, empirical evidence has shown that disfluencies tend to be particularly salient in spontaneous interaction, understood as interaction that unfolds without pre-planning and under real-time processing constraints (Eggers et al. 2017; Eggers 2021). While comparing different interactional settings would be beyond the aims of the present study, the data analysed here are characterised by precisely these conditions, as speakers must continuously manage language selection, monitor interlocutor alignment, and adapt to shifting communicative demands in situ.

Neither context nor content complexity show a statistically significant effect on general disfluency. Differences across speaker nationalities are observed, which may be due to cultural cognitive frameworks (Giles & Hewstone 1982) or random; full nationality coefficients are reported in the supplementary materials which can be found both on Github and OSF[54].

Although speaker nationality was initially included to control for macro-social variation, its contribution to the model proved limited and unsystematic. While a small number of categories reached statistical significance, no coherent pattern emerged that could be meaningfully interpreted in linguistic or typological terms, and the effects were substantially weaker than those of variables such a[55]s turn length, lexical diversity, and plurilingualism. In line with the theoretical orientation of this study, which prioritises practices, repertoires, and trajectories over static sociolinguistic categories, speaker nationality was therefore excluded from subsequent models to avoid overparameterisation and to focus on more significant variables.

Overall model fit statistics are reported in Table 23, indicating that the model explains a substantial proportion of variance at the fixed-effects level (marginal $R^2 = .706$).

Table 23. Model fit statistics for Model 1.

Metric	Value
Observations	7,226
Speakers	47
Marginal R^2	0.706

[54] https://github.com/novella-tedesco/FLO/blob/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026/Supplementary_Tables_Translanguaging_Disfluency.pdf.
[55] Full coefficients are reported in the Supplementary Materials available on the GitHub repository of the project (see link in the preceding note).

Metric	Value
Conditional R^2	0.709
AIC	25,538
BIC	25,745

The results of Model 1 thus provide an initial, aggregate picture of the relationship between translanguaging practices and speech disfluency, while also highlighting the role of other influential variables. At a global level, the presence of translanguaging does not emerge as a significant predictor of increased disfluency, suggesting that the use of multiple linguistic resources within a turn does not, in itself, compromise fluency in spoken interaction.

At the same time, Model 1 highlights the relevance of speaker-level characteristics, most notably plurilingualism, which shows a positive association with baseline disfluency. As discussed above, this effect is best interpreted in terms of increased cognitive and interactional demands associated with managing more complex linguistic repertoires, rather than as evidence of reduced communicative competence.

Crucially, however, Model 1 treats disfluency as a unitary construct. While this provides a necessary baseline, it risks obscuring functionally distinct phenomena. Disfluency encompasses heterogeneous processes ranging from planning-related hesitations to post-articulatory monitoring and pragmatic discourse management, which are unlikely to be uniformly associated with translanguaging phenomena.

For this reason, the analysis proceeds in Model 2 by disaggregating disfluency into three theoretically motivated dimensions – impediments, repairs, and discourse markers – and examining their relationship with translanguaging separately.

6.5.6 Model 2 results: The relationship between translanguaging and different disfluency types

Three separate mixed-effects models were fitted, all retaining the same fixed- and random-effects structure as Model 1, including random intercepts for speakers. Results are summarised in Table 24.

Table 24. Effects of translanguaging on disfluency types (Model 2).

Dependent variable	Estimate (β)	Std. Error	t	p
Impediments	0.12	0.04	3.07	.002
Repairs	0.01	0.01	1.18	.237
Discourse markers	−0.21	0.04	−4.84	< .001

Translanguaging shows a positive and statistically significant association with impediments ($\beta = 0.12$, SE = 0.04, $p = .002$), indicating that translanguaging turns are characterised by a higher incidence of hesitation phenomena typically associated with increased cognitive load. In contrast, translanguaging does not significantly predict repairs ($\beta = 0.01$, SE = 0.01, $p = .237$),

suggesting no systematic link with overt self-monitoring or correction processes. Finally, translanguaging is associated with a significant decrease in discourse markers ($\beta = -0.21$, SE = 0.04, $p < .001$), suggesting reduced reliance on pragmatic devices for discourse organisation in translanguaging turns.

Model 2 reveals a functionally stratified pattern: translanguaging is associated with increased cognitive disfluency (impediments) but reduced pragmatic scaffolding (discourse markers), while leaving repair behaviour unaffected. This pattern suggests a reconfiguration, rather than a simple increase / decrease of disfluency when translanguaging is involved.

The positive association between translanguaging and impediments should not be interpreted exclusively as an effect of translanguaging on speech production. An alternative theoretically plausible interpretation is that impediments may function as cognitive precursors of translanguaging. From this perspective, hesitation phenomena may index moments of intensified lexical search or conceptual restructuring that precede and potentially trigger translanguaging, rather than resulting from it (see Sub-section 6.5.8). This interpretation aligns with psycholinguistic accounts that view filled pauses and truncations as indicators of real-time processing difficulty (Goldman-Eisler 1968; Levelt 1989; Clark & Fox Tree 2002), as well as with interactional approaches that conceptualise translanguaging as a resource mobilised to resolve such difficulties.

At the same time, the observed reduction in discourse markers suggests that once translanguaging is activated, speakers may rely less on conventional pragmatic signalling devices to structure their turns. Translanguaging may thus partially substitute for discourse markers by enabling speakers to maintain progressivity through cross-linguistic resource mobilisation rather than through explicit interactional signalling. However, the steady difference between translanguaging and *English only* turns, in terms of both appearance in the corpus and turn length, should be considered as a possible bias which prevent a clear comparative interpretation around the use of discourse markers in translanguaging vs. standard English turns.

Overall, Model 2 demonstrates that the relationship between translanguaging and disfluency is functionally differentiated. This finding motivates a more fine-grained analysis that distinguishes different types of translanguaging and examines their specific cognitive and interactional correlates.

6.5.7 Model 3 results: Impediments as indexes of cognitive load and/or triggers of translanguaging

Building on Model 2, Model 3 focuses specifically on impediments, conceptualised as a core indicator of cognitive load during speech production. The model examines whether different translanguaging phenomena – codeswitching, codemixing, and translation – are differentially associated with impediments, the only disfluency parameters which increase in a statistically significant way when translanguaging is present. The model includes the same control variables as Models 1 and 2, as well as random intercepts for speakers. Results are reported in Table 25.

Table 25. Fixed effects of Model 3: Translanguaging strategies predicting impediments.

Predictor	Estimate (β)	Std. Error	t	p
Codeswitching	0.05	0.05	0.85	.396
Codemixing	0.20	0.06	3.46	< .001
Translation	0.26	0.15	1.65	.099
Turn length	0.05	0.001	79.36	< .001
Lexical diversity (STTR)	−4.93	0.10	−49.21	< .001

Controls for context, speaker nationality, plurilingualism, and content complexity are omitted for brevity; full results are reported in the Supplementary Materials.

Codemixing shows a positive and statistically significant effect on impediments (β = 0.20, SE = 0.06, $p < .001$). Codeswitching displays a positive but non-significant effect (β = 0.05, SE = 0.05, $p = .396$), as well as translation which shows a positive trend that does not reach conventional levels of statistical significance (β = 0.26, SE = 0.15, $p = .099$).

This graded pattern suggests that translanguaging practices may systematically differ in their relationship with cognitive disfluency. Strategies involving tighter cross-linguistic integration within a single syntactic frame (codemixing) are more strongly associated with impediments than strategies relying on cross-turn language alternation (codeswitching) or translinguistic reformulation (translation)**.**

Two complementary interpretations are compatible with these findings. First, codemixing may be cognitively demanding due to the need to integrate multiple linguistic systems during online planning. Unlike codeswitching, which typically involves language alternation at clause or turn boundaries, codemixing requires the simultaneous activation and integration of multiple linguistic systems within a single syntactic structure. This increased processing demand may manifest as impediments such as hesitations, false starts, or filled pauses, which function as indicators of momentary planning difficulty. Alternatively**,** impediments may precede and trigger codemixing, rather than resulting from it. From this perspective, impediments can be understood as early signals of disruption in speech planning, such as difficulties in lexical retrieval or syntactic formulation, which prompt speakers to mobilise resources from multiple languages as an adaptive response. Under this interpretation, codemixing does not necessarily reflect a default multilingual stance or a stable ideological positioning, but rather a locally emergent strategy activated under cognitive pressure.

The quantitative models employed here do not allow for direct causal inference regarding the interaction of impediments and translanguaging elements. The observed associations should therefore be interpreted as evidence of a systematic coupling between codemixing and cognitive disfluency, compatible with both (i) a planning-based account and (ii) a precursor-based account. In particular, Model 3 revealed a graded relationship between translanguaging strategies and impediments, with codemixing showing the strongest association, but such a pattern is compatible with multiple theoretically plausible interpretations. Further

investigations could examine the presence of impediments in preceding turns by the same speakers, as suggested also in the follow-up qualitative analysis presented in 6.5.8.

6.5.8 A closer look: Qualitative analysis of the relationship between speech impediments and translanguaging

The results of the mixed-effects models suggest a complex relationship between translanguaging strategies and impediments, raising questions about the directionality, temporal organisation, and functional role of the translanguaging phenomena visible in the corpus.

To address these questions, a qualitative follow-up analysis was designed to explore how translanguaging and impediments are sequentially organised across turns, and how different translanguaging strategies are embedded in specific interactional dynamics. This qualitative component is not intended to validate the statistical models, but rather to contextualise and refine the quantitative patterns observed.

Impediments were selected as the primary disfluency phenomenon for qualitative exploration because they are widely interpreted as indicators of increased cognitive load during speech production, including planning difficulty and lexical access (reference). This choice is also motivated by the quantitative results, which identified impediments as the only disfluency type systematically associated with translanguaging presence (Model 2), and as differentially associated with specific translanguaging strategies (Model 3).

Building on the quantitative findings, the analysis concentrates on codeswitching, codemixing and translation separately to capture differences in their cognitive and interactional profiles.

A key methodological decision concerns the explicit distinction between impediments occurring within the same turn as translanguaging, and impediments occurring in the preceding turn by the same speaker. This distinction was introduced for three reasons.

First, it is aimed at separating between temporality and function. Impediments co-occurring with translanguaging may reflect processing difficulty during speech production, whereas impediments in preceding turns may function as early signals of emerging difficulty, which may be resolved by speakers through translanguaging practice.

Second, this distinction helps one to avoid simplistic causal interpretations of the relationship between disfluency and translanguaging. By taking the sequential position of impediments into account, the analysis does not assume that translanguaging is either a direct cause or a direct consequence of disfluency. Instead, it acknowledges that different causal pathways may coexist.

Third, foregrounding sequential position aligns the qualitative analysis with an interactional and process-oriented perspective, in which meaning, planning, and linguistic choice unfold over turns rather than being reducible to isolated units. This approach makes it possible to

interpret translanguaging not only as a coping mechanism, but also as a delayed strategic resource or as an interactionally motivated practice independent of immediate cognitive strain.

Based on this sequential distinction, the qualitative analysis was initially organised around three main groups of samples, each corresponding to one of the identified translanguaging phenomena. Within the macro-groups, two sub-configurations were examined, namely impediments occurring in the same turn as translanguaging, and impediments occurring in the preceding turn by the same speaker. However, as can be seen in Table 27, this sample configuration strategy did not fully adhere to data characteristics.

A small sample was therefore selected, resulting in a total of 6 qualitative excerpts, each analysed in detail with respect to the nature and placement of impediments, the form and function of the translanguaging observable phenomena, the interactional context in which the phenomenon occurs.

Cases were selected through an R script which maximised variation in speakers, contexts, and linguistic repertoires[56]. This strategy ensures analytical contrast while maintaining sensitivity to speaker-specific and interactional variability. Samples are shown in Table 26.

Table 26. Samples from NomadLingo1.1 for qualitative analysis of translanguaging and speech impediments.

Sample_ID	***Analytical group***	***Session ID***	***Speaker ID***	***Transcript***	***Context***
S1	Codeswitching – impediments in the same turn	F1	AR_C_01	ok also du hast **ehm ehm** Deutsch gelernt?	*IT:* German English Italian Neapolitan and- *AR*: sprichst du deutsch? *IT*: ja klar ((chuckles)) *AR*: ah ja? […] *IT*: aber ich habe in Berlin gewohnt und ich arbeite mit Deutsche leute *AR*: **ok also du hast ehm ehm Deutsch gelernt?**
S2	Codeswitching – impediments in the preceding turn	M2	IT_M_02	secondo me un pochino un pochino cioè non tanto	*IT1*: yeah but I can work like**:** yeah like I can**:** yes I nee- I need to be in a**:** country where **ehm** I don't know how to say it the jetlag yeah the time zone makes sense *IT2:* manca un po' di sale vero? A questa lasagna? eh lo sapevo vedi *IT3:* secondo me zero *IT1:* secondo me un pochino un pochino cioè non tanto

[56] The script can be found in the GitHub project repository (https://github.com/novella-tedesco/FLO/tree/Translanguaging-effects-on-speech-(dis)fluency_Tedesco2026).

					IT2: you need more salt in this I always do that because I am I eat with very less salt […]
S3	Codemixing – impediments in the same turn	M7_2	SE_C_01	ah la gomera si ich habe **ehm** I have**: ehm** I stayed there for three days and **ehm** it's been so long that I don't **ehm** I don't actually remember how how it was **ehm**	*IT*: so you were saying? ah ecco fatto *PT*: so we are talking about **ehm** la gomera *GE*: ah la gomera si ich habe **ehm** I have**: ehm** I stayed there for three days and **ehm** it's been so long that I don't **ehm** I don't actually remember how how it was **ehm**
S4	Codemixing – impediments in the preceding turn	F2	FR_C_01	taxes in France **is is** worse like any another country for the taxes **ehm ehm** you pay if you are just **ehm ehm** como se dice esto empleados?	*FR*: yeah **I think** yeah **I think** so and for me is not **ehm** I didn't decide to travel to run away from the tax I just **I I I** don't really care about it because I grew up in France and I had free education I had free wealth **ehm** salud wealth system so**:** I'm I'm proud to have that and to and so I will not **ehm-** *ES*: in France they're way worse than Spain […] *FR*: taxes in France **is is** worse like any another country for the taxes **ehm ehm** you pay if you are just **ehm ehm** como se dice esto empleados?
S5	Translation – impediments in the same turn	GC1_1	AR_C_02	**are the class are are the class are** difficult? las clases son dificiles?	*AR*: ok ok so that's a good percent so you should speak maybe a little bit so we can speak in Spanish *SE*: ((laughs)) little little […] *AR*: **are the class are are the class are** difficult? las clases son dificiles? *SE*: yes and no *AR*: ok wow so *SE*: it's hard for me but not for others

In Sample S1, impediments co-occur with codeswitching within the same turn. One possible interpretation is that the disfluency pattern reflects a speaker-specific trait, rather than being exclusively triggered by translanguaging. However, it is also plausible that the impediments observed are related to the speaker's competence in German and/or to the abrupt shift into a language that is not routinely used in the ongoing interactional context. The hesitation phenomena precede and surround the German segment, suggesting increased planning effort. From an interactional perspective, the codeswitch does not appear to be strategically prepared.

In Sample S2, impediments occur in a preceding turn and are followed by codeswitching. However, while it is true that the Italian speaker does not produce notable disfluencies when speaking Italian and shows more hesitation when speaking English, the codeswitching is initiated by another Italian speaker. The interaction takes place among Italian speakers, and the high fluency in Italian is therefore unsurprising. What is noteworthy is that the interaction becomes more fluent once switching to Italian is available as a resource. This is observable not only in the focal speaker but also in IT2, who, upon returning to English, produces reformulation phenomena such as 'I am I eat'. This suggests that codeswitching here functions less as a response to individual cognitive difficulty and more as a shared interactional resource that facilitates fluency by allowing speakers to operate within their most accessible linguistic repertoire. Such cases highlight the importance of considering participation structure and shared language background when interpreting both disfluency and translanguaging patterns.

In Sample S3, codemixing co-occurs with impediments within the same turn. The speaker's repertoire (German, English, Spanish, Russian, Portuguese) is particularly relevant in this case, as the turn displays words associated with English, Portuguese, and German. It is plausible that the observed disfluencies are partly attributable to increased cognitive load resulting from the dense activation of signs that may occupy distant points within the web of resources available to the speaker. Additionally, the interactional context is characterised by frequent language changes in preceding turns by other interlocutors, which may further increase processing demands. Interestingly, the impediments cluster around the English segments, suggesting that English, despite functioning as a lingua franca, may be the locus of heightened monitoring and planning effort.

In Sample S4, impediments are present both in the preceding turn and in the turn containing translanguaging. In this case, disfluency appears to be a relatively stable characteristic of the speaker's production, rather than an isolated response to a single interactional event. What remains analytically salient is the way in which hesitation phenomena such as 'ehm ehm' precede sourcing resources other than English. This sequential positioning may indicate that translanguaging functions as a solution to locally experienced difficulty, potentially linked to limited competence in English. This interpretation is consistent with interactional approaches that view translingual practices as resources for managing trouble in talk (e.g., Mondada 2018).

In Sample S5, translation co-occurs with impediments within the same turn. The translanguaging practice is explicitly interactional: the Spanish-speaking participant translates in order to support a Swedish interlocutor's language practice. In this case, disfluency may be

caused by multiple interacting factors, including concern for the interlocutor's comprehension, increased monitoring of one's own speech, and the cognitive demands of managing alignment across languages. Translation here does not appear to be cognitively demanding; rather, the effort seems to arise from the need to coordinate linguistic choices with the perceived abilities and expectations of the interlocutor. This aligns with work suggesting that translation in conversational settings is less about linguistic difficulty per se and more about interactional alignment and audience design (Gumperz 1982; Clark 1996).

With respect to translation, this has been noted to combine with impediments almost exclusively in cases where the translation tag occurs within what could also be presented as a codemixing phenomenon, i.e., self-translation within the same turn. In this sense, S5 simultaneously represents both the targeted configuration and its limits. This case suggests that translation can function as a coping mechanism for managing elevated cognitive and interactional load, particularly when speakers anticipate potential misunderstanding or attempt to monitor their speech closely. This observation supports the view that cognitive load is closely tied to alignment and monitoring processes rather than to competence alone (Levelt 1989; Pickering & Garrod 2013).

No cases were retrieved for the configuration *translation preceded by impediments in a prior turn* (Sample S6). While this absence may partly reflect distributional properties of the dataset, it is also analytically meaningful. One plausible interpretation is that translation, unlike codeswitching and codemixing, is less frequently used as a delayed response to prior planning difficulty. From an interactional perspective, this pattern suggests that translation is more closely associated with anticipatory audience design than with the resolution of already manifest production trouble. Speakers appear to translate proactively in order to secure mutual understanding. This interpretation is consistent with the quantitative results of Model 3, where translation showed only a weak and non-significant association with impediments. Taken together, these results support the view that translation in this dataset is not primarily driven by cognitive breakdown, but by interactional considerations related to recipient design, monitoring (Clark & Brennan 1991), and the management of participation frameworks (Clark 1996; Gumperz 1982).

Across all samples, one pattern stands out: impediments are systematically located in English segments and tend to occur immediately before or around translanguaging, rather than within the non-English material itself. With the partial exception of Sample S1 – selected precisely because it contains same-turn impediments – the non-English utterances are comparatively fluent. This distribution suggests that English, as the dominant interactional medium, is the primary site of cognitive and interactional pressure. Translanguaging, in turn, appears to emerge as a locally deployed resource that allows speakers to navigate this pressure, either by redistributing cognitive load or by facilitating alignment with interlocutors.

Summing up, the present follow-up analysis supports the view that the relationship between translanguaging and disfluency is highly sensitive to the local interactional context in which resources are mobilised. Impediments appear to index moments of planning pressure that may

either co-occur with or precede translanguaging, depending on the strategy involved and the interactional contingencies at play.

At the same time, given the nature of this case study, primarily meant to present possible research uses of NomadLingo, the qualitative sample is limited in size and selectively constructed, and the absence or scarcity of certain configurations should be interpreted with caution. These patterns call for further investigation using larger qualitative samples and sequentially richer analytical designs.

6.5.9 Case study 4: Limitations

A main limitation of this study concerns the operationalisation of disfluency, which relies largely on automatic procedures and would benefit from further validation, for instance through random qualitative checks of annotation accuracy. Moreover, the absence of prosodic information in the dataset prevents a finer-grained analysis of temporal and phonetic aspects of fluency. The qualitative analysis is based on a small and deliberately constrained sample; while this allows for analytical precision, future work would benefit from a larger and more flexible sampling strategy. Future studies should further explore the sequential relationship between disfluency and translanguaging, for instance by integrating fine-grained conversation-analytic methods, prosodic annotation, and longitudinal speaker-based modelling. Such follow-up work could clarify whether disfluencies function primarily as triggers, accompaniments, or consequences of translanguaging practices across different mobility trajectories.

6.5.10 Case study 4: Implications for views of plurilingual competence in mobility contexts

The findings of this case study contribute to ongoing debates on plurilingual competence by challenging deficit-oriented interpretations of disfluency in transcultural communication. Rather than signalling communicative failure, disfluencies emerge as part of the adaptive management of complex linguistic repertoires in transcultural contexts. In line with translanguaging perspectives, plurilingual speakers are shown to flexibly mobilise their resources to maintain interactional flow, negotiate meaning, and anticipate interlocutors' needs (García & Wei 2014; Canagarajah 2013). These results support usage-based and dynamic views of competence, in which variability, hesitation, and reformulation are integral to skilled language use rather than deviations from an idealised monolingual norm. In this sense, the study reinforces the need to conceptualise fluency and competence as situated, emergent, and interactionally grounded phenomena (Blommaert 2010; Pennycook 2017).

Chapter 7

Final remarks

This final chapter brings together the main outcomes of the doctoral research here presented and discusses their theoretical, methodological, and practical implications. The aim of the chapter is threefold: first, to provide an integrated summary of the project and its main deliverables, with particular attention to the development of the NomadLingo corpus and the analytical framework built around it; second, to critically assess project limitations, acknowledging constraints related to research design, theoretical positioning, and technical implementation; and third, to outline foreseeable developments and future research directions arising from the findings presented in the thesis.

Section 7.1 offers an overview of the project's objectives and main results. It begins with a discussion of the project deliverables (Section 7.1.1), with particular attention to the theoretical framework and the development of the NomadLingo corpus. Section 7.2 is devoted to a critical reflection on project limitations, distinguishing between issues related to research design, theoretical scope, and technical aspects of data collection and analysis. Section 7.3 then outlines future developments, applications, and research directions, highlighting how the present work can be extended both empirically and theoretically. Finally, Section 7.4 concludes by situating the thesis within current debates on linguistic fluidity, mobility, and contemporary sociolinguistics.

7.1 Project summary, main outcomes and results

This section provides a synthetic overview of the project, bringing together its conceptual foundations, methodological architecture, empirical and theoretical contributions. It foregrounds the internal coherence of the thesis, the ways in which theory and data have been brought into dialogue throughout the research process, the project's strong empirical vocation, its methodological innovations, and the tangible research assets developed within it.

The thesis as a whole set out to explore linguistic fluidity as a characteristic of language produced in transcultural interaction. To this end, it combined the construction of a spoken corpus (NomadLingo), a theoretically informed annotation framework, and a mixed-methods analytical design integrating quantitative modelling and qualitative interactional analysis. Particular emphasis was placed on the development of a FAIR-compliant research infrastructure, ensuring that data collection, annotation, documentation, and dissemination adhere to principles of findability, accessibility, interoperability, and reusability. Within the field of sociolinguistics and mobility studies, this represents an innovative and forward-looking

approach, especially in relation to the systematic documentation of spontaneous, translingual interaction in underexplored mobile contexts, like the one on which the thesis focuses.

The decision to investigate digital nomadism proved to be a valid one for the exploration of linguistic fluidity. Digital nomad communities provide access to interactional environments characterised by high mobility, linguistic diversity, and routine engagement with transcultural communication. The empirical material analysed in this thesis was collected through prolonged engagement within such contexts, made possible by the researcher's sustained presence and participation. As outlined in the Introduction (Section 1.2), this positioning informed the research design and data collection process, allowing for the observation of naturally occurring interaction over time and supporting an approach to linguistic analysis that is situated, relational, and sensitive to lived communicative practices. In doing so, communication and language analysis have proven to be an access gate to sociological findings: an in-depth understanding of identity and community practices of digital nomads can thus be seen as a collateral valuable result of the present thesis.

Sub-section 7.1.1 focuses on the main project deliverables, summarising the theoretical framework developed in Chapter 3 and the methodological design and corpus construction discussed in Chapters 4 and 5, which together constitute both the analytical backbone of the present investigation and a durable research output in their own right. These deliverables extend beyond the scope of the thesis itself, offering a reusable empirical resource and a methodological protocol for future research. Sub-section 7.1.2 then synthesises and expands upon the main empirical findings presented in Chapter 6, discussing their implications for current debates on translanguaging and communicative competence in contexts of mobility. Sub-section 7.1.3 introduces and develops the key theoretical advancements proposed by the thesis, highlighting how an empirically driven, grounded approach has led to the integration of diverse theoretical perspectives, refined and recalibrated on the basis of data. Finally, Sub-section 7.1.4 discusses the social impact of the research from two complementary perspectives**:** on the one hand, the broader societal and educational implications of the findings, particularly in relation to language teaching, inclusion, and the enhancement of linguistic diversity; on the other, the concrete impact-related activities carried out during the doctoral project, including open data practices, training initiatives, and contributions aligned with the universities' third mission and knowledge transfer strategies.

7.1.1 Project deliverables

One of the central contributions of this thesis lies in the development of an integrated theoretical–methodological framework for investigating linguistic fluidity in contexts of mobility. This framework is articulated through two tightly interconnected deliverables: the methodological model outlined in Chapter 4 and the design, construction, and annotation of the NomadLingo corpus presented in Chapter 5.

Chapter 3 established a coherent theoretical foundation by bringing together insights from sociolinguistics, applied linguistics, mobility studies, and transcultural communication.

Chapter 4 then translated this conceptual framework into a concrete research infrastructure through a mixed-methods protocol that integrates quantitative data with qualitative insights. This protocol was designed to allow different analytical scales to interact. In this sense, the methodological design does not treat quantitative and qualitative approaches as separate or sequential stages, but as mutually informing components within an integrated research process, where data collection procedures allow for the compilation of reusable resources and foreground a variety of analytical perspectives across disciplines.

The corpus constitutes a second major deliverable of the project. NomadLingo was designed to capture naturally occurring spoken interaction among highly mobile speakers – digital nomads living across some of the most popular nomadic hubs and communities in Europe – with a particular focus on informal, spontaneous communicative settings. It combines rich contextual metadata (e.g. speaker nationality, linguistic repertoires, interactional settings) with fine-grained linguistic and interactional annotation.

A related key deliverable is the corpus multi-layer annotation scheme, which includes, among others, turn-level variables, translanguaging phenomena, self-repair strategies, and misunderstanding sequences. This layered structure enables both qualitative interactional analysis and quantitative modelling, allowing different analytical perspectives to be integrated within the same empirical object. The structure and encoding of the annotation scheme were designed to be extensible, allowing for the addition of further annotation layers as new research questions emerge. At the same time, the corpus formats support use both with and without annotation, facilitating exploratory analyses as well as reuse by researchers with different theoretical orientations. This design choice reflects an awareness that annotation is itself an interpretive and theory-laden practice, rather than a neutral representation of linguistic reality (McEnery & Hardie 2011), and that analytical transparency and flexibility are therefore essential.

Chapters 4 and 5 deliver not only sets of data ready for analysis but a reusable methodological framework that can be adapted and extended to other transcultural and mobile contexts. The combination of theoretical grounding, methodological pluralism, and empirically driven corpus design constitutes a central asset of the project.

Summing up, beyond the methodological protocol itself, the set of resources produced through the project includes: i) the NomadLingo spoken corpus[57]; (ii) a documented multi-layer annotation scheme, which can be selectively adopted or expanded for related corpora focusing on mobility, multilingualism, or interactional phenomena (Appendix 9); (iii) a structured set of metadata about digital nomads taking part in the FLO projects (Appendices 7 and 8); (iv) two datasets containing responses to two versions of a survey about language attitudes, self-perceived practices and needs within digital nomads communities[58]. Moreover, during the

[57] http://hdl.handle.net/20.500.11752/OPEN-2137.

[58] DOIs: 10.6092/unibo/amsacta/7722; 10.5281/zenodo.15603903.

project ethnographic insights were translated into academic and non-academic articles, blog and instagram posts (as documented in Chapter 4), which enhanced project visibility, active discussion and participation from a diverse audience.

These resources were designed to support future research in sociolinguistics, applied linguistics, discourse studies, and mobility research, as well as comparative and longitudinal investigations of linguistic practices across mobile populations.

7.1.2 Main findings

The main empirical findings of the thesis, presented in Chapters 4 and 6, provide robust evidence for reconceptualising linguistic fluidity as an interactional achievement shaped by practice, rather than as a stable speaker attribute or a deficit relative to monolingual norms. By combining quantitative modelling with qualitative analysis, the case studies in Chapter 6, primarily aimed at exemplifying some of the uses of the resources produced within the FLO project, explore how multiple dimensions of interaction jointly shape fluid communication as represented in the NomadLingo corpus.

Rather than isolating individual variables, the analyses in Sections 6.2, 6.3, and 6.4 collectively illustrate how translanguaging, knowledge negotiation and co-costruction, identity representation and community building are intertwined within spontaneous interaction. These sections show that linguistic resources are mobilised dynamically in response to local interactional demands, participants' awareness of each other's repertoires, and the broader communicative ecology in which interaction unfolds. In this sense, the case studies do not aim to establish deterministic causal relations, but to exemplify patterned tendencies that emerge across different interactional configurations. Phenomena such as humour and irony, for instance, are shown to emerge as interactional outcomes of meaning negotiation and epistemic recalibration, rather than as isolated pragmatic strategies.

One of the central findings presented in Chapter 6 concerns the relationship between translanguaging and disfluency. Contrary to deficit-oriented assumptions that associate bilingualism and translanguaging with increased processing difficulty, the quantitative analyses show that translanguaging is not systematically associated with higher levels of disfluency once interactionally relevant variables are taken into account. This finding supports a view of translanguaging as a functional communicative resource rather than a source of breakdown, aligning with recent practice-based plural approaches to languaging, communicative competence and language learning.

Broadly speaking, the initial findings presented in Chapter 6 suggest that linguistic fluidity emerges from the interplay between cognitive, interactional, and sociobiographical dimensions. It is shaped by speakers' trajectories of mobility, their awareness of co-participants' repertoires, and the interactional ecology in which communication unfolds. The analytical focus on specific phenomena - either pragmatically or linguistically oriented, like misunderstanding or disfluency respectively - should therefore not be interpreted as exhaustive,

but as strategically selective. Other aspects of linguistic fluidity could have been foregrounded with equal analytical legitimacy, depending on the research lens adopted.

The choice to concentrate on these phenomena reflects the exploratory and empirically driven nature of the study, which treats the case studies as illustrative entry points into a broader, multi-dimensional understanding of fluid communication. By empirically demonstrating these relationships, the study advances current debates on ELF, translanguaging, and language socialisation, and provides a data-driven foundation for rethinking fluency beyond monolingual and native-speaker-oriented frameworks.

7.1.3 Empirically-grounded theoretical advancements

One of the central contributions of this thesis lies in the development of an integrated theoretical framework for linguistic fluidity that brings into dialogue translanguaging, intercomprehension, and related repertoire-based approaches, while recognising their different theoretical statuses and analytical scopes. Rather than proposing a unified or homogeneous theory, the study adopts translanguaging as an overarching analytical perspective, within which intercomprehension is conceptualised as a complementary and partially overlapping dimension of fluid meaning-making in transcultural interaction. While these constructs have often been treated separately in the literature, the present study shows that, in naturally occurring interaction among mobile speakers, they frequently co-occur at the level of practice and mutually reinforce one another as part of speakers' flexible deployment of semiotic repertoires.

Building on the theoretical discussion in Chapter 3, translanguaging is applied as a *practical theory of language* (Wei 2018) that foregrounds speakers' situated mobilisation of linguistic, semiotic, embodied, and interactional resources. The thesis argues that phenomena such as intercomprehension, informal translation, language-related metalinguistic reflection, ELF-oriented accommodation practices, and technology-augmented communication are best understood as manifestations of a broader capacity for linguistic fluidity.

These theoretical propositions are not formulated a priori but are progressively refined through a grounded engagement with empirical data, as documented in Chapters 4, 5, and 6. The empirically driven workflow of the study—combining corpus construction, annotation, quantitative modelling, and qualitative interactional analysis—allowed existing theories to be both corroborated and problematised. This process led to a more fine-grained understanding of how transcultural communicative practices unfold in real interactional settings, taking into account the interplay between macro-level sociolinguistic conditions (e.g. mobility regimes, language ideologies) and micro-level interactional dynamics (e.g. accomodation and alignment, self-repair).

The findings presented in Chapter 6, in particular, challenge deficit-oriented interpretations of non-standard varieties of English and other fluid communicative practices. In line with recent work on plurilingualism and transcultural communication, the thesis contributes to a reconceptualisation of transcultural communicative competence as an interactional and

emergent achievement. This reconceptualisation shifts analytical attention away from formal accuracy and adherence to standard norms, and towards situated performance, interactional effectiveness, and the co-construction of transcultural spaces of interaction.

By integrating insights from sociolinguistics, applied linguistics, mobility studies, and translation theory, the thesis advances current debates on linguistic repertoires, communicative competence, and language use in conditions of mobility. More broadly, it exemplifies a research stance in which theory is treated as provisional and revisable, continuously reshaped through systematic engagement with empirical data. This stance underpins the proposed theory of linguistic fluidity and positions the thesis within a growing body of research that conceptualises language as an adaptive, emergent, and socially embedded practice.

7.1.4 Social impact and educational initiatives

The methodological innovations, theoretical advancements and empirical findings presented throughout the thesis carry relevant social and educational implications, particularly with respect to language teaching and learning. In particular, the results discussed in Section 6.5 suggest that adherence to monolingual norms may increase cognitive load and and self-monitoring pressure in spontaneous interaction. This observation aligns with research on language anxiety and bilingual education, which has shown that norm-oriented pedagogical models may negatively affect learners' confidence and communicative engagement (Horwitz 2010; Cummins 2007). Conversely, the observed association between misunderstanding management, translanguaging practices, and successful interaction points to the potential benefits of plural and flexible pedagogical approaches, in which speakers are encouraged to mobilise their full linguistic repertoires as adaptive communicative resources (García & Wei 2014).

From this perspective, the results support educational models that foreground plurilingual competence, translanguaging, and the strategic mobilisation of heterogeneous linguistic and semiotic resources. Rather than framing non-standard forms as deficiencies, such approaches recognise them as adaptive responses to complex communicative ecologies, where meaning-making depends on negotiation, alignment, and creativity. This aligns with European policy frameworks such as CARAP (Candelier et al. 2012) and with recent work in translanguaging pedagogy, which emphasises the pedagogical value of learners' full repertoires and the legitimacy of hybrid practices (García et al. 2017).

Moreover, by highlighting the interactional nature of disfluencies, the research findings could be used to enhance pedagogical approaches in which hesitation, reformulation, and translinguistic negotiation are not penalised but integrated into learning trajectories. In this sense, the study contributes to ongoing debates on creativity in language pedagogy and learner-centred language education (van Lier 2004; García & Wei 2014). Furthermore, the translanguaging practices observed in the corpus – especially informal translation and metalinguistic reflection – provide empirical evidence that pedagogical practices fostering

metatranslinguistic awareness can enhance communicative competence (Cook 2010; Cenoz & Gorter 2015).

These pedagogical principles were translated into concrete educational initiatives during the doctoral project. In particular, the findings informed a series of public-facing and educational activities, including workshops such as those organised within the 2024 European Youth Event in Forlì or activities like the Plurilingual Treasure Hunt (Paone et al. 2025) were used to promote plurilingual awareness, language creativity, and inclusive communication practices. These initiatives demonstrate how empirical research and translanguaging-oriented insights can be operationalised in applied educational contexts, fostering dialogue between academic research and broader public. Moreover, seminar cycles on data fairness and working procedure for the management and processing of complex datasets including personal data and naturally occurring spoken interactions were organised both at the University of Bologna and at hosting universities (i.e., University of Madeira, and KU Leuven).

Beyond formal educational settings, social impact was also pursued through active engagement with the communities under study. Adopting a participatory and community-based ethnographic approach (as extensively discussed in Chapter 1 and 4), the research involved ongoing dialogue with participants, awareness-raising around language practices, and reciprocal knowledge sharing. This engagement extended beyond data collection, as evidenced by sustained informal exchanges with community members concerning language use, dialects, and communicative norms, reinforcing the project's ethical and socially embedded orientation.

Finally, the attention to social impact and public engagement developed in this thesis aligns with the strategic priorities of contemporary academic research, where knowledge transfer, open science, and societal engagement are explicitly valued (European Commission 2016b; ANVUR 2023). In this sense, the doctoral project not only contributes to scientific knowledge but also seeks to enact a form of socially responsible research that places particular emphasis on educational transformation, inclusivity, and public engagement.

7.2 Limitations

Beyond its positive outcomes, the present research is also characterised by a number of limitations that need to be acknowledged. These limitations delineate the scope within which the findings should be interpreted and point towards directions for future refinement.

7.2.1 Research design

The research design adopted in this thesis was intentionally broad and integrative, combining corpus construction, quantitative modelling, qualitative interactional analysis, and ethnographic sensitivity. While this mixed-methods transdisciplinary approach constitutes one of the strengths of the project, it also entails a set of structural limitations that must be explicitly acknowledged. Integrating multiple methods, theoretical traditions, and analytical scales inevitably requires trade-offs in terms of depth, time investment, and disciplinary

specialisation. Given the scope and duration of a doctoral project, it was not possible to pursue all analytical dimensions with the same level of granularity, nor to fully exhaust the potential of each methodological component.

More specific limitations concern corpus design. While the thoroughly documented corpus construction procedure, the annotation scheme, and the granular annotation system allow for further improvements with regards to current limitations (as will be discussed in Section 7.3), some technical and structural constraints remain. In particular, while the corpus seeks to capture language in interaction, its bidimensionality is a limitation, especially considering the advancements made in multimodal corpora in recent times (Knight & Adolphs 2021). Moreover, although the corpus includes extensive contextual metadata, information on participants' sociobiographical trajectories and education could not be systematised to the same degree across all speakers and were therefore excluded from the metadata, despite these variables being recognised as important factors shaping speakers' linguistic attitudes and behaviours in the literature (as discussed in Chapter 4). These limitations do not undermine the validity of the analyses conducted, but they do delimit the range of questions that can be addressed by using the resource, while pointing to directions for further improvements of the current version of the corpus and more careful project design for future translingual spoken corpora. In this sense the construction of NomadLingo can be regarded not as a definitive solution but rather as an exploratory and generative step indicating concrete directions for the further development of the corpus and for more reflexive design choices for future corpora, treating translanguaging as a default condition of interaction rather than a marked deviation (Tedesco et al. 2025).

More broadly, as has been argued in corpus-informed discourse studies, corpora necessarily render visible only what is produced and recordable, while leaving other dimensions of meaning-making - such as gaze, uncertainty, implicit misalignment, or unvoiced misunderstanding - methodologically opaque (Taylor 2013). This limitation is particularly relevant in the analysis of transcultural interaction, where participants may choose not to signal misunderstanding explicitly in order to preserve face, maintain interactional flow, or avoid social friction.

A related methodological issue concerns the interpretation of successful interaction. The frequent absence of overt signals of communicative breakdown in the corpus can be read as an index of high interactional competence and tolerance for partial understanding. At the same time, this absence raises methodological questions about how non-understanding, uncertainty, or latent misalignment can be identified when they are interactionally managed through silence, accommodation, or strategic non-problematisation rather than explicit repair. What appears as smooth interaction in the corpus may therefore, in some cases, conceal underlying interpretative effort or unresolved asymmetries. In this respect, the ethnographic dimension – and my participation in the interactional contexts under study – plays a crucial compensatory role. This confirms that, in highly fluid and trust-based communicative environments, insider positionality can function less as a source of bias than as an epistemic resource, allowing the researcher to triangulate corpus evidence with situated experiential knowledge.

Further limitations emerge when considering the individual case studies presented in Chapter 6. Each case study was designed to exemplify specific analytical affordances of the FLO resources and in particular the NomadLingo corpus, rather than to provide exhaustive coverage of the phenomena under investigation. As such, the scope of each analysis is necessarily selective. In the final Case study focusing on disfluency (Section 6.5), for instance, the operationalisation of disfluency remains necessarily reductive with respect to the full range of temporal, prosodic, and cognitive dimensions involved in speech production. A more fine-grained operationalisation would require either aligned acoustic data or complementary experimental measures, which fall beyond the scope of the present design. Moreover, most elements included in the general disfluency parameter, such as repetitions, were calculated automatically and would benefit from manual revision.

Furthermore, the qualitative excerpts of the case studies rely on a relatively small sample size. While this is consistent with interactional and ethnographic traditions prioritising depth over breadth, it limits the extent to which patterns of stance-taking, alignment, and identity positioning can be generalised. A more extensive qualitative analysis of stance and evaluative positioning across a wider range of speakers and interactional contexts would be necessary to fully capture the variability of these practices within the communities under study.

Finally, in the context of Nomad English and ELF-oriented analyses (Section 6.3), the study prioritised interactionally grounded explanations over large-scale quantitative modelling of linguistic variation. As a result, variationist analyses remain relatively underdeveloped compared to a research design explicitly oriented towards frequency-based comparison across groups.

7.2.2 Theoretical issues

From a theoretical perspective, a key limitation concerns the necessary selectivity of the analytical framework. While the study engages extensively with translanguaging, ELF, and linguistic fluidity, other theoretically relevant perspectives could not be addressed in depth. In particular, critical discourse analysis and the systematic study of ideological positioning were not developed as standalone analytical strands, despite their clear relevance for understanding linguistic and cultural dynamics in digital nomad communities.

Similarly, although the corpus contains a range of lexical innovations and emergent expressions related to digital nomadism, mobility, and platform-based labour, the analysis of neologisms and lexical creativity was not pursued in detail. This decision reflects the linguistic and interactional focus of the thesis, but leaves open important paths for future research at the intersection of sociolinguistics, discourse studies, and lexical innovation.

More broadly, the study adopts a translanguaging perspective that foregrounds practices and repertoires rather than named languages. While this choice is theoretically motivated and consistent with recent sociolinguistic work, it should be acknowledged that, for analytical and operational purposes, the study occasionally draws on more established concepts such as *code-*

switching and *code-mixing*. These terms are used in a pragmatic sense, as heuristic labels that facilitate systematic annotation and comparison, rather than as endorsements of a compartmentalised or monolingual view of linguistic competence. Their use does not imply the assumption of discrete language systems, but reflects the need to engage with existing analytical tools while explicitly reframing them within a repertoire-based and practice-oriented perspective.

7.2.3 Technical issues

Finally, some limitations are linked to technical and methodological constraints. Although the corpus is carefully annotated and designed for reuse, annotation inevitably involves analytical choices that privilege certain surface features over others. In particular, while translanguaging observable phenomena, repairs, and misunderstanding are systematically annotated, other dimensions – e.g. prosody, or interactional patterns – are only indirectly captured or not included at all. This limitation is especially relevant given the increasing importance of multimodality and technology-mediated interaction in the communicative practices of digital nomads. The current version of NomadLingo is primarily audio-based and text-oriented, and future expansions would benefit from the integration of multimodal data, including audio materials as well as other text types like chat interactions.

Moreover, certain constraints emerged during corpus querying and analytical implementation that are not inherent to the corpus itself, but rather reflect broader methodological limitations of current concordance software and corpus-linguistic infrastructures. These tools remain largely grounded in monolingual, fixed representations of language, which are poorly equipped to model fluid, repertoire-based language practices characteristic of translanguaging interaction (Tedesco et al. 2025). This misalignment highlights the need for further methodological innovation in corpus tools if translingual spoken data are to be fully exploited.

In addition to analytical and infrastructural constraints, some limitations are linked to the data collection procedures adopted in the project. Data were gathered in natural settings which inevitably limited the degree of control over recording conditions. While this openness is methodologically consistent with the ethnographic and ecological orientation of the study, it also resulted in uneven recording quality. Moreover, the reliance on audio-only recordings constrained the systematic capture of embodied, multimodal resources that are known to play a central role in meaning-making in face-to-face interaction. The latter limitation was partially addressed through contextual notes taken during data collection; however since data were transcribed later and revisors were not present during data collection, the limitation here acknowledged remains. This reflects the trade-offs inherent in prioritising ecological validity and participant-led interaction, and points to the need for more systematically multimodal and longitudinal data collection strategies in future extensions of the project.

7.3 Foreseeable Developments, Applications and Future Research Directions

The project opens up a wide range of future developments and applications, both within linguistics and across adjacent disciplinary fields. A first area of expansion concerns the corpus itself. NomadLingo is explicitly conceived as an extensible resource, and future iterations could include larger and more balanced samples, longitudinal components tracking speakers' trajectories over time, and the systematic integration of multimodal data. In particular, future developments could incorporate methods of linguistic profiling and speaker modelling, drawing on approaches from corpus linguistics and computational sociolinguistics (Biber 1993b; Brezina 2018), to systematically capture variation across speakers, repertoires, and interactional contexts. Such approaches would make it possible to move beyond descriptive analysis towards more fine-grained modelling of linguistic behaviour across mobile populations. Extending the corpus to include chat-based communication, platform-mediated interaction, and audiovisual material would make it possible to investigate how linguistic fluidity is negotiated across different semiotic channels, allowing for more fine-grained analyses of meaning-making in technology-mediated environments.

While these developments primarily concern corpus design and linguistic analysis, the project also opens up broader interdisciplinary perspectives. Beyond linguistics, the resources developed within the FLO project hold substantial potential for interdisciplinary research. From a migration studies perspective, the corpus and associated survey data could be used to explore processes of integration and belonging, the role of social networks, perceptions of autonomy and precarity, gendered experiences of mobility, and attitudes towards safety, freedom, and stability. Comparative analyses between digital nomads and more sedentary populations could further illuminate how sustained mobility may reshape social relations, identity construction, and everyday linguistic practices, offering valuable insights for both sociological theory and applied policy research.

The corpus also lends itself to sociological and (critical) discourse-analytic approaches. Future research could examine how power relations, inequalities, and ideological positions are negotiated in everyday interaction, as well as how dominant narratives around freedom, flexibility, and entrepreneurship are reproduced or contested in digital nomad discourse. In this sense, NomadLingo provides a valuable empirical basis for connecting micro-level interactional analysis with macro-level social processes.

From an educational perspective, the findings of this thesis point towards the need for more explicitly plural, interdisciplinary, and critical approaches to language education. Teaching, understood as a political act, plays a crucial role in opening up ways of seeing the world and in recognising learners' full repertoires as legitimate resources. As argued by García and colleagues (2017), language practices are deeply intertwined with identity, agency, and representation, and restricting learners to a single named language risks silencing their experiences, stories, and ways of thinking. In this sense, approaches such as CLIL, literature-based language teaching, and translanguaging pedagogy are particularly relevant, as they foreground meaning, creativity, and critical engagement rather than mere formal accuracy. The

resources produced within this project could be practically used in teacher education, curriculum design, classroom-based research, and educational materials development, for instance through the use of annotated corpus excerpts to raise metalinguistic awareness, design data-driven learning activities, or critically reflect on real-world communicative practices in multilingual and mobile contexts. The emphasis on communication as a dynamic, socially embedded process resonates with broader societal transformations driven by digitalisation, real-time translation technologies, artificial intelligence, and intensified mobility patterns, reinforcing the need for educational models that are responsive to contemporary realities.

From an organisational perspective, future iterations of the FLO project could also move towards less centralised and more distributed research models, involving targeted projects focusing on specific communities, participants or settings. Such designs would allow for greater control over contextual variables while maintaining comparability across datasets, and would support the transition from exploratory to more hypothesis-driven research.

Finally, the project gestures towards forms of research valorisation beyond traditional academic outputs. The corpus, analytical tools, and conceptual frameworks developed here may support future collaborations with educational institutions, policy-makers, cultural organisations, and actors involved in fair and sustainable tourism. In line with evolving models of knowledge transfer, the long-term value of the project lies in its contribution to open, reusable research infrastructures, its capacity to inform inclusive language policies, and its potential to foster more equitable and reflexive understandings of multilingual communication in a globalised world.

More broadly, the thesis highlights the urgent need for corpus linguistics to adapt to rapidly changing real-world language practices. Digitalisation, artificial intelligence, real-time translation tools, and intensified patterns of mobility increasingly challenge traditional assumptions about language stability, speaker identity, and bounded linguistic systems. In such a context, the relationship between speaker, language, culture, and nation can no longer be treated as fixed or self-evident, calling for corpus designs and analytical frameworks that are sensitive to fluidity, hybridity, and emergence.

Relatedly, future research could build on the present study to conduct more in-depth investigations of English variants in superdiverse interactional settings, further exploring how new forms of English are shaped, negotiated, and reconfigured in highly mobile and transcultural communicative ecologies. In this direction, more focused and longitudinal research designs – such as following a small number of speakers across different locations and interactional contexts – could provide deeper insight into the dynamics of repertoire expansion and identity construction over time.

7.4 Conclusion

In conclusion, this thesis has investigated linguistic fluidity as an interactional, cognitive, and socially situated phenomenon, using digital nomad communication as a privileged site of

observation. By combining corpus-informed quantitative analysis with fine-grained qualitative investigation, the study has shown that fluid language practices cannot be adequately captured through static models of multilingualism, nor reduced to deviations from standardised norms. Instead, linguistic fluidity emerges as an adaptive resource through which speakers manage meaning-making, cultural alignment, identity representation and community co-construction through languaging processes.

At the theoretical level, the thesis contributes to ongoing debates in ELF research, translanguaging studies, and language socialisation by foregrounding practices over codes, repertoires over named languages, and trajectories over fixed sociolinguistic categories. The findings support recent perspectives that conceptualise English as a flexible resource embedded within broader semiotic ecologies, rather than as a bounded linguistic system. In doing so, the study aligns with approaches that view language not as a stable object, but as a socially and politically constructed category whose material effects are nevertheless real and consequential. This contribution is particularly relevant for research on transcultural communication, where linguistic practices are inseparable from mobility, identity negotiation, and social positioning.

Methodologically, the thesis demonstrates the analytical potential of FAIR-oriented, systematically processed ethnographic data and, in particular, spoken corpora, for the study of translanguaging and transcultural communication. The mixed-methods design adopted shows how surface-level linguistic features – when consistently operationalised and interpreted – can serve as robust proxies for underlying cognitive and social processes. The FLO research infrastructure and the NomadLingo corpus provide a concrete example of how FAIR principles can be implemented in sociolinguistic research, offering reusable, extensible protocols for future work across linguistics and related disciplines. In this respect, the study contributes to ongoing methodological debates on how corpus linguistics can be adapted to capture fluid, mobile, and ethnographically grounded language practices.

Empirically, the results challenge deficit-oriented interpretations of non-standard language use by showing that linguistic fluidity, variation and misunderstanding are not signs of communicative failure, but integral components of meaning construction in transcultural interaction. Far from hindering communication, translangaging practices often support negotiation, creativity, and social alignment, particularly in contexts characterised by superdiversity and shared experiential knowledge that is actively and locally co-constructed. These findings are especially significant for research on digital nomadism, where communication unfolds at the intersection of mobility, informal sociality, and temporary communities of practice.

More broadly, the thesis argues that understanding language in contemporary societies requires a shift towards plural, transdisciplinary, and socially engaged perspectives. In a world shaped by technological mediation, transnational mobility, and increasing linguistic complexity, research, education, and policy must move beyond monolingual ideologies and embrace the full range of resources speakers bring to interaction. By offering empirically grounded insights into digital nomad communication, the study contributes to digital nomadism research by

foregrounding language as a central dimension of mobility, integration, and community life. At the same time, it advances research on translanguaging and transcultural communication by providing data-driven evidence of how linguistic fluidity is enacted in practice. Finally, through its FAIR-oriented design and integrated corpus-based approach, the thesis contributes to methodological innovation in corpus linguistics, particularly in relation to the study of mobile, superdiverse, and underexplored communicative settings.

References

Abney, S. (2011). Data-intensive experimental linguistics. *Linguistic Issues in Language Technology*, 6(2), 1–27. https://doi.org/10.33011/lilt.v6i.1235

Adolphs, S. (2008). Corpus and context: Investigating pragmatic functions in spoken discourse. *John Benjamins Publishing.*

Adolphs, S., & Knight, D. (2020). Building a spoken corpus: What are the challenges? In P. Baker & J. Egbert (Eds.), *Triangulating methodological approaches in corpus-linguistic research* (pp. 103–121). Routledge.

Ahuja, V., Patel, A., & Ahuja, J. (2020). Digital nomads: A systematic literature review and future research agenda. *Journal of Global Mobility*, 8(3), 213–239.

Aijmer, K. (2002). English discourse particles: Evidence from a corpus. *John Benjamins Publishing.*

Aithal, A., & Aithal, P. (2020). Development and Validation of Survey Questionnaire & Experimental Data—A Systematical Review-Based Statistical Approach. *International Journal of Management, Technology, and Social Sciences (IJMTS),* 5, 233-251. https://ssrn.com/abstract=3724105

ALLEA – All European Academies. (2017). *The European Code of Conduct for Research Integrity.* https://allea.org/code-of-conduct/

Allen, M., & Wiles, J. L. (2016). A rose by any other name: Participants choosing research pseudonyms. *Qualitative Research in Psychology, 13*(2), 149–165.

Alma Mater Studiorum – Università di Bologna. (2023). *AMS Acta: Open-access institutional repository for research outputs and data.* https://amsacta.unibo.it/

an Avermaet, P. (2009). The CEFR and minority languages: Impact and challenges. *Language Policy*, 8(2), 125–147.

Anderson, L. (2006). Analytic Auto-Ethnography. *Journal of Contemporary Ethnography - J CONTEMP ETHNOGR*. 35. 373-395. https://doi.org/10.1177/0891241605280449

Androutsopoulos, J. (2008). Potentials and limitations of discourse-centred online ethnography. *Language@Internet, 5*, Article 8. https://www.languageatinternet.org/articles/2008/1613

Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism, 19*(2), 185–205.

Antonini, R., Cirillo, L., Rossato, L., & Torresi, I. (Eds.). (2017). *Non-professional interpreting and translation: State of the art and future of an emerging field of research* (Benjamins Translation Library, Vol. 129). Amsterdam & Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/btl.129

ANVUR. (2023). *Linee guida per la valorizzazione dell'impatto sociale della ricerca*. Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca.

Anzaldúa, G. (2012). *Borderlands / La Frontera: The new mestiza* (4th ed.). Aunt Lute Books.

Arbutina, H. & Šinković, Z. & Pribisalić, L. (2023). The Legal Status of Digital Nomads. 1422-1427. https\\:doi.org\10.23919/MIPRO57284.2023.10159694.

Arnaut, K., Blommaert, J., Rampton, B., & Spotti, M. (2016). *Language and superdiversity*. Routledge.

Aroles, J., Granter, E., & Vaujany, F.-X. de. (2022). *Becoming a digital nomad: The emergence of new work practices*. **Organisation Studies, 43**(4), 555–573.

Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge. Baayen 2008

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). *Mixed-effects modeling with crossed random effects for subjects and items*. Journal of Memory and Language, 59(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Babbie, E. (2016). *The practice of social research* (14th ed.). Cengage Learning.

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). *WhisperX: Time-accurate speech transcription of long-form audio*. arXiv. https://arxiv.org/abs/2303.00747

Baker, C. (1992). *Attitudes and language*. Multilingual Matters.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Baker, W. (2011). *Intercultural awareness: Modelling an understanding of cultures in intercultural communication*. Language and Intercultural Communication, 11(3), 197–214. https://doi.org/10.1080/14708477.2011.577779

Baker, W. (2015). *Researching English as a lingua franca*. Palgrave Macmillan.

Baker, W. (2022). From intercultural to transcultural communication. *Language and Intercultural Communication*, *22*(3), 280–293. https://doi.org/10.1080/14708477.2021.2001477

Baker, W. (2024). Transcultural communication. In **W. Baker & J. Jenkins** (Eds.), *The Routledge handbook of English as a lingua franca*. Routledge. https://doi.org/10.4324/9781003207245-10

Balazs, C. L., & Morello-Frosch, R. (2013). The three Rs: How community-based participatory research strengthens the rigor, relevance, and reach of science. *Environmental Justice, 6(*1), 9–16. https://doi.org/10.1089/env.2012.0017

Ball, J. (2012). Identity and knowledge in indigenous language socialisation research. *Language and Linguistics Compass, 6 (7),* 447–459. https://doi.org/10.1002/lnc3.347

Bare Bones Software, Inc. (2023). *BBEdit* (Version 14). https://www.barebones.com

Barni, M., & Extra, G. (2008). *Mapping linguistic diversity in multicultural contexts*. De Gruyter Mouton.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68* (3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Başaran, A. (2025). Digital Nomads, the New Frontier of Work in the Digital Age: A Bibliometric Analysis. *Sustainability*, *17*(5), 1906. https://doi.org/10.3390/su17051906

Baude, O., Blanche-Benveniste, C., Calas, M., Cappeau, P., Cordereix, P. et al. (2006). *Corpus oraux, guide des bonnes pratiques*. Presses universitaires d'Orléans; CNRS Éditions.

Bauman, Z. (1998). *Globalisation: The Human Consequences.* Cambridge: Polity Press.

Bauman, Z. (2000). *Liquid modernity*. Polity Press.

Bednorz, J. (2024). Working from anywhere? Work from here! Approaches to attract digital nomads. *Annals of Tourism Research, 105*. https://doi.org/10.1016/j.annals.2023.103715

Bennett, J. (1976). *Linguistic Behaviour.* Cambridge University Press.

Benson, M., & O'Reilly, K. (2009). Migration and the search for a better way of life: A critical exploration of lifestyle migration. *The Sociological Review*, 57(4), 608–625. https://doi.org/10.1111/j.1467-954X.2009.01864.x

Benson, M., & Osbaldiston, N. (2014). *Understanding lifestyle migration: Theoretical approaches to migration and the quest for a better way of life*. London: Palgrave Macmillan.

Bhatia, T. K., & Ritchie, W. C. (2004). Codeswitching: Introduction. In W. C. Ritchie & T. K. Bhatia (Eds.), *The handbook of bilingualism* (pp. 375–390). Malden, MA: Blackwell Publishing.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3–11. https://doi.org/10.1017/S1366728908003477

Biber, D. (1993a). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (1993b). Using register-diversified corpora for general language studies. *Computational Linguistics, 19*(2), 219–241.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817–848. https://doi.org/10.1093/poq/nfq058

Birtchnell, T. (2019). Digital nomads and (im) mobile identities. In *Routledge handbook of identity studies* (pp. 274-282). Routledge.

Blackledge, A., & Creese, A. (2017). Translanguaging and the linguistic repertoire. In S. May (Ed.), *The Routledge handbook of migration and language* (pp. 23–35). Milton Park: Routledge.

Blanchard, M. & Buysse, L. (2021). Like in Discourse Marker Combinations in Spoken Interaction. *Corpus Pragmatics. 5.* https://doi.org/10.1007/s41701-021-00105-4

Blommaert, J. (2010). *The sociolinguistics of globalisation*. Cambridge: Cambridge University Press.

Blommaert, J. (2013). Ethnography, Superdiversity and Linguistic Landscapes: Chronicles of Complexity. Bristol: Multilingual Matters

Blommaert, J. (2014). From mobility to complexity in sociolinguistic theory and method. *Tilburg Papers in Culture Studies*, Paper 100. Tilburg University.

Blommaert, J., & Rampton, B. (2011). Language and superdiversity. *Diversities*, 13(2), 1–21.

Blommaert, J., & Rampton, B. (2011). Language and superdiversity. *Diversities*, 13(2), 1–21.

Blommaert, J., & Varis, P. (2015). Enoughness, accent, and light communities: Essays on contemporary identities. *Tilburg Papers in Culture Studies*, 139.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston. **Boersma, P., & Weenink, D. (2024).** *Praat: Doing phonetics by computer* (Version 6.4). Amsterdam: University of Amsterdam. https://www.praat.org

Bødker, S. (2016). Rethinking technology on the boundaries of life and work. *Personal and Ubiquitous Computing*, *20*(4), 533-544.

Bonneau, C., Aroles, J., & Estagnasié, C. (2023). Romanticisation and Monetisation of the Digital Nomad Lifestyle: The Role Played by Online Narratives in Shaping Professional Identity Work. Organisation, 30(1), 65-88.

Bordieu 1991

Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9604.001.0001

Borin, L. (2011). *Language technology: A survey of the field*. In J. Simpson (Ed.), *The Routledge handbook of applied linguistics* (pp. 77–92). London: Routledge.

Bornstein, M. H., Putnick, D. L., & Lansford, J. E. (2013). Parenting attributions and attitudes in cross-cultural perspective. *Parenting: Science and Practice*, 13(3), 213–247. https://doi.org/10.1080/15295192.2013.832456

Bourdieu, P. (1991). *Language and symbolic power.* Harvard University Press.

Bourdieu, P. (1999). Scattered remarks. *European Journal of Social Theory*, *2*(3), 334-340.

Bowern, C. (2018). Computational phylogenetics. *Annual Review of Linguistics*, *4*(1), 281-296.

Bozzi, A. (2024). Digital nomadism from the perspective of places and mobilities: a literature review. *European Transport Research Review*. 16. https://doi.org/10.1186/s12544-024-00663-z.

Brenner, N. (2001). The limits to scale? Methodological reflections on scalar structuration, *Progress in Human Geography, 15, 4 (2001): 525-548*.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316410899

British Educational Research Association (BERA). (2018). *Ethical guidelines for educational research* (4th ed.). London: BERA.

Broeder, D., & Odijk, J. (2023). *CLARIN taxonomy of tools and services tasks.*

Brooks, G. (2025). Putting Corpora in Context. Plenary talk at Birmingham Conference on Linguistics. Birmingham Aston University.

Brookes, G., & Baker, P. (2021). *Obesity in the News: Language and Representation in the Press.* Cambridge University Press.

Brusseau, R. A. (2021). *Digital nomads: In search of freedom, community, and meaningful work in the new economy*. Chicago: University of Chicago Press.

Bucholtz, M. (2001). The politics of transcription. *Journal of Pragmatics, 32*(10), 1439–1465. https://doi.org/10.1016/S0378-2166(99)00094-6

Bucholtz, M. (2007). Variation in transcribability: On the methodological problems of representing speech. *Journal of Sociolinguistics, 11*(6), 784–808. https://doi.org/10.1111/j.1467-9841.2007.00345.x

Bucholtz, M., & Hall, K. (2005). *Identity and interaction: A sociocultural linguistic approach. Discourse Studies, 7*(4–5), 585–614. https://doi.org/10.1177/1461445605054407

Bucholtz, M., Liang, A. C., & Sutton, L. A. (2014). Language and identity. In T. A. Van Dijk (Ed.), *Discourse studies: A multidisciplinary introduction* (2nd ed., pp. 173–204). London: Sage.

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open science, open data, and open scholarship: European policies to make science fit for the 21st century. *Frontiers in Big Data, 2*, 43. https://doi.org/10.3389/fdata.2019.00043

Burnard, L. (2004). *What is the Text Encoding Initiative?* (Rev. ed.). Providence, RI: Brown University & Oxford: Oxford University Press.

Burrell, J. (2009). The field site as a network: A strategy for locating ethnographic research. *Field Methods, 21*(2), 181–199. https://doi.org/10.1177/1525822X08329699

Busch, B. (2017). Expanding the notion of the linguistic repertoire: On the concept of Spracherleben. *Applied Linguistics, 38*(3), 340–358. https://doi.org/10.1093/applin/amv030

Buysse, A. (2012). *Developing intercultural communicative competence in a multilingual context*.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge University Press. https://doi.org/10.1017/S0008413100002425

Bybee, J. L. (2006). *Frequency of use and the organisation of language.* Oxford University Press.

Byers-Heinlein, K., Jardak, A., Fourakis, E., & Lew-Williams, C. (2022). Effects of language mixing on bilingual children's word learning. *Bilingualism: Language and Cognition*, *25*(1), 55-69.

Byram, M. (2008). *From foreign language education to education for intercultural citizenship: Essays and reflections*. Clevedon: Multilingual Matters.

Byram, M. (2021). *Teaching and assessing intercultural communicative competence: Revisited*. Bristol: Multilingual Matters.ByteDance Ltd. 2023 Capcut

ByteDance Ltd. (2023). *CapCut* [Computer software]. https://www.capcut.com

Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., et al. (2020). Language resources for the future: An overview of ELRA activities. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 6–12. Marseille: ELRA.

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*, 4(6), 377–389. https://doi.org/10.1111/j.1749-818X.2010.00201.x

Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal*, 95(3), 401–417. https://doi.org/10.1111/j.1540-4781.2011.01207.x

Canagarajah, S. (2012). *Translingual practice: Global Englishes and cosmopolitan relations*. New York: Routledge.

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations.* Routledge.

Canagarajah, S. (Ed.). (2017). *The Routledge handbook of migration and language*. London: Routledge.

Candelier, M., Camilleri-Grima, A., Castellotti, V., de Pietro, J.-F., Lőrincz, I., Meißner, F.-J., Schröder-Sura, A., & Noguerol, A. (2012). *CARAP – A framework of reference for pluralistic approaches to languages and cultures*. Strasbourg: Council of Europe.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Cenoz, J., & Gorter, D. (2015). *Multilingual education: Between language learning and translanguaging*. Cambridge: Cambridge University Press.

Cenoz, J., & Gorter, D. (2021). *Pedagogical translanguaging*. Cambridge: Cambridge University Press.

Cervini, C., & Paone, R. (2025). Annotazione di dati orali in contesti di interazione plurilingue: insegnare l'intercomprensione a studenti di scienze mediche veterinarie. In *Insegnare l'intercomprensione*. Roma: Aracne Editrice. https://doi.org/10.53136/97912218188264Charmaz 2014

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.

Chen, P., & Pun, N. (2025). 'Re-coupling site'of social media and the workplace in digital nomadism: Chinese female workers' self-produced vlogs with class and gender connotations. *Information, Communication & Society*, *28*(3), 435-451.

Chevalier, J. M., & Buckles, D. J. (2013). *Participatory action research: Theory and methods for engaged inquiry*. London: Routledge.

Chevtaeva, E. (2024). Workplace Transformation: Exploring the Impact of Digital Nomad-Inspired Travel Experiences. In: Berezina, K., Nixon, L., Tuomi, A. (eds) *Information and Communication Technologies in Tourism 2024*. ENTER 2024. Springer Proceedings in Business and Economics. Springer, Cham. https://doi.org/10.1007/978-3-031-58839-6_25

Chevtaeva, E., Denizci-Guillet, B. (2021). 'Digital nomads' lifestyles and coworkation', *Journal of Destination Marketing & Management, 21, 100633.* Amsterdam: Elsevier.

CLARIN. (2020). *CLARIN recommendations for corpus metadata and documentation*. CLARIN ERIC.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.Clark 1996

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*(3), 201–242. https://doi.org/10.1006/cogp.1998.0693

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography*. Berkeley: University of California Press.

Cocola-Gant, A., Gago, A., & Jover, J. (2020). Tourism, gentrification and neighbourhood change: An analytical framework–Reflections from Southern European cities. *The overtourism debate: NIMBY, nuisance, commodification* (pp. 121-135). Emerald publishing limited.

Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*. London: Continuum.

Cohen, S. A., Duncan, T., & Thulemark, M. (2015). Lifestyle mobilities. *Journal of Transport Geography, 43*, 163–175. https://doi.org/10.1016/j.jtrangeo.2014.12.001

Cook, D. (2020). The freedom trap: digital nomads and the use of disciplining practices to manage work/leisure boundaries. *Inf Technol Tourism* **22**, 355–390 (2020). https://doi.org/10.1007/s40558-020-00172-4

Cook, D. (2023). What is a digital nomad? Definition and taxonomy in the era of mainstream remote work. *World Leisure Journal*, *65*(2), 256–275. https://doi.org/10.1080/16078055.2023.2190608

Cook, G. (2010). *Translation in language teaching*. Oxford: Oxford University Press.

Copland, F., & Creese, A. (2015). *Linguistic ethnography*. London: Sage.

CoreTrustSeal. (2023). *CoreTrustSeal requirements 2023–2025*.

Costello, A. B., & Osborne, J. W. (2017). Best practices in exploratory factor analysis. *Practical Assessment, Research & Evaluation, 10*(7).

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford: Stanford University Press.

Coulmas, F. (1998). Language contact. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 314–329). Oxford: Blackwell.

Council of Europe (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume, Council of Europe Publishing, Strasbourg.

Council of Europe. (2000). *Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin*. Official Journal of the European Communities.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.*Cambridge University Press.

Council of Europe. (2008). *White Paper on Intercultural Dialogue: Living together as equals in dignity*. Strasbourg: Council of Europe.

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume with new descriptors*. Strasbourg: Council of Europe.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Strasbourg: Council of Europe Publishing.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Strasbourg: Council of Europe Publishing.

Couper-Kuhlen, E., & Selting, M. (2018). *Interactional linguistics*. Cambridge: Cambridge University Press.

Couper, M. P. (2008). *Designing effective web surveys*. Cambridge: Cambridge University Press.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.

Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics, 32*, 5–20.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Cronin, M. (2013). *Translation in the digital age*. London: Routledge.

Cummins, J. (2007). Rethinking monolingual instructional strategies. *Canadian Journal of Applied Linguistics, 10*(2), 221–240.

Dalisaymo, Lyn & Patalay, Myra. (2024). Artificial Intelligence Impact on Digital Nomad Work Life Balance: The Role of Automation. 1-6.

Davies, B. (2008). Reflexivity, methodology, and ethics. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 725–740). Thousand Oaks, CA: Sage.

Davies, M. (2008). Data collection and interpretation in corpus linguistics. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 247–259). Berlin: Mouton de Gruyter.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 520 million words, 1990–present. *International Journal of Corpus Linguistics, 14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

De Decker, P. (2020). Housing and digital nomadism: Urban change, mobility, and precarity. *Housing Studies, 35*(6), 1121–1139.

De Fina, A. (2021). Doing identity in narrative. In A. De Fina & A. Georgakopoulou (Eds.), *The handbook of narrative analysis* (pp. 351–368). Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781119577111.ch18

De Paiva Duarte, M. (2013). Communities of practice and language learning. *Revista Brasileira de Linguística Aplicada, 13*(1), 37–57.

de Sousa, B. V., Bastida, M., & Vieira, E. (2025). Unraveling digital nomadism: a comprehensive bibliometric review of emerging research and trends. *Journal of Global Mobility: The Home of Expatriate Management Research*, *13*(1), 107-132.

Deardorff, D. K. (2009). *The SAGE handbook of intercultural competence*. Thousand Oaks, CA: Sage.

Demaj, M., Shehu, E., & Kotorri, A. (2021). Digital nomadism and new forms of work mobility. *Journal of Tourism Futures, 7*(3), 322–336.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (5th ed.). Thousand Oaks, CA: Sage.

Di Garbo, F. (2023). Language contact and morphosyntactic change. *Annual Review of Linguistics, 9*, 109–127. https://doi.org/10.1146/annurev-linguistics-031422-111455

Dijck, J. van. (2013). *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.

Dillman, D. A. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: Wiley.

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). Language friction and multilingual policies in higher education: the stakeholders' view. *Journal of Multilingual and Multicultural Development*, *35*(4), 345-360.

Dörnyei, Z. (2016). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in SLA research. In A. Mackey & S. Gass (Eds.), *Research methods in second language acquisition* (pp. 74–94). Malden, MA: Wiley-Blackwell. https://doi.org/10.1002/9781444347340.ch5

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). London: Routledge.

Dreher, S., & Triandafyllidou, A. (2023). Digital nomadism and the reconfiguration of mobility regimes. *Global Networks, 23*(2), 189–206.

Dreher, S., & Triandafyllidou, A. (2025). Digital nomads and mobile citizenship. *Journal of Ethnic and Migration Studies*.

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse* (pp. 139–182). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164.07du

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000–2005). *Santa Barbara Corpus of Spoken American English (Parts 1–4).* Linguistic Data Consortium.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data* (pp. 45–89). Hillsdale, NJ: Lawrence Erlbaum.

Duff, P. A. (2020a). Informal language learning. In S. May (Ed.), *The Routledge handbook of language and migration* (pp. 245–258). Milton Park: Routledge.

Duff, P. A. (2020b). Language socialisation: Linking micro practices and macro structures. In A. De Fina & A. Georgakopoulou (Eds.), *The handbook of narrative analysis* (pp. 343–350). Oxford: Wiley-Blackwell.

Duff, P. A., & Anderson, T. (2015). Academic language and literacy socialisation for second language students. *The handbook of classroom discourse and interaction*, 337-352.

Duffy, B. E. (2017). *(Not) getting paid to do what you love: Gender, social media, and aspirational work*. New Haven, CT: Yale University Press.

Duranti, A. (1997). *Linguistic anthropology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511810190

Duranti, A., Ochs, E., & Schieffelin, B. (2012). The handbook of language socialisation. Oxford: Wiley-Blackwell.

Dwyer, S. C., & Buckle, J. L. (2009). The space between: On being an insider-outsider in qualitative research. *International Journal of Qualitative Methods, 8*(1), 54–63. https://doi.org/10.1177/160940690900800105

Eckert, P. (2016). Variation, meaning and social change. *English Language and Linguistics, 20*(2), 317–342. https://doi.org/10.1017/S1360674316000111

Eggers, J. (2021). Digital nomads and the reconfiguration of work-life boundaries. *Organisation Studies*.

Eggers, J., Hatak, I., Kraus, S., & Niemand, T. (2017). Technologies That Support Marketing and Market Development in SMEs—Evidence from Social Networks*, Journal of Small Business Management*, 55(2), 270-302. Taylor & Francis Journals.

Ehn, K., Jorge, A., & Marques-Pita, M. (2022). Digital nomads and the COVID-19 pandemic: Narratives about relocation in a time of lockdowns and reduced mobility. *Social Media+ Society*, *8*(1).

Ehn, M., Nilsson, E. M., & Topgaard, R. (2023). Digital nomadism and everyday infrastructures. *Mobilities*.

Ellis, C., & Bochner, A. P. (2006). Analyzing analytic autoethnography. *Journal of Contemporary Ethnography, 35*(4), 429–449. https://doi.org/10.1177/0891241605280449

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 12*(1).

Ely, R., & Gleason, J. B. (1996). Socialisation across contexts. *The Handbook of Child Language* (eds. Fletcher, P., MacWhinney, B.). https://doi.org/10.1111/b.9780631203124.1996.00010.x

Errattahi, R., et al. (2018). Automatic speech recognition for multilingual contexts. *Procedia Computer Science, 128*, 305–312.

ESFRI. (2021). *ESFRI roadmap 2021: Strategy report on research infrastructures*. European Commission.

European Commission. (2007). *Multilingualism: An asset for Europe and a shared commitment* (COM(2007) 566 final). Brussels: European Commission.

European Commission. (2016a). *H2020 programme: Guidelines on data management*. Brussels: European Commission.

European Commission. (2016b). *Responsible research and innovation: Europe's ability to respond to societal challenges*. Brussels: European Commission.

European Commission. (2017). *Communication on a comprehensive approach to the teaching and learning of languages.* Publications Office of the European Union.

European Commission. (2017). *Strengthening European identity through education and culture* (COM/2017/673). Brussels: European Commission.

European Commission. (2018). *Horizon Europe: Investing to shape our future*. Brussels: European Commission.

European Commission. (2018). *Recommendation on access to and preservation of scientific information* (2018/790/EU). Brussels: European Union.

European Commission. (2020). *White paper on artificial intelligence: A European approach to excellence and trust*. Brussels: European Commission.

European Data Protection Board. (2020). *Guidelines on the processing of personal data under GDPR*. Brussels: EDPB.

European Union. (2007). *United in diversity*. Luxembourg: Publications Office of the European Union.

Eurostat. (2022). *Migration and migrant population statistics*. Luxembourg: Eurostat.

Extra, G., & Spotti, M. (2009). Language diversity in Europe today. *European Journal of Applied Linguistics, 1*(1), 1–20.

Falzon, M.-A. (Ed.). (2009). *Multi-sited ethnography: Theory, praxis and locality in contemporary research*. Farnham: Ashgate.

Fasold, R. (1984). *The sociolinguistics of society*. Oxford: Blackwell.

Fasold, R. (1990). *The sociolinguistics of language*. Oxford: Blackwell.

Favell, A. (2008). *The new face of East–West migration in Europe*. Journal of Ethnic and Migration Studies, 40(5), 701–716.

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science* (pp. 17–47). Cham: Springer. https://doi.org/10.1007/978-3-319-00026-8_2

Fechter, A.-M., & Walsh, K. (2013). Examining 'expatriate'continuities: Postcolonial approaches to mobile professionals. *The New Expatriates* (pp. 9-22). Routledge.

Firth, A. (1996). The discursive accomplishment of normality. *Journal of Pragmatics, 26*(2), 237–259. https://doi.org/10.1016/0378-2166(96)00014-8

Flatio. (2023). *Report on digital nomads*. https://www.flatio.com

Flatio. (2024). *Digital nomad report 2024*. https://www.flatio.com

Flew, T. (2018). *Globalisation, media and culture*. London: Palgrave.

Flores, N., & García, E. S. (2020). Power, language, and bilingual learners. In *Handbook of the cultural foundations of learning* (pp. 178-192). Routledge.

Floridi, L. (2015). *The onlife manifesto*. Springer Nature.

Flowerdew, J. (2014). Ethnographically inspired approaches to the study of academic discourse. In *Academic discourse* (pp. 245-262). Routledge.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words. *Journal of Memory and Language, 34*(6), 709–738. https://doi.org/10.1006/jmla.1995.1032

Franzke, A. S., Bechmann, A., Zimmer, M., & Ess, C. (2020). Internet research: Ethical guidelines 3.0. *Association of Internet Researchers*.

Fried, M., Van Trijp, R. (2004-2026). *Constructional Approaches to Language* [Book series]. John Benjamins Publishing Company.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. London: Palgrave Macmillan.

García, O., Johnson, S. I., & Seltzer, K. (2017). *The translanguaging classroom: Leveraging student bilingualism for learning*. Philadelphia: Caslon.

Garrett, P. (2010). *Attitudes to language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511844713

Gattiker, U. E., & Stollemmaier, R. S. (Eds.). (2012). Technology-mediated communication (Vol. 3). Walter de Gruyter.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Germann Molz, J. (2021). *The world is our office: Digital nomads and the globalisation of work*. Polity Press.

Getman, N. (2021). Developing a Digital Nomads Destination from the Ground Up. Master's Thesis in Tourism Development and Culture. University of Glasgow, University of Malta.

Giles, H., & Hewstone, M. (1982). Cognitive structures in intergroup communication. In H. Giles (Ed.), *Intergroup behaviour* (pp. 161–183). Oxford: Blackwell.

Gilquin, G. (2016). The place of corpora in grammatical theory. *Lingua, 175–176*, 5–18. https://doi.org/10.1016/j.lingua.2015.07.008

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Glick Schiller, N., & Salazar, N. B. (2013). Regimes of mobility. *Journal of Ethnic and Migration Studies, 39*(2), 183–200. https://doi.org/10.1080/1369183X.2013.723253

Global Policy Institute. (2024). *Global policy report on digital mobility*.

Godfrey, J. J., & Zampolli, A. (1997). *Readings in spoken language processing*. San Francisco: Morgan Kaufmann.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalisation in language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 640–665. https://doi.org/10.1037/a0014981

Götz, S. (2013). Fluency in native and non-native English speech. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.53

Graffigna, G., & Barello, S. (2018). Spotlight on the Patient Health Engagement model (PHE model): a psychosocial theory to understand people's meaningful engagement in their own health care. *Patient preference and adherence,* 1261-1271.

Green, A. (2020). Digital nomads and the reconfiguration of work. *Work, Employment and Society*.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Guirado, O., Dosil-Santamaria, M., & Miguelena, J. (2025). Emotional and conscious translanguaging in non-formal education: linguistic socialisation in residential care homes in a minority language setting. *Journal of Multilingual and Multicultural Development*, 1-19.

Gumperz, J. J. (1968). The speech community. *International Encyclopedia of the Social Sciences*. New York: Macmillan.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511611834

Hammarström, H. (2016). Linguistic diversity and language documentation. *Language Documentation & Conservation, 10*, 1–25.

Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in practice* (4th ed.). London: Routledge.

Hannonen, O. (2020). In search of a digital nomad. *Information Technology & Tourism, 22*, 335–353. https://doi.org/10.1007/s40558-020-00177-z

Hannonen, O. (2024). Digital nomadism as mobility regime.

Haraway, D. (1988). Situated knowledges. *Feminist Studies, 14*(3), 575–599. https://doi.org/10.2307/3178066

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems. *Psychological Bulletin, 136*(2), 206–228. https://doi.org/10.1037/a0017782

Hatton, J., Holton, G., Seyfeddinipur, M., Thieberger, N. (2021). Lameta [software] https://github.com/onset/laMETA/releases

Haugen, E. (1972). *The ecology of language*. Stanford: Stanford University Press.

Haugh, M. (2013). Disentangling face, facework and politeness. *Sociocultural Pragmatics, 1*(1), 46–73

Heller, M. (2008). Language and the nation-state: Challenges to sociolinguistic theory and practice 1. *Journal of sociolinguistics*, *12*(4), 504-524.

Heller, M., Bell, L., Daveluy, M., McLaughlin, M., & Noël, H. (2015). *Sustaining the nation.* Oxford: Oxford University Press.

Hermann, I., Paris, C. M. (2020). Digital Nomadism: the nexus of remote working and travel mobility, *Information Technology & Tourism, vol. 22(3)*, pages 329-334. Springer.

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry, 3*(3), 274–294.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet, 4*.

Hett, G., & Hett, J. (2013). Ethics in intercultural research: Reflections on the challenges of conducting field research in a Syrian context. *Compare: A Journal of Comparative and International Education*, *43*(4), 496-515.

Hine, C. (2015). *Ethnography for the internet*. London: Bloomsbury.

Hjorth, L., Horst, H. A., Galloway, A., & Bell, G. (Eds.). (2017). *The Routledge companion to digital ethnography* (pp. 21-28). New York: Routledge.

Holliday, A. (2006). Native-speakerism. *ELT journal, 60(4), 385-387.*

Holliday, A. (2018). *Understanding intercultural communication*. Taylor & Francis.

Holmes, P., Fay, R., Andrews, M., & Attia, M. (2022). *Researching intercultural learning*. London: Routledge. https://doi.org/10.4324/9781003020899

Horwitz, E. K. (2010). Foreign language anxiety. *Language Teaching, 43*(2), 154–167. https://doi.org/10.1017/S026144480999036X

House, J. (1999). Misunderstanding in intercultural communication: Interactions in English as a lingua franca and the myth of mutual intelligibility. In C. Gnutzmann (Ed.), *Teaching and learning English as a global language* (pp. 73–89). Tübingen: Stauffenburg Verlag.

House, J. (1996). Developing pragmatic fluency. *Studies in Second Language Acquisition, 18*(2), 225–252. https://doi.org/10.1017/S0272263100014910

Huston, S., & Francis, G. (2000). Pattern grammar. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.4

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hymes, D. (1964). Toward ethnographies of communication. *American Anthropologist, 66*(6), 1–34. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00020

Hymes, D. (1972). Models of the interaction of language and social life. In J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics* (pp. 35–71). New York: Holt, Rinehart & Winston.

Hymes, D. (1974). *Foundations in sociolinguistics*. Philadelphia: University of Pennsylvania Press.

Hymes, D. (1996). *Ethnography, linguistics, narrative inequality*. London: Taylor & Francis.

Ide, N., & Pustejovsky, J. (2017). Handbook of linguistic annotation. Dordrecht: Springer. https://doi.org/10.1007/978-94-024-0881-2

Ide, N., & Romary, L. (2004). International standard for linguistic annotation. *Natural Language Engineering, 10*(3–4), 211–225. https://doi.org/10.1017/S1351324904003465

Ingvarsson, C. (2023). Connecting and Disconnecting : How Digital Nomads Manage Work in Absence of a Workplace. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 6270–6279.

Isaac, R. K., & Dorrestein, J. (2025). The interaction of digital nomads with the host community and its effects on their quality of life: Chiang Mai as a case study. In *Tourism and Hospitality for Humanity* (pp. 235-259). Edward Elgar Publishing.

Iverson, J.M. (2010). Developing language in a developing body: the relationship between motor development and language development. *J Child Lang*. 37(2):229-61.

Iwamoto, K., Ochiai, T., Delcroix, M., Ikeshita, R., Sato, H., Araki, S., & Katagiri, S. (2022). How bad are artifacts?: Analyzing the impact of speech enhancement errors on ASR. *arXiv preprint arXiv:2201.06685*.

Jain, N., Schwarzschild, A., Wen, Y., Somepalli, G., Kirchenbauer, J., Chiang, P. Y., ... & Goldstein, T. (2023). Baseline defenses for adversarial attacks against aligned language models. *arXiv preprint arXiv:2309.00614*.

Jaspers, J., & Madsen, L. (2019). Critical perspectives on linguistic fixity and fluidity. *Abingdon: Routledge*, *10*.

Jasso, G. (Ed.) (2024). *Frontiers in Sociology*, *8*.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–23). Amsterdam: John Benjamins.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century.* Cambridge, MA: MIT Press.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice 2(3).* 49–85.

Ji, Y., Kim, S. M., & Kim, Y. (2024). A way to attract digital nomads to tourist destinations in the new normal era. *Sustainability*, *16*(6), 2336.

Katz, D. S., et al. (2021). GitHub and FAIR research software. *Computing in Science & Engineering, 23*(2), 70–76.

Kendon, A. (2022). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kharitonov, E., et al. (2023). Self-supervised speech representation learning. *IEEE Signal Processing Magazine*.

Kharitonov, E., et al. (2023). Self-supervised speech representation learning. *IEEE Signal Processing Magazine*.

Kindsmüller, M., Melzer, A., Mentler, T. (2009). Online Communities and Online Community Building. *Encyclopedia of Information Science and Technology*. Hershey, Information Science Publishing.

King, R. (2012). Theories and Typologies of Migration: An Overview and A Primer. *Willy Brandt Series of Working Papers in International Migration and Ethnic Relations, 12.* 1-43.

Kinginger, C. (2013). Identity and language learning in study abroad. *Foreign Language Annals, 46*(3), 339–358. https://doi.org/10.1111/flan.12037

Klein, J. T. (2010). *A taxonomy of interdisciplinarity*. In R. Frodeman et al. (Eds.), *The Oxford handbook of interdisciplinarity*. Oxford: Oxford University Press.

Knight, D., & Adolphs, S. (2021). Multimodal corpora. In *A practical handbook of corpus linguistics* (pp. 353-371). Cham: Springer International Publishing.

Koenecke, A., et al. (2020). Racial disparities in speech recognition. *Proceedings of the National Academy of Sciences, 117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Koester, A. (2010). *Workplace discourse.*

Kopper, S., Parry, K., Kannan, H. (2023). Survey Design. https://www.povertyactionlab.org/resource/survey-design

Koskela, K., & Beckers, P. (2024). Typologizing Digital Nomad Visas: Comparing Policy Rationales from Tourism to Wealth Accumulation to Immigration. *Journal of Comparative Policy Analysis: Research and Practice*, *26*(5), 407–425. https://doi.org/10.1080/13876988.2024.2343696

Kozak, M., Cetin, G., & Alrawadieh, Z. (2024). Repositioning work and leisure: Digital nomads versus tourists. *International Journal of Tourism Research*, *26*(4), e2732.

Kozinets, R.V. (2015). Netnography. *The international encyclopedia of digital communication and society.* Sage.

Kozinets, R.V. (2020). *Netnography: The essential guide to qualitative social media research.* Sage Publications Limited.

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication.* Routledge.

Kroll, J. F. (2017). The bilingual lexicon: A window into language dynamics and cognition. In *Bilingualism: A framework for understanding the mental lexicon* (pp. 27-48). John Benjamins Publishing Company.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*(3), 213-236.

Krzyżanowski, M., & Wodak, R. (2011). Political strategies and language policies: The European Union Lisbon strategy and its implications for the EU's language and multilingualism policy. *Language Policy*, *10*(2), 115-136.

Kunz, S. (2016). Privileged mobilities: Locating the expatriate in migration scholarship. *Geography Compass, 10(3), 89-101.*

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1992). *Language variation and change*. Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change, Vol. 2: Social factors*. Oxford: Blackwell.

Labrecque, L. I., et al. (2011). Consumer perceptions and brand personality. *Journal of Marketing, 75*(4), 37–53. https://doi.org/10.1509/jmkg.75.4.37

Lagidze, N., & Pirtskhalava, E. (2024). The Ties That Bind: Exploring Place Identity and Attachment Among Digital Nomads.

Lassiter, L. E. (2005). *The Chicago guide to collaborative ethnography*. Chicago: University of Chicago Press.

Lee, A., Toombs, A. L., Erickson, I., Nemer, D., Ho, Y. S., Jo, E., & Guo, Z. (2019). The social infrastructure of co-spaces: Home, work, and sociable places for digital nomads. *Proceedings of the ACM on human-computer interaction*, *3*(CSCW), 1-23.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2*(4), 34–46. https://doi.org/10.1111/j.1540-4560.1946.tb02295.x

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Evolution of Wenger's concept of community of practice. *Implementation science*, *4*(1), 11.

Lickley, R. (2015). Fluency and disfluency. *Journal of Phonetics, 52*, 1–14. https://doi.org/10.1016/j.wocn.2015.06.002

Linao, P. A., Heimtun, B., Morgan, N. (2024). Digital nomadism, gender and racial power relations, *Annals of Tourism Research,* Volume 106.

Lindström, J. (2025). Constructions in spoken interaction. *The Cambridge Handbook of Construction Grammar*, 309-334.

Love, N. (2017). *Language and its study*. London: Routledge. https://doi.org/10.4324/9781315462992

Love, R., et al. (2017). The spoken BNC2014. *International Journal of Corpus Linguistics, 22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Ma, R., Van Hoey, T., Szmrecsanyi, B. (2025). Isomorphism-inspired theorising about optionality and variation is unsupported by data from English grammar. English Language and Linguistics. 1-21.

Maastricht Treaty. (1992). Treaty on European Union. Official Journal of the European Communities.

Mahlberg, M. (2014). Corpus linguistics and discourse analysis. In K. P. Schneider & A. Barron (Eds.), *Pragmatics of discourse* (pp. 215–238). Berlin: De Gruyter Mouton.

Makimoto, T., & Manners, D. (1997). *Digital nomad*. Chichester: John Wiley & Sons.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.

Mancinelli, F. (2020). Digital nomadism: Freedom, responsibility and the neoliberal subject. *Mobilities, 15*(3), 417–430.

Mancinelli, F., & Germann Molz, J. (2023). Living and working on the move: Digital nomadism, mobility and lifestyle migration. In J. Germann Molz & C. Paris (Eds.), *Lifestyle mobilities: Intersections of travel, leisure and migration*. London: Routledge.

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology, 24*, 95–117. https://doi.org/10.1146/annurev.an.24.100195.000523

Markham, A. (2018). *Critical approaches to ethics in digital research*. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection*. London: SAGE.

Mauranen, A. (2003). The corpus of English as a lingua franca in academic settings. *TESOL Quarterly, 37*(3), 513–527. https://doi.org/10.2307/3588402

Mauranen, A. (2006*).* Signaling and preventing misunderstanding in English as lingua franca communication. *International Journal of the Sociology of language, 177.* https://doi.org/10.1515/IJSL.2006.008

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

Mauranen, A., & Ranta, E. (2006). *English as a lingua franca: Studies and findings*. Cambridge Scholars Publishing.

Mauranen, A., & Metsä-Ketelä, M. (Eds.). (2006). English as a Lingua Franca*. Special Issue of The Nordic Journal of English Studies.*

Mauranen, Anna. (2023). Reflexively Speaking: Metadiscourse in English as a Lingua Franca. 10.1515/9783110295498.

Mazzaferro, G. (2018). Translanguaging and linguistic repertoires in multilingual practices. *Applied Linguistics Review, 9*(4), 595–620. https://doi.org/10.1515/applirev-2017-0060

MBO Partners (2024). *State of Independence Report 2024*. Washington, DC: MBO Partners.

MBO Partners (2025). *State of Independence Report 2025*. Washington, DC: MBO Partners.

McCarthy, M. (2006). Spoken language and applied linguistics. Cambridge: Cambridge University Press.

McCarthy, M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Messenger, J.C., Gschwind, L. (2016). Three generations of Telework: New ICTs and the (R)evolution from Home Office to Virtual Office. *New Technology, Work and Employment in the era of Covid-19: the contribution of the journal and its past publications to the current debate* 31 (3), pp. 195-208. https://doi.org/10.1111/ntwe.12073

Microsoft Corporation (2023). *Visual Studio Code* (Version 1.x) [Computer software]. https://code.visualstudio.com

Miguel, C., Lutz, C., & Perez-Vega, R. (2024). Digital nomads and place attachment: Negotiating mobility, belonging and work. *Journal of Sustainable Tourism*.

Milroy, L., & Milroy, J. (1992). Social network and social class. *Language in Society, 21*(1), 1–26. https://doi.org/10.1017/S0047404500015013

Miočević, D., et al. (2025). Digital nomadism, entrepreneurial ecosystems and transnational careers. *Entrepreneurship & Regional Development*.

Mohlin, A., & Chatzipanagiotou, N. (2023). Digital Nomads' Experiences on the Support of Digital Technologies in Relation to Social Isolation. In *Proceedings of the 21st European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centered Computing on the Design of Cooperation Technologies* - Exploratory Papers, Reports of the European Society for Socially Embedded Technologies https://doi.org/10.48340/ecscw2023_ep01

Monachini, M., & Frontini, F. (2005). TEI-based spoken corpora annotation: Principles and practices. In *Proceedings of the Corpus Linguistics Conference 2005*. Birmingham: University of Birmingham.

Moret, J. (2020). Mobility capital: Somali migrants' trajectories of (im)mobilities and the negotiation of social inequalities across borders. *Geoforum*, *116*, 235-242.

Murphy-Lejeune, E. (2002). *Student mobility and narrative in Europe: The new strangers*. London: Routledge.

Nash, C., Hine, C., & Miller, D. (2021). Netnography. In *The SAGE handbook of qualitative research* (5th ed.). London: Sage.

Nero, S. J. (2024). English as a Lingua Franca communication today. *Invited talk*, ELF Communication Today Conference.

Nicolescu, B. (2014). *From modernity to cosmodernity: Science, culture, and spirituality*. Albany: SUNY Press.

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: a multimodal interaction analysis. *Visual communication, 10(2), 129-147.*

Nowicka, M., & Ryan, L. (2015). Beyond insiders and outsiders in migration research: Rejecting a priori commonality. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 16*(2).

Nowicka, M., & Ryan, L. (2015). Beyond insiders and outsiders in migration research: Rejecting a priori commonality. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 16*(2).

O'Hagan, M. (2016). Massively open translation: Unpacking the relationship between technology and translation. *International Journal of Communication, 10*, 929–946.

O'Hagan, M. (2016). Massively open translation: Unpacking the relationship between technology and translation. *International Journal of Communication, 10*, 929–946.

Ochs, E. (1988). *Culture and language development*. Cambridge: Cambridge University Press.

Ochs, E., & Schieffelin, B. (2011). The theory of language socialisation. In A. Duranti et al. (Eds.), *The handbook of language socialisation*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444342901.ch1

Orel, M. (2019). Coworking environments and digital nomadism: The role of workspaces in new work practices. *Review of Managerial Science, 13*(3), 649–669.

Orel, M., Dvouletý, O., Ratten, V. (2021). *The flexible workplace*. Springer.

Otsuji, E., & Pennycook, A. (2010). Metrolingualism: Fixity, fluidity and language in flux. *International Journal of Multilingualism, 7*(3), 240–254. https://doi.org/10.1080/14790710903414331

Özyürek, A., et al. (2024). Mobility, language practices and identity among digital nomads. *Language in Society*.

Paone, E., Tedesco, N., Cervini, C., Höhmann, D. (2025). *Plurilingual Treasure Hunt.* Paper presented at DILLE 2025 Conference, Tirana.

Qosim, N., Rukmana, A. Y., Usup, U. (2025). The Impact of Digital Nomad Economy on Local Businesses Study of MSME Market Changes in Popular Destinations in Indonesia. *The Eastasouth Journal of Social Science and Humanities (ESSSH). Vol. 2 No. 02 (2025).* Jakarta_ Eastasouth Institute. https://doi.org/10.58812/esssh.v2i02.474.

Tedesco, N., Paone, E., Cervini, C. (forth.). Exploring plurilingual communication in spontaneous and task-based interactions: what insights for language education?. *Bridging Voices in Plurilingual Education* (eds. Cortés Velásquez, D., Gkaintartzi, A.). Peter Lang.

Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001

Parreño-Castellano, J., Domínguez-Mujica, J., & Moreno-Medina, C. (2022). Reflections on digital nomadism in Spain during the COVID-19 pandemic—effect of policy and place. *Sustainability*, *14*(23), 16253.

Partington, M., Smith, T. (2022). The world's first digital nomad village, one year on. Sifted.eu

Pennycook, A. (2007). *Global Englishes and transcultural flows*. London: Routledge.

Pennycook, A. (2010). *Language as a local practice*. London: Routledge. https://doi.org/10.4324/9780203846223

Pennycook, A. (2017). *The cultural politics of English as an international language.* Routledge.

Pennycook, A., & Otsuji, E. (2015). *Metrolingualism*. London: Routledge.

Pescosolido, B. A. (2007). The sociology of social networks. In *21st century sociology* (pp. I-208). SAGE Publications, Inc.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Piano Nazionale per la Scienza Aperta (2022). *PNAS*. Rome: Ministero dell'Università e della Ricerca.

North, B. & Piccardo, E. (2016). Developing illustrative descriptors of aspects of mediation for the CEFR. *Education Policy Division, Council of Europe.*

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., Tacchi, J. (2016). *Digital ethnography: Principles and practice*. London: Sage.

Polson, E. (2024). From the tag to the# hashtag: Street art, Instagram, and gentrification. *Space and Culture, 27(1)*. SAGE Pubblications.

Poole, D. (1992). Language socialisation and informal learning. *Applied Linguistics, 13*(2), 185–205.

Prengler, S., et al. (2025). Digital nomadism and regional development. *Regional Studies*.

Prinsloo, M. (2023). Semiotic repertoires and mobility. *Applied Linguistics, 44*(2), 215–234. https://doi.org/10.1093/applin/amac040

Pym, A. (2021). Translation and language learning as policy options. Questions of costs and literacy development. *Translation and Interpreting, 13*(1): 24-37. https://doi.org/10.12807/ti.113201.2021.a02.

Qotim, A., et al. (2025). Digital nomads, remote work and transnational lifestyles. *Journal of Global Mobility*.

Radford, A. & Kim, J. & Xu, T. & Brockman, G., McLeavey, C. & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. 10.48550/arXiv.2212.04356.

Raja, D. N., & Giri, K. J. (2025). AUDIRE: a comprehensive review of speech recognition technologies–methods, uses, and challenges. *International Journal of Speech Technology*, 1-27.

Rampton, B. (1990). Displacing the "native speaker": Expertise, affiliation and inheritance. *ELT Journal, 44*(2), 97–101. https://doi.org/10.1093/elt/44.2.97

Rampton, B. (2007). Linguistic ethnography, interactional sociolinguistics and the study of identities. *Working papers in Urban Language and Literacies*, 43.

Rampton, B., Tusting, K., Maybin, J., Barwell, R., Creese, A., Lytr, V. (2004). UK linguistic ethnography: A discussion paper. *Working Papers in Urban Language & Literacies, 13*. London: King's College.

Rampton, B., Maybin, J., & Roberts, C. (2015). Theory and method in linguistic ethnography. In *Linguistic ethnography: Interdisciplinary explorations* (pp. 14-50). London: Palgrave Macmillan UK.

Rasnača, Z. (2023). Digital nomad visas and labour law implications. *European Labour Law Journal, 14*(3), 347–365.

Reason, P., & Bradbury, H. (2008). *The SAGE handbook of action research* (2nd ed.). London: Sage.

Reichenberger, Ina (2018). Digital nomads – a quest for holistic freedom in work and leisure. *Annals of Leisure Research*, 21:3, 364-380. Routledge.

Richards 2021 digital nomadism

Richter, S., & Richter, A. (2020). Digital Nomads. *Business & Information Systems Engineering, 62(*1), 77–81.

Riesgo Gómez, V., Cortez, P., Gil, J., Sequera, J. (2025). Beyond the digital nomad: Transnational digital workers in Lisbon, *Digital Geography and Society*, Volume 9, 100148. https://doi.org/10.1016/j.diggeo.2025.100148.

Rowe, D. (2018). Digital media, sport and culture. *Communication & Sport, 6*(4), 452–468. https://doi.org/10.1177/2167479517719879

RTVE (2025). Turismo e lavoro da remoto in Spagna. *RTVE News*.

Sardinha, L., Vieira, M. & Sousa, A. & Leite, E. (2023). DIGITAL NOMADS — THE CASE OF THE AUTONOMOUS REGION OF MADEIRA. *Journal of Entrepreneurial Researchers.* 1.

Saunders, B., Kitzinger, J., & Kitzinger, C. (2015). Anonymising interview data: Challenges and compromise. *Qualitative Research, 15*(5), 616–632. https://doi.org/10.1177/1468794114550439

Scatena, A., & Gaggioli, A. (2020). Immersive technologies and social interaction. *Frontiers in Psychology, 11*, 564678.

Schegloff, E. A. (2007). *Sequence organisation in interaction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791204

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*(4), 289–327. https://doi.org/10.1515/semi.1973.8.4.289

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organisation of repair in conversation. *Language*, 53(2), 361–382.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D. (Ed.) (2012). *The handbook of discourse analysis*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118584194

Schilling, N. (2012). *Sociolinguistic fieldwork*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165976

Schilling, N. (2013). Survey methods in sociolinguistics. *Journal of Linguistic Geography, 1*(2), 114–120.

Schlagwein, D. (2017). How remote work and digital nomadism transform organizing. *Information Systems Journal, 27*(3), 345–378. https://doi.org/10.1111/isj.12145

Schlagwein, D. (2018). Escaping the 'rat race': Justifications in digital nomadism. In *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK. https://aisel.aisnet.org/ecis2018_rp/131

Schleef, E. (2014). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.

Schmidt, T. (2014). EXMARaLDA. In *Handbook on Corpus Phonology* (pp. 402–419). Oxford: Oxford University Press.

Schmidt, T. (2016). TEI-based transcription and annotation of spoken language. *Journal of the Text Encoding Initiative*, Issue 10. https://doi.org/10.4000/jtei.1568

Scholte, J. A. (2007). Defining globalisation. *Clm. Economia, 10*, 15-63. Asociación de Economía de Castilla-La Mancha.

Scholz, R. W., Köckler, H., Zscheischler, J., Czichos, R., Hofmann, K.-M., & Sindermann, C. (2024). Transdisciplinary knowledge integration, Part II: Experiences of five transdisciplinary processes on digital data use in Germany. *Technological Forecasting and Social Change*, 199, 123123. https://doi.org/10.1016/j.techfore.2023.122981

Science Europe (2021). *Practical guide to the international alignment of research data management*. Brussels: Science Europe.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. London: Routledge. https://doi.org/10.4324/9780203851357

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.

Seidlhofer, B., et al. (2013). The Vienna–Oxford International Corpus of English (VOICE). *Corpus Linguistics and Linguistic Theory, 9*(2), 241–251. https://doi.org/10.1515/cllt-2013-0010

Sequera, J. (2024). Digital nomadism and urban transformations. *Urban Geography*.

Sequera, J., & Nofre, J. (2019). Touristification, transnational gentrification and urban change. *Urban Studies, 56*(15), 3121–3138. https://doi.org/10.1177/0042098019840890

Shawkat, S., Abd Rozan, M. Z., Salim, N. B., & Shehzad, H. M. F. (2021). Digital nomads: A systematic literature review. *2021 7th international conference on research and innovation in information systems (ICRIIS).*

Sheller, M. (2018). *Mobility justice*. London: Verso.

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A, 38*(2), 207–226. https://doi.org/10.1068/a37268

Shriberg, E. (2001). To "errrr" is human. *Journal of the International Phonetic Association, 31*(1), 153–169. https://doi.org/10.1017/S0025100301001128

Silverstein, M. (1993). Metapragmatic discourse and metapragmatic function. In J. Lucy (Ed.), *Reflexive language* (pp. 33–58). Cambridge: CUP.

Šímová, L. (2023). A research framework for digital nomadism: a bibliometric study. *World Leisure Journal*, *65*(2), 175-191.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Skehan, P. (2009). Modelling second language performance. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Spolsky, B. (2011). *Sociolinguistics*. Oxford: Oxford University Press.

Spradley, J. P. (1980). *Participant observation*. New York: Holt, Rinehart & Winston.

Startup Madeira (2021). *Digital Nomad Village Madeira report*. [institutional report] Funchal: Startup Madeira. Startup Madeira 2021 on digital nomads

Stebbins, R. A. (2011). Exploratory research in the social sciences. *Qualitative Sociology Review, 7*(1), 6–25.

Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.

Szmrecsanyi, B. (2014). *Grammatical variation in British English dialects*. Cambridge: CUP. https://doi.org/10.1017/CBO9781139342810

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752. https://doi.org/10.1126/science.aat5991

Tagg, C. (2023). Digital language and communication. In *The Routledge handbook of applied linguistics* (pp. 68-80). Routledge.

Tagliamonte, S. (2012). *Variationist sociolinguistics*. Oxford: Wiley-Blackwell.

Taguchi, N. (2019). *The Routledge handbook of second language pragmatics*. London: Routledge. https://doi.org/10.4324/9781351164085

Taltavull, P. (2024). Digital nomads and real estate markets. *Journal of European Real Estate Research*, *17*(1), 1-3.

Tavakoli, P., & Skehan, P. (2005). Strategic Planning, Task Structure, and Performance Testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239-273). John Benjamins. https://doi.org/10.1075/lllt.11.15tav

Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora, 8*(1), 81–113.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling. *Journal of Mixed Methods Research, 1*(1), 77–100. https://doi.org/10.1177/1558689806292430

Tedesco, N. (2024a). A sociolinguistic perspective on digital nomadism in Europe. Paper presentation at *Colloque International Migrations, Soci&ter,* Université de Mons.

Tedesco, A. (2024b). *Where's your language at? Extended pilot*. [dataset]. https://doi.org/10.6092/unibo/amsacta/7628.

Tedesco, N. (2025a) Translanguaging in the era of digital nomadism: a sociolinguistic perspective on voluntary mobility in Europe. Discov glob soc 3, 50 (2025). https://doi.org/10.1007/s44282-025-00191-8

Tedesco, N. (2025b). *Where's your language at?* [dataset]. https://doi.org/10.5281/zenodo.15603903

Tedesco, N. Bernardini, S., Cervini, C. (2026). *NomadLingo 1.1 open* [dataset]. http://hdl.handle.net/20.500.11752/OPEN-2137

Tedesco, N. (forth.). Nomad English: Nomad English: Redefining acceptability and accessibility in the era of mobility and digitalization Acceptability, Accessibility, and Inclusivity in English Communication Globally, ed. By Aiello, J., Iaia P. L. Routledge. (Eds. by Aiello, J., Iaia, P.) *ELF Communication Today: Acceptability and Accessibility in Focus.* Routledge.

TEI Consortium (2025). *TEI Guidelines*. https://tei-c.org

Thompson, A. (2018). Digital nomadism and lifestyle mobility: Work, leisure and the pursuit of freedom. *Mobilities*, 13(1), 1–15. https://doi.org/10.1080/17450101.2017.1323628

Thompson, A. (2019). Intentionality and lifestyle in digital nomadism. *Journal of Lifestyle Studies*, 10(2), 123–139.

Thu Ha, P., Dinh, T. T. H., & Nguyen, T. M. H. (2021). Language, mobility and identity in transnational contexts. *Journal of Multilingual and Multicultural Development*, 42(6), 497–511.

Thurlow, C., & Mroczek, K. (2011). Digital discourse: Language in the new media. Oxford: Oxford University Press.

Tognini-Bonelli, E. (2001). Corpus linguistics at work. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.6

Tusting, K. (Ed.). (2020). *The Routledge handbook of linguistic ethnography.* New York: Routledge.

Tusting, K., & Maybin, J. (2007). Linguistic ethnography and interdisciplinarity. *Journal of Sociolinguistics*, 11(5), 575–597. https://doi.org/10.1111/j.1467-9841.2007.00334.x

Urry, J. (2007). Mobilities. Cambridge: Polity Press.

Van Lier, L. (2004). The ecology and semiotics of language learning: A sociocultural perspective. Boston: Kluwer Academic. https://doi.org/10.1007/978-1-4020-7912-1

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, 16(40), 33–36. https://doi.org/10.7748/ns2002.06.16.40.33.c3214

Varis, P., & Blommaert, J. (2020). Conviviality and collectives on social media. *Multilingual Margins*, 7(1), 1–15.

Vertovec, S. (2007). Super-diversity and Its Implications. *Ethnic and Racial Studies,* 30(6), 1024-1054.

Vesala, H. (2020). Digital nomads: Mobility, work and lifestyle. *Information Technology & Tourism*, 22(3), 335–353. https://doi.org/10.1007/s40558-020-00174-2

Vicente-Saez, R., & Martínez-Fuentes, C. (2018). Open science now: A systematic literature review. *Journal of Business Research*, 88, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

Vogel, S., & García, O. (2017). Translanguaging. In *Oxford Research Encyclopedia of Education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.181

von Zumbusch, J. Sin, H., Lalicic, L. (2020). The role of co-living spaces in digital nomads' well-being. *Information Technology & Tourism* 22:439–453. Springer.

Wang, B., Schlagwein, D., Cecez-Kecmanovic, D., & Cahalane, M. (2019). Digital Nomadism and the Market Economy: Resistance and Compliance. In ICIS.

Webb, A. K. (2024). Digital nomadism and the challenge to social citizenship. *Global Policy 15 (2).* Wiley. https://doi.org/10.1111/1758-5899.13377.

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1), 9–30. https://doi.org/10.1093/applin/amx039

Wenger, E. (1998). *Communities of practice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803938

Wibowo, T.O. (2024). Dynamic of migration lifestyle and settlement of nomads in Bali, Indonesia. *The Indonesian Journal of Geography Studies*.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

Winata, G. I., Lin, Z., Xu, P., et al. (2022). Multilingual speech recognition with self-supervised learning. *Proceedings of Interspeech 2022*, 1096–1100. https://doi.org/10.21437/Interspeech.2022-1095

Windhouwer, M., & Wright, S. (2012). Linking resources using TEI. *Journal of the Text Encoding Initiative, 2*. https://doi.org/10.4000/jtei.527

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of LREC 2006*, 1556–1559.

Woldoff, R. A., & Litchfield, R. C. (2021). *Digital nomads: In search of freedom, community, and meaningful work in the new economy*. Oxford: Oxford University Press.

Woodcock, J., & Graham, M. (2019). *The gig economy: A critical introduction*. Cambridge: Polity Press.

Woolard, K. A. (2020). Language ideology. In J. Stanlaw (Ed.), *The international encyclopedia of linguistic anthropology*. Hoboken: Wiley. https://doi.org/10.1002/9781118786093.iela0210

Xiong, W., Droppo, J., Huang, X., et al. (2016). Toward human parity in conversational speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*(12), 2410–2423. https://doi.org/10.1109/TASLP.2017.2756434

Yılmaz, E., van den Heuvel, H., & van Leeuwen, D. A. (2018). Investigating bilingual code-switching in ASR. *Speech Communication, 102*, 10–22. https://doi.org/10.1016/j.specom.2018.05.002

Zappavigna, M. (2018). *Searchable talk: Hashtags and social media metadiscourse*. London: Bloomsbury.

Zhou, L., Buhalis, D., Fan, D., Ladkin, A. & Lian, X. (2024). Attracting digital nomads: Smart destination strategies, innovation and competitiveness. Journal of Destination Marketing & Management. 31. 10.1016/j.jdmm.2023.100850

Zhu, H. (2017). Transcultural communication and language practices. *Applied Linguistics Review, 8*(4), 543–565. https://doi.org/10.1515/applirev-2016-1046

Zikic, J., Zupic, I., & Černe, M. (2025). Digital Nomads in Conversation: Reddit-based Analysis and the Future of Nomadic versus Migrant Career Journeys. *International Migration Review*, *0*(0). https://doi.org/10.1177/01979183241306343

Appendices

For a better view of the appendices in their original formats and size visit the OSF project repository (https://doi.org/10.17605/OSF.IO/UK9NY).

Appendix 1. *Literature review entries*[59]

Author(s)	Year of publication	Name of publication	Type of publication	Journal/Book /Conference	DOI (if existing)	Main disciplinary field
Getman, N.	2021	Developing a digital nomads destination from the ground up	Master's Thesis	Iscte - Instituto Universitário de Lisboa	http://hdl.handle.net/10071/26267	Tourism Development
Lee, J.	2024	The motivations of Chinese digital nomads in moving to Chiang Mai, Thailand	Doctoral Thesis	Chulalongkorn University	https://digital.car.chula.ac.th/chulaetd/11080	China Studies / Tourism
Zhai, Y.	2024	The motivations of Chinese digital nomads in moving to Chiang Mai, Thailand	Master's Thesis	Chulalongkorn University	https://digital.car.chula.ac.th/chulaetd/11080	China Studies / Tourism
Nash, C., Jarrahi, M.H., Sutherland, W.	2021	Nomadic work and location independence: The role of space in shaping the work of digital nomads	Journal Article	Human Behavior and Emerging Technologies	https://doi.org/10.1002/hbe2.234	Sociology of Technology
Kozak, M., Cetin, G., Alrawadieh, Z.	2024	Repositioning work and leisure: Digital nomads versus tourists	Journal Article	International Journal of Tourism Research	https://doi.org/10.1002/jtr.2732	Tourism Studies
Foley, A.M., Moncada, S., Mycoo, M., Nunn, P., Tandrayen-Ragoobur, V., Evans, C.	2022	Small Island developing states in a post-pandemic world: Challenges and opportunities for climate action	Journal Article	WIREs Climate Change	https://doi.org/10.1002/wcc.769	Climate Studies / Development

[59] https://osf.io/uk9ny/files/thu95.

Patokorpi, E.	2006	Abductive reasoning and ICT-enhanced learning: Towards the epistemology of digital nomads	Book Chapter	The Information Society: Emerging Landscapes	https://doi.org/10.1007/0-387-31168-8_7	Philosophy of Technology
Chevtaeva, E.	2024	Workplace transformation: Exploring the impact of digital-nomad-inspired travel experiences	Conference Paper	Information and Communication Technologies in Tourism 2024 (ENTER 2024)	https://doi.org/10.1007/978-3-031-58839-6_25	Tourism / Information Technology
Lezcano-Calderón, Y., Palos-Sánchez, P.R.	2025	Neo-nomads and sustainability	Book Chapter	Transitioning to a Circular Economy	https://doi.org/10.1007/978-3-031-77661-8_3	Sustainability Studies
de Almeida, M.A., Correia, A., Barbosa, C.E., de Souza, J.M., Schneider, D.	2025	AI and digital nomads: Glimpsing the future human–computer interaction	Conference Paper	Computer-Human Interaction Research and Applications (CHIRA 2024)	https://doi.org/10.1007/978-3-031-83845-3_12	Human-Computer Interaction
Nash, C., Jarrahi, M.H., Sutherland, W., Phillips, G.	2018	Digital nomads beyond the buzzword: Defining digital nomadic work and use of digital technologies	Conference Paper	Transforming Digital Worlds (iConference 2018 Proceedings)	https://doi.org/10.1007/978-3-319-78105-1_25	Information Science
Thompson, B.Y.	2019	'I get my lovin' on the run': Digital nomads, constant travel, and nurturing romantic relationships	Book Chapter	The Geographies of Digital Sexuality	https://doi.org/10.1007/978-981-13-6876-9_5	Geography of Relationships
Sun, Y., Zhou, C.	2024	A study of digital nomad culture and local social practices: Based on fieldwork research in a certain area of Southwest China	Conference Paper	Big Data and Social Computing (BDSC 2024)	https://doi.org/10.1007/978-981-97-5803-6_18	Cultural Anthropology / China Studies
Wang, B., Schlagwein, D., Cecez-Kecmanovic, D.	2025	'Emancipation' in digital nomadism vs in the nation-state: A comparative	Journal Article	Journal of Business Ethics	https://doi.org/10.1007/s10551-024-05699-8	Business Ethics / Political Theory

		analysis of idealtypes				
Richter, S., Richter, A.	2020	Digital nomads	Journal Article	Business & Information Systems Engineering	https://doi.org/10.1007/s12599-019-00615-1	Information Systems
Schlagwein, D.	2018	The history of digital nomadism	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-018-0111-1	History of Technology / Tourism
Gretzel, U.	2019	#VanLife: Materiality, makeovers and mobility amongst digital nomads	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-019-00157-9	Tourism / Material Culture
Nash, C.	2019	Is mobile work really location-independent ? The role of space in the work of digital nomads	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-019-00158-8	Geography of Work
Cook, D.	2020	The freedom trap: Digital nomads and the use of disciplining practices to manage work/leisure boundaries	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-020-00172-4	Tourism / Sociology of Work
Willment, N.	2020	The travel blogger as digital nomad: (Re-)imagining workplace performances of digital nomadism within travel blogging work	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-020-00173-3	Media Studies / Tourism
Mancinelli, F.	2020	Digital nomads: Freedom, responsibility and the neoliberal order	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-020-00174-2	Political Economy / Tourism
Hannonen, O.	2020	In search of a digital nomad: Defining the phenomenon	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-020-00177-z	Tourism Studies
Hermann, I., Paris, C.M.	2020	Digital nomadism: The nexus of remote working and travel mobility	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-020-00188-w	Tourism Studies
Li, J., Zhong, S.	2025	Understanding the well-being of digital nomads from the	Journal Article	Information Technology & Tourism	https://doi.org/10.1007/s40558-025-00331-5	Psychology / Tourism

		perspective of self-determination theory: The case of China				
Thompson, B.Y.	2019	The digital nomad lifestyle: (Remote) work/leisure balance, privilege, and constructed community	Journal Article	International Journal of the Sociology of Leisure	https://doi.org/10.1007/s41978-018-00030-y	Sociology of Leisure
Bednorz, J.	2024	Working from anywhere? Work from here! Approaches to attract digital nomads	Journal Article	Annals of Tourism Research	https://doi.org/10.1016/j.annals.2023.103715	Tourism Economics
Patricia Aida Linao, P., Heimtun, B., Morgan, N.	2024	Digital nomadism, gender and racial power relations	Journal Article	Annals of Tourism Research	https://doi.org/10.1016/j.annals.2024.103770	Gender Studies / Critical Race Theory
Rainoldi, M., Ladkin, A., Buhalis, D.	2025	Digital nomads' work–leisure management practices	Journal Article	Annals of Tourism Research	https://doi.org/10.1016/j.annals.2025.103904	Tourism / Work-Life Balance
Miguel, C., Lutz, C., Majetić, F.	2024	Alone on the road: Loneliness among digital nomads and the use of social media to foster personal relationships	Journal Article	Computers in Human Behavior	https://doi.org/10.1016/j.chb.2023.107843	Psychology / Social Media
Bahri, M.T.	2024	Evidence of the digital nomad phenomenon: From "reinventing" migration theory to destination countries' readiness	Journal Article	Heliyon	https://doi.org/10.1016/j.heliyon.2024.e36655	Migration Studies
Tyutyuryukov, V., Guseva, N.	2021	From remote work to digital nomads: Tax issues and tax opportunities of digital lifestyle	Conference Paper	IFAC-PapersOnLine	https://doi.org/10.1016/j.ifacol.2021.10.443	Taxation
Chevtaeva, E., Denizci-Guillet, B.	2021	Digital nomads' lifestyles and coworkation	Journal Article	Journal of Destination Marketing & Management	https://doi.org/10.1016/j.jdmm.2021.100633	Tourism Management

Zhou, L., Buhalis, D., Fan, D.X.F., Ladkin, A., Lian, X.	2024	Attracting digital nomads: Smart destination strategies, innovation and competitiveness	Journal Article	Journal of Destination Marketing & Management	https://doi.org/10.1016/j.jdmm.2023.100850	Destination Marketing
Jinhee, L., Shin, H., Kang, J.	2024	Investigating the managerial effects of workcations (work plus vacations) on digital nomad employees: Workcation satisfaction, work engagement, innovation behavior, intention to stay, and revisit intention	Journal Article	Journal of Hospitality and Tourism Management	https://doi.org/10.1016/j.jhtm.2024.05.003	Hospitality Management
Schlagwein, D.	2020	The mobilities of digital work: The case of digital nomadism	Journal Article	The Journal of Strategic Information Systems	https://doi.org/10.1016/j.jsis.2020.101618	Information Systems
Lacárcel, F.J.S.	2025	Digital technologies, sustainable lifestyle, and tourism: How digital nomads navigate global mobility?	Journal Article	Sustainable Technology and Entrepreneurship	https://doi.org/10.1016/j.staae.2025.100096	Sustainable Technology / Tourism
Lacárcel, F.J.S., Huete, R., Zerva, K.	2024	Decoding digital nomad destination decisions through user-generated content	Journal Article	Technological Forecasting and Social Change	https://doi.org/10.1016/j.techfore.2023.123098	Technology Studies / Tourism
Tiberius, V., Chen, N., Bartels, M., von der Oelsnitz, D.	2024	Breaking out! A netnography study on motives of a digital nomad lifestyle	Journal Article	Technology in Society	https://doi.org/10.1016/j.techsoc.2024.102678	Sociology of Technology
Miguel, C., Perez-Vega, R., Lutz, C., Majetić, F., Sánchez-Razo, M.	2023	It's not all shiny and glamorous: Loneliness and fear of missing out among digital nomads	Journal Article	Tourism Management Perspectives	https://doi.org/10.1016/j.tmp.2023.101049	Tourism Psychology

Miocevic, D., Kock, F.	2025	Socio-cultural adaptation in cross-cultural transitions of digital nomads: A self-discrepancy view	Journal Article	Tourism Management	https://doi.org/10.1016/j.tourman.2025.105153	Cross-Cultural Management / Tourism
Koll, E., Wang, X., Lin, M.S.	2025	Digital nomadism as a form of travel: A phenomenological study of going on a workation	Journal Article	Tourism Management	https://doi.org/10.1016/j.tourman.2025.105185	Tourism / Phenomenology
Schlagwein, D.	2017	'Escaping the rat race': Different orders of worth in digital nomadism	Journal Article	Journal of Information Technology	https://doi.org/10.1057/s41265-017-0043-9	Information Systems
Richards, G.	2015	The new global nomads: Youth travel in a globalizing world	Journal Article	Tourism Recreation Research	https://doi.org/10.1080/02508281.2015.1075724	Youth Tourism
Miocevic, D., Kvasina, A.	2025	A parsimonious framework of digital nomads' local food avoidance in destination: The explanatory role of travel motivation and emotions	Journal Article	Tourism Recreation Research	https://doi.org/10.1080/02508281.2025.2450996	Tourism / Food Studies
Hong, R.	2023	Road warriors to digital nomads: Portable computers, habitats, and remote work	Journal Article	Cultural Studies	https://doi.org/10.1080/09502386.2021.1992462	Cultural Studies / Technology
Reichenberger, I.	2018	Digital nomads – A quest for holistic freedom in work and leisure	Journal Article	Annals of Leisure Research	https://doi.org/10.1080/11745398.2017.1358098	Leisure Studies
Yu, K., Zhang, J., Hannam, K.	2024	Welcome to 'Dalifornia': Chinese digital nomads as mobile entrepreneurs	Journal Article	Current Issues in Tourism	https://doi.org/10.1080/13683500.2024.2376884	Tourism / Entrepreneurship

Chen, P., Pun, N.	2024	'Re-coupling site' of social media and the workplace in digital nomadism: Chinese female workers' self-produced vlogs with class and gender connotations	Journal Article	Information, Communication & Society	https://doi.org/10.1080/1369118X.2024.2410254	Gender Studies / Media Studies
Dreher, N., Triandafyllidou, A.	2025	Understanding digital nomadism: A three-level framework for migration studies	Journal Article	Journal of Ethnic and Migration Studies	https://doi.org/10.1080/1369183X.2025.2517353	Migration Studies
Koskela, K., Beckers, P.	2024	Typologizing digital nomad visas: Comparing policy rationales from tourism to wealth accumulation to immigration	Journal Article	Journal of Comparative Policy Analysis: Research and Practice	https://doi.org/10.1080/13876988.2024.2343696	Comparative Public Policy
Holleran, M., Notting, M.	2023	Mobility guilt: Digital nomads and COVID-19	Journal Article	Tourism Geographies	https://doi.org/10.1080/14616688.2023.2217538	Tourism Geography
Hannonen, O.	2024	Emergent geographies of digital nomadism: Conceptual framing, insights and implications for tourism	Journal Article	Tourism Geographies	https://doi.org/10.1080/14616688.2023.2299845	Tourism Geography
Bergan, T.L., Gorman-Murray, A., Power, E.R.	2021	Coliving housing: Home cultures of precarity for the new creative class	Journal Article	Social & Cultural Geography	https://doi.org/10.1080/14649365.2020.1734230	Human Geography / Sociology
Orel, M.	2019	Coworking environments and digital nomadism: Balancing work and leisure whilst on the move	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2019.1639275	Leisure Studies
Šímová, T.	2022	A research framework for digital nomadism: A	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2022.2134200	Bibliometrics / Leisure Studies

		bibliometric study				
Pacheco, C., Azevedo, A.	2023	Mapping the journey of the coliving experience of digital nomads using verbal and visual narratives	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2022.2156593	Leisure Studies / Narrative Analysis
Matsushita, K.	2023	How the Japanese workcation embraces digital nomadic work style employees	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2022.2156594	Leisure Studies / Japanese Studies
Sánchez-Vergara, J.I., Orel, M., Capdevila, I.	2023	"Home office is the here and now." Digital nomad visa systems and remote work-focused leisure policies	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2023.2165142	Leisure Studies / Public Policy
Cook, D.	2023	What is a digital nomad? Definition and taxonomy in the era of mainstream remote work	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2023.2190608	Leisure Studies
Jaiswal, R., Gupta, S., Gupta, S.K.	2024	The impending disruption of digital nomadism: Opportunities, challenges, and research agenda	Journal Article	World Leisure Journal	https://doi.org/10.1080/16078055.2024.2346091	Leisure Studies / Futures Studies
Green, P.	2020	Disruptions of self, place and mobility: Digital nomads in Chiang Mai, Thailand	Journal Article	Mobilities	https://doi.org/10.1080/17450101.2020.1723253	Mobility Studies / Anthropology
Fiedler, C.	2022	Locked-down vanlife: Identity projects of digital nomads amid the challenges of the global COVID-19 pandemic	Journal Article	Mobilities	https://doi.org/10.1080/17450101.2022.2125137	Mobility Studies

Mancinelli, F., Germann Molz, J.	2023	Moving with and against the state: Digital nomads and frictional mobility regimes	Journal Article	Mobilities	https://doi.org/10.1080/17450101.2023.2209825	Mobility Studies / Political Science
Huang, Z.	2023	Digital nomads in China: A case study of Anji's digital nomad community	Journal Article	Journal of China Tourism Research	https://doi.org/10.1080/19388160.2023.2190012	China Studies / Tourism
Methorst, J.J., Johnston, L., Collins, F.L.	2025	Digital nomads and coworking spaces' infrastructures: Disruptions and repairs of place, bodies and labour during pandemic times	Journal Article	Applied Mobilities	https://doi.org/10.1080/23800127.2025.2466883	Mobility Studies
Ahuja, S., Nikolova, N., Clegg, S.	2020	Identities, digital nomads, and liquid modernity	Book Chapter	The Oxford Handbook of Identities in Organisations	https://doi.org/10.1093/oxfordhb/9780198827115.013.20	Business / Organisation Studies
Al-Hadi, N.A., Al-Aufi, A.S.	2019	Information context and socio-technical practice of digital nomads	Journal Article	Global Knowledge, Memory and Communication	https://doi.org/10.1108/GKMC-10-2018-0082	Information Science
Miocevic, D., Kovacevic, A.K.	2025	Gastronomic journeys: Understanding food variety-seeking behavior in the digital nomad lifestyle	Journal Article	International Journal of Contemporary Hospitality Management	https://doi.org/10.1108/IJCHM-09-2024-1439	Hospitality / Food Studies
Orel, M.	2020	Life is better in flip flops: Digital nomads and their transformational travels to Thailand	Journal Article	International Journal of Culture, Tourism and Hospitality Research	https://doi.org/10.1108/IJCTHR-12-2019-0229	Tourism / Cultural Studies
Kulshrestha, R.	2025	Living life the nomadic way: Identifying skills of a digital nomad in new forms of organisational employment	Journal Article	International Journal of Organisational Analysis	https://doi.org/10.1108/IJOA-12-2024-2071	Human Resource Management

Schlagwein, D.	2018	'Escaping the rat race': Justifications in digital nomadism	Journal Article	Information Technology & People	https://doi.org/10.1108/ITP-02-2017-0047	Information Systems / Sociology
Vohra, V., Banishetty, S., Dutta, T., Joshi, A.	2025	Work, wander and repeat: Modeling the enablers of digital nomadic lifestyle	Journal Article	Journal of Global Mobility	https://doi.org/10.1108/JGM-05-2024-0044	Mobility Studies
Narayanaswami, P.	2020	Branding the digital nomad lifestyle: A genre analysis of coworking retreat websites	Journal Article	Journal of Tourism Futures	https://doi.org/10.1108/JTF-10-2019-0120	Marketing / Tourism
Heuser, C., Reintjes, K.	2019	The identity quest of digital nomads	Journal Article	Journal of Tourism Futures	https://doi.org/10.1108/JTF-12-2018-0066	Tourism Futures / Psychology
Nichols, A.	2022	Digital nomads: A savvy enterprise's newest HR frontier	Journal Article	Strategic HR Review	https://doi.org/10.1108/SHR-08-2022-0049	Human Resource Management
Merkel, J.	2022	Coworking spaces as destinations and new stakeholders in hospitality ecosystems	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-02-2022-0007	Hospitality Management
Günay, F., Toksöz, D., Aslan, Z.	2024	The causes, process and result of becoming a digital nomad family: A phenomenological case study	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0042	Family Studies / Tourism
Yayla, Ö., Göde, M.Ö., Ekincek, S.	2024	Global palates: Unraveling digital nomads' culinary journeys and gastro-tourist profiles	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0045	Food Tourism
Arslan, E.	2024	From tourists to nomads: Unveiling the transformative potential of digital nomads in travel and tourism	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0046	Tourism Studies

Kaya Sayarı, B., Coşkun, İ.O.	2024	Deconstructing digital nomads: Are they the last frontiers of the post-tourist?	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0049	Tourism Theory
Ay, E., Güzel, M.O.	2024	Digital nomads in destination marketing	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0050	Tourism Marketing
Özoğul Balyali, T.	2024	Female digital nomads' destination experiences cape perspective	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0051	Gender Studies / Tourism
Ünal, G.	2024	The relationship between regional development and digital nomadism	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0056	Regional Development / Tourism
Arisoy, B.	2024	Digital nomadism and the tourism industry labor markets	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0062	Tourism / Hospitality Management
Günay, S., Akgiş İlhan, Ö.	2024	Conclusion: Exploring the intersection of digital nomadism and tourism in a postmodern world	Journal Article	Worldwide Hospitality and Tourism Themes	https://doi.org/10.1108/WHATT-03-2024-0066	Tourism Studies
de Almeida, M.A., Correia, A., Schneider, D., de Souza, J.M.	2021	COVID-19 as opportunity to test digital nomad lifestyle	Conference Paper	2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)	https://doi.org/10.1109/CSCWD49262.2021.9437685	Computer-Supported Cooperative Work
de Almeida, M.A., Correia, A., de Souza, J.M., Schneider, D.	2022	Digital nomads during the COVID-19 pandemic: Evidence from narratives on Reddit discussions	Conference Paper	2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)	https://doi.org/10.1109/CSCWD54268.2022.9776240	Computer-Supported Cooperative Work
de Almeida, M.A., Correia, A., de Souza, J.M., Schneider, D.	2023	Mapping tokenomics arrangements to expand the digital nomad ecosystem	Conference Paper	2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)	https://doi.org/10.1109/CSCWD57460.2023.10152859	Computer-Supported Cooperative Work

Dalisaymo, L.M., Patalay, M.M.	2024	Artificial intelligence impact on digital nomad work–life balance: The role of automation	Conference Paper	2024 2nd International Conference on Computing and Data Analytics (ICCDA)	https://doi.org/10.1109/ICCDA64887.2024.10867307	Information Systems / AI
Shawkat, S., Abd Rozan, M.Z., Bt Salim, N., Shehzad, H.M.F.	2021	Digital nomads: A systematic literature review	Conference Paper	2021 7th International Conference on Research and Innovation in Information Systems (ICRIIS)	https://doi.org/10.1109/ICRIIS53035.2021.9617008	Information Systems (Literature Review)
Makimoto, T.	2013	The age of the digital nomad: Impact of CMOS innovation	Journal Article	IEEE Solid-State Circuits Magazine	https://doi.org/10.1109/MSSC.2012.2231498	Electrical Engineering / Semiconductors
de Almeida, M.A., de Souza, J.M., Correia, A., Schneider, D.	2022	The role of wannabes in the digital nomad ecosystem in times of pandemic	Conference Paper	2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)	https://doi.org/10.1109/SMC53654.2022.9945599	Information Systems
Sequera, J.	2024	Struggling with the digital nomad: Transnational teleworkers as the new 'creative class' in the urban marketplace?	Journal Article	International Journal of Urban and Regional Research	https://doi.org/10.1111/1468-2427.13293	Urban Studies
Webb, A.K.	2024	Digital nomadism and the challenge to social citizenship	Journal Article	Global Policy	https://doi.org/10.1111/1758-5899.13377	Political Science / Social Policy
Sciuva, E.	2025	Geographies of digital nomadism: A research agenda	Journal Article	Geographical Journal	https://doi.org/10.1111/gec3.70016	Geography
Wang, B., Cecez-Kecmanovic, D., Schlagwein, D., Cahalane, M.	2019	Digital nomadism and the market economy: Resistance and compliance	Journal Article	Information Systems Journal	https://doi.org/10.1111/isj.12251	Information Systems
Aroles, J., Granter, E., Vaujany, F.	2020	'Becoming mainstream': The professionalisation and corporatisation of digital nomadism	Journal Article	New Technology, Work and Employment	https://doi.org/10.1111/ntwe.12158	Sociology of Work / Management

Sutherland, W., Jarrahi, M.H.	2017	The gig economy and information infrastructure: The case of the digital nomad community	Journal Article	Proceedings of the ACM on Human-Computer Interaction	https://doi.org/10.1145/3134732	Human-Computer Interaction
Al-Zobaidi, S.	2009	Digital nomads: Between homepages and homelands	Journal Article	Middle East Journal of Culture and Communication	https://doi.org/10.1163/187398509X12476683126707	Media Studies / Cultural Studies
Prengler, M.K., Klotz, A.C., Murphy, C.	2025	Place iteration and integration: How digital nomads navigate the mobile worker paradox	Journal Article	Administrative Science Quarterly	https://doi.org/10.1177/00018392241302795	Organisational Studies
Khawas, A., Kumar, P., Kumar, S.	2025	Gentrification and the rise of digital nomads: Economic, social, and cultural impacts on Delhi's urban landscape	Journal Article	Urban Studies	https://doi.org/10.1177/00420980231207999	Urban Studies
Miocevic, D.	2025	Baby come back: Resident–digital nomad conflicts, destination identification, and revisit intention	Journal Article	Journal of Travel Research	https://doi.org/10.1177/00472875231220945	Tourism Management
Xiao, Y., Lutz, C.	2025	Wayfarers in cyberspace: A temporal investigation of digital nomads based on liquid modernity theory	Journal Article	Journal of Travel Research	https://doi.org/10.1177/00472875231224242	Tourism / Social Theory
Zikic, J., Zupic, I., Černe, M.	2025	Digital nomads in conversation: Reddit-based analysis and the future of nomadic versus migrant career journeys	Journal Article	Work, Employment and Society	https://doi.org/10.1177/01979183241306343	Career Studies / Sociology
Hari, K.C., Triandafyllidou, A.	2025	Digital nomadism and the emergence of digital	Journal Article	International Migration Review	https://doi.org/10.1177/01979183241306367	Migration Policy

		nomad visas: What policy objectives do states aim to achieve?				
Willment, N.	2025	Travel bloggers as 'digital nomads': How can understanding this lifestyle migration help us to think about the future of work, migration and technology?	Journal Article	International Migration Review	https://doi.org/10.1177/01979183251319	Migration Studies
Cook, D.	2022	Breaking the contract: Digital nomads and the state	Journal Article	Critique of Anthropology	https://doi.org/10.1177/0308275X221120172	Anthropology
Aroles, J., Bonneau, C., Bhankaraully, S.	2022	Conceptualising 'meta-work' in the context of continuous, global mobility: The case of digital nomadism	Journal Article	Work, Employment and Society	https://doi.org/10.1177/09500170211069797	Sociology of Work
Bonneau, C., Aroles, J., Estagnasié, C.	2023	Romanticisation and monetisation of the digital nomad lifestyle: The role of online narratives in shaping professional identity work	Journal Article	Organisation	https://doi.org/10.1177/13505084221131638	Organisation Studies
Miguel, C., Lutz, C., Perez-Vega, R.	2024	Working from paradise? An analysis of the representation of digital nomads' values and lifestyle on Instagram	Journal Article	New Media & Society	https://doi.org/10.1177/14614448231205892	Media Studies
Rasnača, Z.	2023	Enforcing the rights of remote workers: The case of digital nomads	Journal Article	European Labour Law Journal	https://doi.org/10.1177/20319525231177638	Labor Law
Polson, E.	2024	A global sense of workplace?	Journal Article	Cultural Studies ↔ Critical	https://doi.org/10.1177/20501579251352991	Cultural Studies

		Rethinking the (dis)locality of digital nomads		Methodologies		
Bozzi, N.	2020	#digitalnomads, #solotravellers, #remoteworkers: A cultural critique of the traveling entrepreneur on Instagram	Journal Article	Social Media + Society	https://doi.org/10.1177/20563051209266 44	Media Studies / Cultural Studies
Ehn, K., Jorge, A., Marques-Pita, M.	2022	Digital nomads and the COVID-19 pandemic: Narratives about relocation in a time of lockdowns and reduced mobility	Journal Article	Social Media + Society	https://doi.org/10.1177/20563051221132894	Media Studies / Sociology
Bozzi, A.	2024	Digital nomadism from the perspective of places and mobilities: A literature review	Journal Article	European Transport Research Review	https://doi.org/10.1186/s12544-024-00663-z	Transport Studies / Geography
Hensellek, S., Puchala, N.	2021	The emergence of the digital nomad: A review and analysis of the opportunities and risks of digital nomadism	Journal Article	Zeitschrift für Tourismuswissenschaft	https://doi.org/10.1515/tw-2021-0013	Tourism Studies
Gurvičius, T.	2021	The connectedness of digital nomads with the host environment: Medellín, Colombia	Journal Article	Regional Formation and Development Studies	https://doi.org/10.15181/rfds.v34i2.2159	Human Geography
Von Burg, A.B.	2016	CITIZENSHIP DESERTS: REFUGEES AS DIGITAL NOMADS	Working Paper	N.A.	https://doi.org/10.17223/9785946216104/55	Political Science / Law
Bruni, R., Mladenović, D.	2016	Defining the “figure of merit” for places in the age of digital nomadism	Conference Paper	N.A.	https://doi.org/10.17223/9785946216104/8	Urban Planning
McElroy, E.	2019	Digital nomads and	Journal Article	Imaginations: Journal of	https://doi.org/10.17742/IMAGE.CR.10.1.8	Critical Race Theory /

		settler desires: Racial fantasies of Silicon Valley imperialism		Cross-Cultural Image Studies		Postcolonial Studies
Adepoju, O.D., Agunbiade, T.O.	2018	Digital nomads and the changing work life of technology entrepreneurs in Nigeria	Journal Article	International Journal of Digital Society	https://doi.org/10.20533/ijds.2040.2570.2018.0173	Information Technology / Sociology
Imam, L.S., Wibowo, G.D.H., Cahyowati, C.	2024	Arrangements for granting visas and immigration stay permits for digital nomads in Indonesia in view of legal certainty	Journal Article	Indonesian Journal of Law and Economics Review	https://doi.org/10.21070/ijler.v19i1.1001	Law / Economics
Lagidze, N., Pirtskhalava, E.	2024	The ties that bind: Exploring place identity and attachment among digital nomads	Journal Article	Migration Studies Review	https://doi.org/10.21203/rs.3.rs-4004495/v1	Migration Studies / Human Geography
Pirtskhalava, E.	2024	The ties that bind: Exploring place identity and attachment among digital nomads	Preprint	Research Square	https://doi.org/10.21203/RS.3.RS-4004495/v1	Human Geography / Psychology
Toivanen, M.	2023	Countercultural lifestyle no more? Digital nomadism and the commodification of neo-nomadic mobilities	Journal Article	Mobility Humanities	https://doi.org/10.23090/MH.2023.07.2.2.070	Mobility Studies / Cultural Studies
Arbutina, H., Šinković, Z., Pribisalić, L.	2023	The legal status of digital nomads	Conference Paper	2023 46th MIPRO ICT and Electronics Convention (MIPRO)	https://doi.org/10.23919/MIPRO57284.2023.10159694	Law / Information Technology
Elcokany, N.M., Jaramillo, J., Llaguno, M.B.B., Seweid, M.M., Mousa, O.	2022	Student nurses as modern digital nomads: Developing nursing skills competence and confidence through	Journal Article	International Journal of Innovative Research in Medical Science	https://doi.org/10.23958/ijirms/vol07-i11/1536	Nursing Education

		video-based learning				
Li, X.	2024	Going where the wind blows: A study on media practices and identity formation of digital nomads	Journal Article	Media and Communication Research	https://doi.org/10.23977/mediacr.2024.050327	Media Studies
Frick, N., Marx, J.	2021	Integrating digital nomads in corporate structures: Managerial contemplations	Conference Paper	Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021)	https://doi.org/10.24251/HICSS.2021.260	Management / Information Systems
Stumpf, T.S., Califf, C., Lancaster, J.	2022	Digital nomad entrepreneurship and lifestyle design: A process theory	Conference Paper	Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS 2022)	https://doi.org/10.24251/HICSS.2022.634	Entrepreneurship
Gibadullin, O., Melnychenko, O., Zolotarev, V.	2024	Substantiation of the motivation of global digital nomads for their engagement in the post-war industrial recovery of Ukraine	Journal Article	State Formation	https://doi.org/10.26565/1992-2337-2024-1-01	Economics / Political Science
Stumpf, T.S.	2022	Becoming a digital nomad entrepreneur: A process theory	Journal Article	E-Service Journal	https://doi.org/10.2979/eservicej.13.3.02	Entrepreneurship
Gede, I.G.K., Marhaeni, K.E., Putrana, I.W., Sanjiwani, I.G.A.M.	2021	A model of community-based development in digital nomad tourism in Intaran Village Bali	Conference Paper	Proceedings of the International Conference on Applied Science and Technology on Social Science (ICAST-SS 2020)	https://doi.org/10.2991/assehr.k.210424.073	Community Development / Tourism
Shofi'unnafi, S., Kusuma, B.M.A.	2023	Digital nomads as a strategy to accelerate the religious tourism recovery in post-pandemic	Journal Article	International Journal of Islamic Business Ethics	https://doi.org/10.30659/ijibe.8.1.26-39	Religious Tourism
Sukma Winarya Prabawa, I.W., Ratih Pertiwi, P.	2020	The digital nomad tourist motivation in Bali: Exploratory	Journal Article	Athens Journal of Tourism	https://doi.org/10.30958/ajt.7-3-3	Tourism Motivation

		research based on push and pull theory				
Bishop, N.	2024	Making sense of digital nomadism, workplace surveillance and resistance	Working Paper	SocArXiv	https://doi.org/10.31235/osf.io/jz54n	Sociology of Work
Gorliński-Kucik, P.T., Jędrzejko, P.	2021	On liberatory strategies of digital nomads	Journal Article	Er(r)go: Teoria – Literatura – Kultura	https://doi.org/10.31261/ergo.9155	Cultural Studies / Philosophy
Christiansen, R.A.	2023	Developing digital nomads as a new market segment and their role in the local digital ecosystem	Journal Article	Journal of System and Management Sciences	https://doi.org/10.33168/JSMS.2023.0201	Management / Information Systems
Mourato, I., Dias, Á., Pereira, L.	2023	Estimating the impact of digital nomads' sustainable responsibility on entrepreneurial self-efficacy	Journal Article	Social Sciences	https://doi.org/10.3390/socsci12020097	Entrepreneurship / Sustainability
Parreño-Castellano, J., Domínguez-Mujica, J., Moreno-Medina, C.	2022	Reflections on digital nomadism in Spain during the COVID-19 pandemic: Effect of policy and place	Journal Article	Sustainability	https://doi.org/10.3390/su142316253	Geography / Public Policy
Ji, Y., Kim, S.-M., Kim, Y.	2024	A way to attract digital nomads to tourist destinations in the new normal era	Journal Article	Sustainability	https://doi.org/10.3390/su16062336	Sustainable Tourism
Başaran, A.	2025	Digital nomads, the new frontier of work in the digital age: A bibliometric analysis	Journal Article	Sustainability	https://doi.org/10.3390/su17051906	Information Science / Sociology
Myskiv, G.	2025	Digital nomads as a contemporary trend in the global labor market	Journal Article	Economics, Finance and Management Review	https://doi.org/10.36690/2674-5208-2025-2-94-108	Labor Economics
Bozzi, A.	2024	Digital nomadism	Journal Article	Netcom	https://doi.org/10.4000/11su8	Urban Studies

		and cities: Exploring the production of mobility infrastructures in Barcelona				
Chavarria, A.D.	2024	Sustainable tourists, generous consumers, lifestyle inspirers: The attitudes and expectations of Zadar's local authorities vis-à-vis digital nomads	Journal Article	Netcom	https://doi.org/10.4000/vy1m	Urban Policy / Tourism
Albayrak, E.S.	2023	Changing society order with digitalisation: Global village, global workers—Digital nomads	Book Chapter	Handbook of Research on Perspectives on Society and Technology Addiction	https://doi.org/10.4018/978-1-6684-8397-8.ch006	Sociology / Technology Studies
Thandayuthapani, S., Thirumoorthi, P.	2026	The rise of digital nomad visas: Transforming global work, travel, and local economies	Book Chapter	AI and Digital Nomads Shaping Global Industrial Technology Transitions	https://doi.org/10.4018/979-8-3373-0159-4.ch005	Economics / Policy
Mariotti, I., Bednář, P., Di Marino, M.	2022	The COVID-19 pandemic and the future of working spaces	Book	Routledge	https://doi.org/10.4324/9781003181163	Urban Studies / Work Studies
Atanasova, A., Bardhi, F., Eckhardt, G.M., Mimoun, L.	2022	Digital nomadism as temporal privilege	Book Chapter	The Routledge Handbook of Digital Consumption	https://doi.org/10.4324/9781003317524-3	Consumer Research / Sociology
Betre, H.K.	2022	Reconceptualising the potential of digital nomads to transform community-based and sustainable tourism	Book Chapter	Handbook of Tourism and Consumer Behavior	https://doi.org/10.4337/9781035309801.00009	Sustainable Tourism
Mariati, S., Gilitasha, A., Rahmanita, M., Djati, S.P., Adawiyah, R.	2023	Analysis of sustainable tourism destination development for digital nomads	Journal Article	Technium Social Sciences Journal	https://doi.org/10.47577/tssj.v43i1.8773	Sustainable Tourism

		(Comparative study: Lisbon-Portugal and Canggu, Bali-Indonesia)				
Ersöz, B.	2024	Removing the boundaries of flexibility: A critical evaluation on digital nomadism	Journal Article	The Journal of social sciences Research	https://doi.org/10.48145/gopsbad.1540618	Sociology
Mohlin, A., Chatzipanagiotou, N.	2023	Digital nomads' experiences on the support of digital technologies in relation to social isolation	Conference Paper	Proceedings of the 21st European Conference on Computer-Supported Cooperative Work	https://doi.org/10.48340/ecscw2023_ep01	Computer-Supported Cooperative Work
Wang, B., Schlagwein, D., Cecez-Kecmanovic, D., Cahalane, M.C.	2018	Digital work and high-tech wanderers: Three theoretical framings and a research agenda for digital nomadism	Conference Paper	Proceedings of the Australasian Conference on Information Systems (ACIS 2018)	https://doi.org/10.5130/acis2018.bl	Information Systems
Hemsley, J., Erickson, I., Jarrahi, M.H., Karami, A.	2020	Digital nomads, coworking, and other expressions of mobile work on Twitter	Journal Article	First Monday	https://doi.org/10.5210/fm.v25i3.10371	Information Science / Sociology
Damayanti, C., Nugroho, A.Y., Rohmadiena, Q.	2024	Digital nomad visa as a development instrument for Indonesian tourism diplomacy	Journal Article	Global Insight Journal	https://doi.org/10.52447/gij.v9i2.7762	International Relations / Tourism
Kim, N.	2024	Digital nomadism: Definition, supporting factors, motivations and challenges	Working Paper	Zenodo	https://doi.org/10.5281/zenodo.11265155	Sociology (Literature Review)
Peña, R.V., Pérez, J.A.	2025	Digital nomadism as a structural phenomenon in the global post-pandemic era	Journal Article	Journal of Management and Human Resources	https://doi.org/10.5281/zenodo.15501605	Management

Thomaz Pignatari, L.	2023	The taxation of 'digital nomads' and the '3 W's': Between tax challenges and heavenly beaches	Journal Article	Intertax	https://doi.org/10.54648/TAXI2023033	Law / Taxation
Ouro, M.	2023	There must be fifty ways to tax a digital nomad: Does taxation constrain the geographical freedom of the digital nomad?	Journal Article	Intertax	https://doi.org/10.54648/TAXI2023125	Law / Taxation
Özyürek, H., Babacoglu, S., Polat, M., Turen, U.	2024	Mapping the field of digital nomadism: A bibliometric analysis using VOSviewer and R	Conference Paper	The Eurasia Proceedings of Science, Technology, Engineering and Mathematics	https://doi.org/10.55549/epstem.1518797	Bibliometrics / Science Studies
Qosim, N., Rukmana, A.Y., Usup, U.	2025	The impact of digital nomad economy on local businesses: Study of MSME market changes in popular destinations in Indonesia	Journal Article	The Eastasouth Journal of Social Science and Humanities	https://doi.org/10.58812/esssh.v2i02.474	Economics / Development Studies
Angiello, G.	2022	European cities embracing digital nomads	Journal Article	TeMA – Journal of Land Use, Mobility and Environment	https://doi.org/10.6093/1970-9870/9033	Urban Studies / Geography
Sun, K.-K., Deng, S.-H.	2025	Spatiotemporal arbitrage: How digital nomads leverage global mobility for financial optimisation	Journal Article	Economics and Management Innovation	https://doi.org/10.71222/72pw7j07	Economics / Finance
Haking, J.	2022	Digital nomad lifestyle: A field study in Bali	Master's Thesis	University of Gothenburg	https://gupea.ub.gu.se/handle/2077/71071	Sociology / Anthropology
Methorst, J.J.	2024	Location-independence? Exploring the digital geographies	Doctoral Thesis	The University of Waikato	https://hdl.handle.net/10289/16501	Human Geography

		of work, mobilities and place				
Hill, T.	2023	User research interviews with digital nomads on what they want from an internet country	Working Paper	Internet Country Project	https://internet.country/research/digitalnomads.pdf	Technology / Sociology
Appell, C.R.	2021	A digital nomad strategy for Prince Edward Island	Policy Report	University of Prince Edward Island / Community Foundation of PEI	https://www.cfpei.ca/wp-content/uploads/2022/06/Final-Digital-Nomad-Strategy.pdf	Public Policy / Tourism
Ahlberg, E.	2021	The development of the digital nomad during the course of the pandemic	Master's Thesis	Malmö University	https://www.diva-portal.org/smash/get/diva2:1596331/FULLTEXT02	Sociology / Media Studies
Chova, J.B.	2023	Nomadism without borders: Exploring connections in digital nomad destinations	Master's Thesis	Jönköping University	https://www.diva-portal.org/smash/get/diva2:1758057/FULLTEXT01.pdf	Tourism / Sociology
Casi, E., Mardan, M., Stage, B.M.	2024	Citizenship/residence by investment and digital nomad visas: The golden era of individual tax evasion and avoidance?	Book Chapter	Research Handbook on the Economics of Tax Havens	https://www.elgaronline.com/abstract/book/9781803929743/chapter10.xml	Economics / Law
Ali, S.Z.	2025	Digital nomadism and environmental sustainability: A paradox of mobility	Journal Article	Journal of Social Science and Management Studies	https://www.imdpi.com/index.php/jssms/article/view/40	Environmental Studies / Management
Aufschnaiter, C., Schwarz, S., Hemetsberger, A.	2021	Anchors on the move: Digital nomads' solid footholds in liquidity	Conference Paper	Advances in Consumer Research	https://www.researchgate.net/publication/354871628_Anchors_on_the_Move_Digital_Nomads'_Solid_Footholds_in_Liquidity	Consumer Research / Marketing
Dreher, N., Triandafyllidou, A.	2023	Digital nomads: Toward a future research agenda	Journal Article	Journal of Ethnic and Migration Studies	https://www.torontomu.ca/content/dam/centre-for-immigration-and-settlement/tmcis/publications/workingpapers/2023-04-WP-Dreher-Triandafyllidou.pdf	Migration Studies
Abbas, Y.	2006	Neo-nomads: Designing environment	Doctoral Thesis	Harvard University, Graduate	N.A.	Architecture / Design

		s for living in the age of mental, physical and digital mobilities		School of Design		
Barroso, C.F.	2020	Backpacking culture: Digital nomads in Porto—A Selina case study	Master's Thesis	University of Porto	N.A.	Tourism / Sociology
Beaumont, J.	2019	Digital nomads and sense of place: A case study of Lisbon	Master's Thesis	Wageningen University	N.A.	Human Geography
Brusseau, J.	2021	What does it mean to be a digital nomad?	Journal Article	Philosophy Now	N.A.	Philosophy
Demaj, E., Hasimja, A., Rahimi, A.	2021	Digital nomadism as a new flexible working approach: Making Tirana the next European hotspot for digital nomads	Journal Article	Albanian Journal of Business and Economics	N.A.	Economics / Urban Development
Farmer, B.P.	2024	The economic impacts of digital nomads in Medellín, Colombia	Bachelor's Thesis	University of New Hampshire, Durham	N.A.	Economics
Fernandes Barroso, C.	2020	Backpacking culture: Digital nomads in Porto—A Selina case study	Master's Thesis	University of Porto	N.A.	Tourism / Sociology
Filiposki, O., Metodijeski, D., Cuculevski, N.	2024	Digital nomads as contemporary form of business tourism	Journal Article	International Journal of Economics, Management and Tourism	N.A.	Business Tourism
Gourlay, F., Mahéo, C., Marinos, C.	2024	Coworkation and digital nomadism: Normalisation and/or differentiation of working and living practices?	Conference Paper	35th International Geographical Congress	N.A.	Geography / Work Studies
Huang, T.	2021	Technology-triggered social affordances: A netnography study of	Journal Article	Journal of Information Technology Case and Application Research	N.A.	Information Systems

		digital nomads (Part 2)				
Ihrig, K.	2024	How to be a digital nomad: Build a successful career while travelling the world	Book	Kogan Page	N.A.	Business / Guidebook
Kalansuriya, D.I.	2024	Sri Lanka's transition to a digital nomad hotspot: A netnographic analysis	Conference Paper	Proceedings of the International Conference on Real Estate Management and Valuation (ICREMV 2024)	N.A.	Real Estate / Tourism
Kocherbaeva, A.	2024	Return migration from Russia to Kyrgyzstan: Digital nomads in Kyrgyzstan	Journal Article	Bulletin of the Kazakh National University. International Relations and International Law Journal	N.A.	Migration Studies / Central Asian Studies
Lipniewicz, R.	2024	Digital nomads and international taxation of income from employment	Journal Article	Krytyka Prawa	N.A.	Law / Taxation
Lonardi, S., Buffa, F., Martini, U.	2024	Digital nomadism and tourism development: Stakeholders' perceptions of an inner area	Journal Article	Sinergie – Italian Journal of Management	N.A.	Management / Tourism
Marugg, M.	2022	Combining coliving and coworking in Amsterdam	Master's Thesis	University of Amsterdam	N.A.	Urban Studies / Real Estate
Mouratidis, G.	2018	Travel, remote work and alternative lifestyles	Master's Thesis	Lund University	N.A.	Sociology
Nair, M.K., Gokhale, D.S.	2024	The rise of digital nomadism and its impact on X and Z generation	Journal Article	Journal of Advance and Future Research	N.A.	Generational Studies / Sociology
Nanopulos, K., Demiroglou, N., Iliopoulou, V.	2021	The rise of 'digital nomads' in Greece	Journal Article	Tourismos: An International Multidisciplinary Journal of Tourism	N.A.	Tourism Studies
Narotzki, D., Kuperberg, V.	2023	Taxation of digital nomads	Journal Article	Tax Notes Federal	N.A.	Law / Taxation
Nithitaknakin, N., Khampuong	2023	Digital learning space	Journal Article	International Journal of Instruction	N.A.	Education Technology

, P., Chairungruang, S.		management for digital nomads				
Novella, T.	2025	Translanguaging in the era of digital nomadism: A sociolinguistic perspective on voluntary mobility in Europe	Journal Article	Journal of Sociolinguistics	N.A.	Sociolinguistics
Noviarini, N., Samputra, P.L.	2024	Digital nomad and analysis of regional economic resilience of tourism sector in Bali province after COVID-19 pandemic	Journal Article	Jurnal Indonesia Sosial Teknologi	N.A.	Regional Economics / Tourism
O'Brien, M.	2014	Finding a home for the 'digital nomad'	Journal Article	Journal of Mobile Media	N.A.	Media Studies
Pramudya, A.W.E., Prasetyo, H.B.D., Wempi, J.A.	2024	Digital nomad work trends in Indonesia after the pandemic	Journal Article	International Journal of Science and Society	N.A.	Sociology of Work
Prayana, I.K.R., Darma, G.S.	2023	Dynamic governance as an effort to develop quality tourism: Heaven for digital nomads	Journal Article	Tec Empresarial	N.A.	Governance / Tourism
Prester, J.	2022	Belonging on the move: How digital nomads synchronise socio-temporal rhythms	Journal Article	Ephemera: Theory & Politics in Organisation	N.A.	Organisation Studies
Prester, J., Cecez-Kecmanovic, D., Schlagwein, D., Cahalane, M.	2021	Legitimation as the correspondence of practice: An ethnographic study of digital nomad work practices	Conference Paper	Proceedings of the 42nd International Conference on Information Systems (ICIS 2021)	N.A.	Information Systems
Puusaar, S.	2024	Precariously beyond border: The life of a digital nomad. An ethnographi	Master's Thesis	University of Lisbon	N.A.	Anthropology

		c exploration of the tensions in self-making and being made in the remote working community in Lisbon, Portugal				
Rainoldi, M.	2024	Managing work and leisure in the digital age: A practice exploration of digital nomadism	Doctoral Thesis	Bournemouth University	N.A.	Leisure Studies / Sociology
Rakhmadi, R.	2022	Role of digital nomad in supporting tourism in Indonesia: Case study Bali	Journal Article	Journal of Tourism and Hospitality Management	N.A.	Tourism Management
Sapuán, M.A., Razali, M.A.H.M., Ariffin, H.F., Harun, N.H.	2024	Nomad Nirvana	Journal Article	Gading Journal for the Social Sciences	N.A.	Social Sciences
Smee, G.	2024	A 'digital nomad' visa is coming to South Africa	Journal Article	Personal Finance Magazine	N.A.	Finance / Policy
Smercina, V.	2019	Living on the move: The digital nomad mobile phenomenon, identity and practice	Master's Thesis	University of Kansas	N.A.	Sociology / Anthropology
Sousa, B.V. de, Bastida, M., Vieira, E.	2025	Unraveling digital nomadism: A comprehensive bibliometric review of emerging research and trends	Journal Article	Journal of Global Mobility	N.A.	Bibliometrics / Mobility Studies
Stickel, M.D.	2020	Challenges and opportunities of digital nomadism and its implications for tomorrow's workforce	Journal Article	Journal of Management Research	N.A.	Management
Susanti, P.H.	2023	Readiness of Canggu Village, Bali, as digital nomad tourism	Journal Article	Journal of Tourism Economics, Policy and Hospitality Management	N.A.	Tourism Economics

Syrbe, S.	2022	Digital nomadism as a new form of remote work for companies	Journal Article	Journal of Business and Economics	N.A.	Business
Taltavull, P.	2024	Digital nomads and real estate markets	Journal Article	Journal of European Real Estate Research	N.A.	Real Estate
Thornell, C.	2022	Hotels reconfigure to meet growing needs of the digital nomad	Trade Article	Hotel News Now	N.A.	Hospitality Management
Tousarvandani, K.	2025	Destination decision-making: A study of digital nomads' preferences and motivations	Master's Thesis	University of Padova	N.A.	Tourism Studies
Tsvetcoff, R., Patrício, R., Alçada, S.	2024	Attracting digital nomads to rural areas: An attempt to bridge	Book Chapter	Sustainable Development of Territories in Contexts of Uncertainty: External Shocks and Risks	N.A.	Rural Development
Viana de Sousa, B., Bastida, M., Vieira, E.	2025	Unraveling digital nomadism: A comprehensive bibliometric review of emerging research and trends	Journal Article	Journal of Global Mobility	N.A.	Bibliometrics / Mobility Studies
Wang, H.	2020	A city of placelessness	Book	Routledge	N.A.	Urban Studies
Wibowo, T.O.	2024	Dynamic of migration lifestyle and settlement of nomads in Bali, Indonesia	Journal Article	The Indonesian Journal of Geography	N.A.	Geography / Migration Studies
Woldoff, R.A., Litchfield, R.C.	2021	Paradise paradox: Constructing a digital nomad community	Book	Oxford University Press	N.A.	Sociology
Yılmaz, O., Gürbüz, P.G.	2025	Digital nomads' leisure time and examination of recreation perceptions	Journal Article	Journal of Sport for All and Recreation	N.A.	Leisure Studies
Zaki, H.S.	2024	Digital nomad tourists at	Journal Article	Journal of the Faculty of Tourism and	N.A.	Tourism Studies

		the Egyptian destination: Motivations and challenges		Hotels – University of Sadat City		
Zeinstra, S.	2023	Digital nomads and perceived professionalism	Master's Thesis	University of Amsterdam	N.A.	Professional Studies
Zhou, Y., Li, L., Shi, Y., Luo, J., Yu, B.	2024	Adaptive behavior in digital nomads: An interactive spatial structure study using machine learning	Conference Paper	iBAGUI CBCN 2024 Conference Proceedings	N.A.	Computer Science / Human Geography

Appendix 2. ***FLO Data Management Plan***[60]

Data Management Plan [1]

Project title: FLO – The European Fluid Language Observatory

Creator: Novella Tedesco

Affiliation: University of Bologna – Department of Interpreting and Translation

Project abstract:

In the era of digital transformation, unprecedented migratory flows and mobility, super-diversity has become the essence of today's society. As fundamental elements in cultural and social processes, languages are our primary means of communication with the world, they are at the basis of community building: languages represent who we are. And, in the ever-changing *liquid* reality which we all are living our languages assume a fluid shape, they change along with the people we meet, the place where we live, the books we read, the films we watch, the networks we are part of, and the experiences we make (Bauman 2000; Blommaert 2010; Garcia & Wei 2014; Canagarajah 2017; Wang et al. 2023).

FLO[2], The European Fluid Languages Observatory, is a Ph.D. project which aims at bringing together research, documentation, practices about cultural and linguistic fluidity. It was conceived as both a conceptual framework and an infrastructural project aimed at documenting and analysing the linguistic practices of digital nomad communities in Europe. Its mission is twofold: (i) to provide a systematic methodology for collecting, curating, and sharing language resources including but not limited to natural language data, in accordance with FAIR and Open Science principles (Wilkinson et al. 2016; European Commission 2018), and (ii) to enhance knowledge co-creation and sharing about transcultural communication beyond the boundaries of academia.

Digital nomads are people who work remotely, often in the digital labour market, while adopting a travelling lifestyle. At present, communities of digital nomads, based on digital networking and communal life habits and practices, are rising around the world. In Europe, the biggest communities of digital nomads can be found in the Canary Islands (Spain), in Madeira and the Algarve Peninsula (Portugal) and in Bankso, a Bulgarian village, but many more communities are getting larger and larger. From a sociolinguistic perspective, FLO's purpose is thus to collect different types of evidence regarding the way languages are used within digital nomad communities around Europe. Indeed, these communities have been regarded as a good example of transcultural superdiverse contexts of communication (Reichenberger 2018; Woldoff &Lichtfield 2021).

The data collection protocol outlined in this document takes inspiration from theoretical and methodological reflections in contemporary sociolinguistics (see, for example, Blommaert 2014, or Tusting2020).

[60] https://osf.io/uk9ny/files/p79j6.

Start date: 01/11/2022 **End date:** 31/10/2025

Copyright information:

This work is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

Data description, collection and re-use

What data will be collected: Ethnographic data; questionnaire data; spoken language data.

How new data will be collected: The ethnographic data will be collected through on field and online observation, with a participatory community-based approach (Beckett et al., 2022)[3]; the ethnographic data will also include recorded interviews, photographs of events and places, screenshots from social network posts and communications; the questionnaire data will be collected online using LimeSurvey, participants are selected on a voluntary basis; the spoken language data will be collected through audio-recording of natural-occurring conversations in informal settings (in particular, at social events in the communities of digital nomads based in Canary Islands and Madeira Islands). The latter typology of data will then be transcribed and transcriptions will be treated as a separate dataset.

Documentation and data quality

- ***Questionnaire responses***

Documentation accompanying data: Questionnaire's structure and questions, readme.txt

Metadata accompanying data: Authors and Contributors in the study, Corresponding Author's Email, Titleand Description of the dataset, key-words, Bibliography, Academic field, University Department, Date, Type of License, time and date of response submission.

Data quality control: All responses are collected upon agreement with the Privacy Policy and Legal documentation, included in the online questionnaire (Appendix A.0). The settings on LimeSurvey, the software used to develop the questionnaire and collect responses, ensure that every response is related to a unique participant. For privacy and data preservation reasons LimeSurvey was chosen as the software to develop and administrate the questionnaire because it can be installed on the University servers.Participants, despite being chosen on a voluntary basis, are considered to be representative of the group under study through a set of preliminary questions. Moreover, date and time sampling has been performed in some co-workings and digital nomad meet-up events in Madeira and Canary Islands to randomize the sample of participants.
Empty responses are cancelled from the dataset, and incomplete responses are signaled.

From a technical perspective, data FAIRness is ensured by applying data sharing and formats standards of the scientific community in the field of corpus and socio-linguistics.

- ***Audio-recorded interviews***

Documentation accompanying data: Legal documentation (Appendix A.1), transcript summaries (which will be used to write blog articles).

Metadata accompanying data: Date, location, Author, Personal information of the interviewed.

Data quality control: Data quality is ensured by the use of a highly performing software for audio recording (MOTIV Audio) and external noise reduction will be applied if needed. These data will not be deposited on any online repository but will be transcribed and pieces of transcripts will be used for analysis or published according to research needs.

- ***Pictures from ethnographic research***

Documentation accompanying data: Legal documentation (Appendix A.2), of the people involved – in case there are participants who are recognizable.

Metadata accompanying data: Date, location (also on the web), contextual notes.

Data quality control: These data will not be deposited on any online repository but will be used in presentations and publications if relevant. In these cases, the highest possible quality of images will be used.

- ***Spoken language data***

Documentation accompanying data: Legal documentation (Appendix A.3).

Metadata accompanying data: Authors and Contributors in the study, Corresponding Author's Email, Titleand Description of the dataset, key-words, Bibliography, Academic field, University Department, Date, Type of License, Contextual metadata (Appendix B).

Data quality control: Data quality is ensured by the use of a highly performing software for audio recording (MOTIV Audio) and external noise reduction will be applied if needed through the software WavePad Audio Editor. Microphones are placed in the best possible position to get the best audio quality and locations with little noise are chosen as settings for the data collection. The researcher checks that microphones maintainthe most convenient position throughout the data collection procedure.

- ***Transcriptions of spoken language data***

Documentation accompanying data: Legal documentation (Appendix A.3), linguistic annotation schema, readme.txt.

Metadata accompanying data: Authors and Contributors in the study, Corresponding Author's Email, Titleand Description of the dataset, key-words, Bibliography, Academic field, University Department, Contextual metadata, Date, Type of License.

Data quality control: After automatic transcription which will be done employing the best-performing software for speech-to-text transcription based on LLMs, transcripts will be revised also through the use ofthe software ELAN by the researchers. In case of doubts, help and further revision will be provided by external experts supervised by the responsible researcher.

Storage and back-up during the research process

How data and metadata will be stored and backed-up during the research process: During the research process data are stored in an encrypted folder on the personal laptop of the researcher, on external hard-disk and on the cloud (Microsoft Onedrive) of the university account.

How data security and protection of sensitive data will be taken care of during the research: An encrypted folder is used to avoid unauthorized access to data on the researcher's laptop. Security on the computers used for data processing is periodically reinforced by strong passwords and antivirus scan. Only GDPR-compliant cloud storage is used (i.e., located in European servers). Data processing is carried out in compliance with GDPR, as explained in the consent forms (Appendices A). _____

Legal and ethical requirements, codes of conduct

How compliance with legislation on personal data and data security will be ensured:
Data processing is carried out in compliance with GDPR, as explained in the consent forms (Appendices A). These documents have been analysed and approved by the Bioethical Committee of the University of Bologna (23rd June 2023).
More specifically, survey data will be anonymised before publication. Language data (transcripts of naturally-occurring conversations) will be pseudo-anonymised and recordings will only be published with restricted access. If some parts of a recording carrying personal information are found to be so interesting that their publication would substantially improve the results of the research, the participants will be contacted via email and their explicit consent will be asked.

How other legal issues, such as intellectual property rights and ownership, will be managed: Every step of the research which involves data has been discussed with the Privacy Office of the University of Bologna. As for intellectual property and ownership, which particularly relates to the use of images and short phrases, they will be paraphrased and described in publications and will only be published after explicit consent given by the person who produced the image/text and for scientific purposes only.

Ethical issues and codes of conduct: The code of conduct of the researcher throughout research is in line with the procedures of informed consensus discussed with the Bioethical Committee of the University of Bologna. Any ethical issues have been discussed with the Bioethical Committee of the University of Bologna, which gave its anonymous approval.
Moreover, the researcher is devoted to describing the ethical aspects of participatory ethnographic approach to research in the Ph.D. thesis, as well as her position within the community under study.

Data sharing and long-term preservation

How and when data will be shared: Qualitative analysis involving the presentation of small part of data will be shared during conference presentations or in academic publications before the end of the project (December 2025). Field notes and pictures from the ethnographic research may be used within the thesis, which, once discussed, will be published in open-access mode. Academic publications which may derive from the project might contain data of the same nature.
Survey data, once anonymised, will be deposited on AmsActa (institutional repository) and Zenodo. The transcripts of the spoken data will be organised into a corpus and will be shared on the Italian CLARIN repository after the end of the project. The spoken data will be shared on the same location, but with restricted accesses.

What type of data will be preserved: After anonymisation or pseudo-anonymisation procedures as stated in each of the privacy notices, all data will be preserved long-term.

Where data will be preserved long-term: Some of the datasets will be stored on institutional repositories, as follows.
Pilot Survey Data: AMSActa – the institutional repository of the University of Bologna. These data will be preserved in an open-access format.
Survey Data: Zenodo - These data will be preserved in an open-access format.
Spoken Data: ILC4CLARIN
Transcripts of spoken data: ILC4CLARIN
The present document, as well as the links to all datasets, and the codes for data processing will be shared on two research sharing platforms, namely GitHub and OSF. These will be updated with state-of-the-artdocumentation and all datasets version periodically.
Information about data will also be published on the project website thenomadlinguist.eu, property domain of the main researcher (Novella Tedesco).

What methods or software tools are needed to access and use data: The main data product, i.e. the corpus of transcribed spoken interactions, can be downloaded as simple .txt / .xml, as .csv, or explored through the online tool NoSketchEngine (hosted on the University of Bologna servers).
Survey data can be accessed through Excel or any other .csv readers, as they are available in different formats.
Spoken data can be accessed through an audio player, after the data are released to the requesting personby the responsible researcher.

How the application of a unique persistent identifier to each dataset will be ensured: Through deposit in qualified repositories (either a DOI or a Handle will be assigned to each dataset).

Data management responsibilities and resources

Who is responsible for the data management: Novella Tedesco

What resources will be dedicated to data management and ensuring that data will be FAIR: Data management is a fundamental part of the FLO project. Indeed, one of the primary aims of the project is building resources to investigate sociolinguistic variation in English as well as translingual and plurilingual practices for transcultural communication happening in the superdiverse contexts of investigation, i.e.digital nomad communities in Madeira and Canary Islands. For this reason, the present document is constantly updated and has been developed by the responsible researcher with the precious help of the Privacy Office and the Data Steward team of the University of Bologna, particularly with the researcher Bianca Gualandi. Moreover, the repositories where data are deposited apply quality checks on the data. Finally, it should be noted that each dataset includes a detailed readme.txt file where indications about compiling procedures and possible uses of the datasets and accompanying documentation, such as metadata, are explained.
In Figure 1 a summary of FLO's data management is provided.

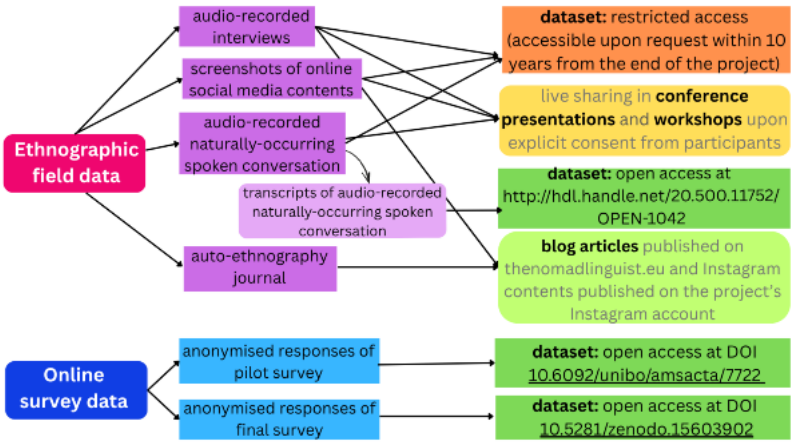


Figure 1. FLO's data overview and accessibility.

The two main sources of data collection, i.e. field research and online survey, led to the creation of different types of datasets, some of which are published in open access and fully adhere to the FAIR data principles (in the green boxes at the right), while access to some data is restricted to guarantee respect of participants' privacy. Nevertheless, these type of data have informed blog articles and live presentations (which in the figure appear enclosed in the rounded boxes in light green and yellow, respectively).

References

Bauman, Z. (2000). *Liquid modernity*. Polity Press.

Canagarajah, S. (Ed.). (2017). *The Routledge handbook of migration and language*. London: Routledge.
Commission Recommendation (EU) 2018/790. Commission Recommendation on access to and preservation of scientific information. EUR-Lex.
García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education.* Palgrave Macmillan.
Reichenberger, I. (2018). Digital nomads – A quest for holistic freedom in work and leisure. *Annals of Leisure Research,* 21(3), 364–380. Taylor & Francis.
Tusting, K. (2020). Ethnography and sociolinguistics in educational research. *Journal of Educational Research*, 15(3), 112–127.
Wang, J., Jin, G., & Li, W. (2023). Changing perceptions of language in sociolinguistics. *Humanities and Social Sciences Communications*, 10, 91. Nature. https://doi.org/10.1057/s41599-023-01574-5.
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3, 160018.*
Woldoff, R. A., & Litchfield, R. C. (2021). *Digital nomads: In search of freedom, community, and meaningful work in the new economy*. Oxford: Oxford University Press.

Appendix A.0 – Consent form and privacy notice for questionnaire responses

Participant information sheet and consent form: questionnaire
Project title: "FLO – The European Fluid Languages Observatory"

1. **Project and responsible people**

You are kindly invited to take part in the online survey 'Where's your language at?', part of the wider scientific project named "FLO – the European Fluid Languages Observatory". The responsible researchers for the study are Silvia Bernardini, Professor at the University of Bologna, and Novella Tedesco, Ph.D. candidate at the University of Bologna, who is also the contact person for this study.
Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. A member of the team can be contacted if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

2. **Project description and objectives**

The FLO project aims to investigate new communication practices and the use of English as a lingua franca (ELF) within the European communities of digital nomads. On the one hand, ethnographic and linguistic data are collected to gain insights about processes of identity construction and (self)representation, knowledge sharing and community building within the European communities of digital nomads, in particular focusing on the communities based in some regions of Spain and Portugal. On the other hand, the project has been conceived with the objective of proposing a translanguaging approach to foreign language teaching focused on successful intercultural communication strategies and the development of a transcultural consciousness. To achieve these goals, the researchers involved in the project are engaged in data collection and analysis through the online questionnaire aimed at gaining ethnographic, sociological and sociolinguistics information about digital nomads in Europe, their social and educational background, their linguistic attitudes and behaviours.

3. **Why have I been chosen?**

Participation is open to any people who regard themselves as digital nomad.

4. **What do I have to do?**

Participation in the FLO research project by filling in the questionnaire involves answering multiple-choice questions as well as other types of questions. It is voluntary and free of charge. Participation in the survey by means of a questionnaire does not require personal data and contact details to be provided, but this possibility is envisaged in the event that the participant is willing to be contacted in order to participate in other stages of the research. The questions concern the definition of 'digital nomad' and the phenomenon of digital nomadism, travel habits, lifestyle, the use of English as a lingua franca and communicative strategies employed in intercultural communication contexts, attendance to language courses, level of education and occupation of the participants.

5. **What will happen to me if I take part?**

Participation in the study is voluntary and free of charge. For the participants, collaboration does not entail any kind of risk or inconvenience. The researchers expect information on the habits, behaviour and lifestyle of the digital nomad community under investigation, as well as information on the linguistic behaviour and communicative strategies adopted by the participants in transcultural communication situations taking place in informal contexts.

6. **Withdrawl from partecipation**

You have the right to withdraw your consent to participate in this study at any time, even without prior notice or specific reasons.

7. **What will happen to the results of the project?**

You have the right to request information on the results and outcome of the research. If you wish to stay up-to-date on the progress and preliminary results of the research, you can follow our updates via social networks, on Instagram (the.nomad.linguist), or visit our website thenomadlinguist.eu. You can also contact the project responsible at any time by emailing novella.tedesco2@unibo.it.

In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

8. **Anonymity**

Data will be processed to eliminate any references that might allow individual statements to be linked to a specific person. Under no circumstances will any brief quotations be traced back to individual persons.

8. Contacts

For any information and clarification on this study or for any needs, please contact researchers Novella Tedesco (novella.tedesco2@unibo.it) and Silvia Bernardini (silvia.bernardini@unibo.it) who are at your disposal for further information or clarification.

Informed consent form

Name _____ Address _____

☐ I confirm that I have read and understand the Participant Information Sheet and any information given orally by the investigator;

☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason;

☐ I have understood all the requirements and risks deriving from my participation in the project;
☐ I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified (except as might be required by law);
☐ I agree that data gathered in this study may be stored anonymously and securely, and may be used for future research.
☐ I agree to receive information regarding any incidental results that may concern me.
☐ I AGREE TO TAKE PART IN THE STUDY, aware that such consent is freely given and may be revoked at any time without disadvantage or prejudice.

_____, _____
(place and date)

_____
(signature of the participant)

_____
(signature of the investigator)

Privacy Notice in compliance with UE GDPR

You are participating in scientific research about the communities of digital nomads in Europe. This privacy notice informs you about the processing of your personal data as part of the research. You have a legal right to receive this information in accordance with the European Union legislation.

Data Controller

- **Data Holder**

Alma Mater Studiorum – Università di Bologna (sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: privacy@unibo.it; PEC: scriviunibo@pec.unibo.it).

- **Data Protection – University of Bologna**

Sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: dpo@unibo.it; PEC: scriviunibo@pec.unibo.it.

Responsible person for personal data processing and research team

- **Novella Tedesco** (novella.tedesco2@unibo.it)
- Silvia Bernardini (silvia.bernardini@unibo.it)
- Cristiana Cervini (cristiana.cervini@unibo.it).

DATA PROCESSING

Your personal data will be handled confidentially and without disclosing them to any outsiders. They will be processed and anonymised by the researchers before result publications of any kinds.

(A) PARTICIPATION IN RESEARCH, RELATED OPERATIONS AND ACTIVITIES

If you agree to participate in this research, some of your personal data will be collected through an online survey accessible at thenomadlinguist.eu/wheresyourlanguageat .
The information you provide will be processed by Novella Tedesco, which will replace all identification data with a code. Please note that your data will only be subject to anonymisation after two years as, within that period,
1 - You may be asked to participate in other questionnaires, interviews or focus groups, which will also be compared and processed in the light of the answers already provided;

2 - It is in the University's interest to be able to contact you again to discuss certain aspects of the research in greater detail and to keep you informed, should you wish, about the results of the research;
3 - It is in the University's interest to be able to contact you again if you have expressed a willingness to participate in linguistic data collection;
4 - It is in the University's interest to contact you again if you have expressed a willingness to participate in targeted interviews or focus groups.
The data collected and archived for the purpose of the study in question, once anonymised, will be retainedindefinitely and will constitute reusable resources for institutions, individuals and the scientific community.
Age and nationality of participants may be disclosed while maintaining anonymity.

(A1) Information on dissemination

Data will only be disseminated in a strictly anonymous form, e.g. through scientific publications, statisticsand scientific conferences.

(B) DATA STORAGE FOR FUTURE RESEARCH ACTIVITIES

Considering that new findings may indicate unprecedented research opportunities for researchers or allow for further study and research on data particular to the study in question, it is possible to allow prolonged storage of the data in non-anonymised form for a period of time of more than two years after the conclusion of this study for possible future research activities and, if necessary, to contact you again to have you express your specific consent for new research, if you so wish.
If, on the other hand, you refuse to consent to the processing described herein, your data will be deleted or anonymised immediately upon expiry of the retention period indicated in (A).

(C) Data storage for administrative purposes

Your data will also be processed for administrative accounting purposes and retained for as long as is strictly necessary for the pursuit of those purposes, with the exception of the ten-year period to ensure the tax, accounting and administrative requirements required by law and, probably, for a longer period, which cannot be determined in advance, as a consequence of various conditions for the lawfulness of the processing (e.g. legal actions making processing necessary for more than ten years).

Legal basis and nature of the data provision

The legal basis for the processing referred to in points (A), (A1) and (B) above lies in consent, pursuant to Article 6, first paragraph, letter a) of Regulation (EU) 2016/679 (General Data Protection Regulation) and, in the case of data falling into special categories, Article 9, second paragraph, letter a) of Regulation (EU) 2016/679. Consent to processing for the purposes described above is optional.
However, it is specified that the provision of data for the purposes set out in points (A) and (A1) is not obligatory but is essential for the realisation of the purposes described. Refusal to grant them will therefore not allow you to participate in the research in question.
The provision of data for the purposes described in (B) is optional, not resulting from a regulatory obligation, but it is necessary to allow the data to be kept for a longer period than the period envisaged for the conclusion of the present study in order to guarantee future research activities and, if necessary, to contact you again so that you may express your

specific consent for a new research. Failure to provide data for these purposes will only result in the impossibility of carrying out what has been described above.
The provision of data for the purposes described in point (C) is necessary in order to comply with legal obligations regarding the retention of administrative and accounting documents. In this case, the legal basis can be found in Article 6(1)(c) and (e) of Regulation (EU) 2016/679.

rights of the data subject

Cancellation of consent (GDPR, Article 7)
You have the right to cancel your consent if the processing of personal data is based on consent. Such a cancellation has no impact on the lawfulness of consent-based processing conducted before the cancellation of consent.
Right to access your personal data (GDPR, Article 15)
You have the right to get to know whether and which personal data of yours are processed. If you wish, you can also request a copy of your personal data to be processed.
Right to rectification (GDPR, Article 16)
If there are any inaccuracies or errors in your personal data to be processed, you are entitled to request that these be rectified or supplemented.
Right to erasure (GDPR, Article 17)
You have the right to demand in some cases that your personal data be erased. However, the right of erasure is not applicable if the erasure would prevent or greatly hinder reaching the goal of processing in a scientific research.
Right to restriction of processing (GDPR, Article 18)
You have the right to restrict the processing of your personal data in some cases, like when you challenge the correctness of your personal data.
Right to data portability (GDPR, Article 20)
You have the right to receive your submitted personal data in an organised, generally used and machine-readable format, and also the right to transfer the data to another Data Controller if possible and processing takes place automatically.
Deviating from the rights
In some individual cases, it is possible to deviate from the described rights on the grounds stipulated in the GDPR and the Data Protection Act insofar as the rights would prevent or greatly hinder reaching the goals of scientific or historical research or statistical purposes. The need for deviating from the rights is always assessed case-specifically. It is also possible to deviate from the rights if the data subject cannot, or cannot any longer, be identified.
Archiving
When personal data are processed for archiving purposes serving a public interest, the data subjects do not generally have the above-described rights. The authenticity, reliability, and research value of archived materials would be at risk if the personal data included were changed. Deviation from data subject rights is based on the Data Protection Act, section 32, and the GDPR, Article 17.3.
Profiling and automatised decision-making
In this research, your personal data will not be used for any automatic decision-making. In this research, the purpose of the processing of personal data is not to assess your personal qualities, i.e. profiling, but personal data and qualities are considered from the perspective of broader scientific research.

Incidental results

In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella

Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

PERSONAL DATA PROCESSING CONSENT FORM

Name _____

Date of Birth _____

Address _____

☐ I agree to my personal data being stored and processed for scientific and statistical research for the purposes above described (Section A) – NECESSARY for the participation in the research;
☐ I agree to my personal data being processed and disclosed in publications for the research purposes above described (Section A1) – NOT NECESSARY for the participation in the research;
☐ I agree to my personal data being stored and used after the present research has concluded (Section B) – NOT NECESSARY for the participation in the research.

Date Signature

_____ _____

Appendix A.1 – Consent form and privacy notice for recorded interviews

Participant information sheet and consent form: recorded interview
Project title: "Fluid languages for fluid communities: a translanguaging approach to the study of digital nomads' communication in Europe"

1. **Project and responsible people**

You are kindly invited to take part in an interview for FLO, the European Fluid Languages Observatory, in the framework of a research into informal plurilingual and intercultural communication within the communities of digital nomads in Europe. The responsible researchers for the study are Silvia Bernardini, Professor at the University of Bologna, and Novella Tedesco, Ph.D. candidate at the University of Bologna, who is also the contact person for this study.
Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please contact us if there is anything that is not clear or if you would like more information, and take time to decide whether or not you wish to take part.

2. **Project description and objectives**

FLO aims to investigate new communication practices and the use of English as a lingua franca (ELF) within the European communities of digital nomads. On the one hand, ethnographic and linguistic data are collected to gain insights about processes of identity

construction and (self)representation, knowledge sharing and community building within the European communities of digital nomads, in particular focusingon the communities based in some regions of Spain and Portugal. On the other hand, the project has been conceived with the objective of proposing a translanguaging approach to foreign language teaching focused on successful intercultural communication strategies and the development of a transculturalconsciousness. To achieve these goals, the researchers involved in the project record interviews about topics related to digital nomads, to understand how these emerging communities use languages to build intercultural relations and communicate with the always-changing surrounding environment.

3. **Why have I been chosen?**

We would like to interview you, as you are considered to have information and knowledge from which the research would benefit.

4. **What do I have to do?**

You are asked to engage in the discussion and answer questions from the interviewer. Participation is voluntary and free of charge. Interviews will be recorded for research purposes only. If you prefer, you can record the interview on your own device. Participation in the interview does not require personal data and contact details to be provided, but this possibility is envisaged in the event that the participant is willing to be contacted in order to participate in other stages of the research. The questions concern the definition of 'digital nomad' and the phenomenon of digital nomadism, travel habits, lifestyle, the use of English as a lingua franca and communicative strategies employed in intercultural communication contexts, attendance to language courses, level of education and occupation of the participants.

5. **What will happen to me if I take part?**

For the participants, collaboration does not entail any kind of risk or inconvenience. The researchers expect information on the habits, behaviour and lifestyle of the digital nomad community under investigation, as well as information on the linguistic behaviour and communicative strategies adopted by the participants in transcultural communication situations taking place in informal contexts.

6. **Withdrawl from partecipation**

You have the right to withdraw your consent to participate in this study at any time, even without prior notice or specific reasons.

7. **What will happen to the results of the project?**

You have the right to request information on the results and outcome of the research. If you wish to stay up-to-date on the progress and preliminary results of the research, you can follow our updates on Instagram or reading the posts on our blog. You can also contact the project responsible at any time by emailing novella.tedesco2@unibo.it.
In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

8. **Contacts**

For any information and clarification on this study or for any needs, please contact researchers Novella Tedesco (novella.tedesco2@unibo.it) and Silvia Bernardini (silvia.bernardini@unibo.it) who are at your disposal for further information or clarification.

Informed consent form

Name _____ Address _____

☐ I confirm that I have read and understand the Participant Information Sheet and any information given orally by the investigator;

☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason;

☐ I have understood all the requirements and risks deriving from my participation in the project;

☐ I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified (except as might be required by law);

☐ I agree that data gathered in this study may be stored anonymously and securely, and may be used for future research.

☐ I agree to receive information regarding any incidental results that may concern me.

☐ I AGREE TO TAKE PART IN THE STUDY, aware that such consent is freely given and may be revoked at any time without disadvantage or prejudice.

_____, _____
(place and date)

_____
(signature of the participant)

_____
(signature of the investigator)

Privacy Notice in compliance with UE GDPR

You are participating in scientific research about the communities of digital nomads in Europe. This privacy notice informs you about the processing of your personal data as part of the research. You have a legal right to receive this information in accordance with the European Union legislation.

Data Controller

- **Data Holder**

Alma Mater Studiorum – Università di Bologna (sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: privacy@unibo.it; PEC: scriviunibo@pec.unibo.it).

- **Data Protection – University of Bologna**

Sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: dpo@unibo.it; PEC: scriviunibo@pec.unibo.it.

Responsible person for personal data processing and research team

- **Novella Tedesco** (novella.tedesco2@unibo.it)
- Silvia Bernardini (silvia.bernardini@unibo.it)
- Cristiana Cervini (cristiana.cervini@unibo.it).

DATA PROCESSING

Your personal data will be handled confidentially and without disclosing them to any outsiders. They will be processed and anonymised by the researchers before result publications of any kinds.

(A) PARTICIPATION IN RESEARCH, RELATED OPERATIONS AND ACTIVITIES

If you agree to participate in this research, some of your personal data will be collected through audio- or video-recorded interviews.

The information you provide will be processed by Novella Tedesco, which will replace all identification data with a code. Please note that your data will only be subject to anonymisation after two years as, within that period,
1 - You may be asked to participate in questionnaires, other interviews or focus groups, which will also be compared and processed in the light of the answers already provided;
2 - It is in the University's interest to be able to contact you again to discuss certain aspects of the research in greater detail and to keep you informed, if you wish, about the results of the research;
3 - It is in the University's interest to be able to contact you again if you have expressed a willingness to participate in linguistic data collection;
4 - It is in the University's interest to contact you again if you have expressed a willingness to participate in targeted interviews or focus groups, or to compile the questionnaire online.
The data collected and archived for the purpose of the study in question, once anonymised, will be retainedindefinitely and will constitute reusable resources for institutions, individuals and the scientific community.
Age and nationality of participants may be disclosed while maintaining anonymity.

(A1) Information on dissemination

1) Data will only be disseminated in a strictly anonymous form, e.g. through scientific publications, statisticsand scientific conferences.
2) If the interview, in any case without any identification data, proves to be socially useful and usable to show the linguistic behaviour of the population under consideration, it may - with your consent - be the subject of communication and/or dissemination. The communication or dissemination of the data described above will only take place after an assessment of the relevance and not of the excess of processing with respect to the purpose of the collection or if the non-publication of what emerged in the interview could negatively affect the quality of the research/study. The recording of the interview and/or the transcription of the content, in any case without any identification, could be communicated to the project partners and be subject to dissemination/publication (e.g. in scientific journals, Internet, databases accessible to other researchers, institutional or disciplinary data repositories i.e. document management information systems maintained by individual research institutions or associations and consortia of research institutions for a particular scientific community, where data and their meta-information are stored in a secure and reliable manner that guarantees their authenticity and inalterability and distributed to the public in a free and open form, etc.).
3) If you are a well-known person, or if you perform a public function in a certain role or position, it may be in the public interest or socially valuable to let third parties know the information from the interview concerning the institution or role you represent (including, for example, your first and last name). You may therefore be asked for consent to your personal data be processed and disseminated for the following purposes
- for the purpose of enabling the processing and dissemination of information that, due to the nature of your role, enables researchers to discover, interpret and revise theories, facts and behaviours relating to the field of knowledge on the above topics;

- deposit of project documents, containing your data, on institutional or disciplinary data repositories;
- so that the information you provide contributes to the formation of public opinion about your views and/or facts relevant to the community.

(B) DATA STORAGE FOR FUTURE RESEARCH ACTIVITIES

Considering that new findings may indicate unprecedented research opportunities for researchers or allow for further study and research on data particular to the study in question, it is possible to allow prolonged storage of the data in non-anonymised form for a period of time of more than two years after the conclusion of this study for possible future research activities and, if necessary, to contact you again to have you express your specific consent for new research, if you so wish.
If, on the other hand, you refuse to consent to the processing described herein, your data will be deleted or anonymised immediately upon expiry of the retention period indicated in (A).

(C) Data storage for administrative purposes

Your data will also be processed for administrative accounting purposes and retained for as long as is strictly necessary for the pursuit of those purposes, with the exception of the ten-year period to ensure the tax, accounting and administrative requirements required by law and, probably, for a longer period, which cannot be determined in advance, as a consequence of various conditions for the lawfulness of the processing (e.g. legal actions making processing necessary for more than ten years).

Legal basis and nature of the data provision

The legal basis for the processing referred to in points (A), (A1) and (B) above lies in consent, pursuant to Article 6, first paragraph, letter a) of Regulation (EU) 2016/679 (General Data Protection Regulation) and, in the case of data falling into special categories, Article 9, second paragraph, letter a) of Regulation (EU) 2016/679. Consent to processing for the purposes described above is optional.
However, it is specified that the provision of data for the purposes set out in points (A) and (A1) is not obligatory but is essential for the realisation of the purposes described. Refusal to grant them will therefore not allow you to participate in the research in question.
The provision of data for the purposes described in (B) is optional, not resulting from a regulatory obligation, but it is necessary to allow the data to be kept for a longer period than the period envisaged for the conclusion of the present study in order to guarantee future research activities and, if necessary, to contact you again so that you may express your specific consent for a new research. Failure to provide data for these purposes will only result in the impossibility of carrying out what has been described above.
The provision of data for the purposes described in point (C) is necessary in order to comply with legal obligations regarding the retention of administrative and accounting documents. In this case, the legal basis can be found in Article 6(1)(c) and (e) of Regulation (EU) 2016/679.

rights of the data subject

Cancellation of consent (GDPR, Article 7)

You have the right to cancel your consent if the processing of personal data is based on consent. Such a cancellation has no impact on the lawfulness of consent-based processing conducted before the cancellation of consent.

Right to access your personal data (GDPR, Article 15)

You have the right to get to know whether and which personal data of yours are processed. If you wish, you can also request a copy of your personal data to be processed.

Right to rectification (GDPR, Article 16)

If there are any inaccuracies or errors in your personal data to be processed, you are entitled to request that these be rectified or supplemented.

Right to erasure (GDPR, Article 17)

You have the right to demand in some cases that your personal data be erased. However, the right of erasure is not applicable if the erasure would prevent or greatly hinder reaching the goal of processing in a scientific research.

Right to restriction of processing (GDPR, Article 18)

You have the right to restrict the processing of your personal data in some cases, like when you challenge the correctness of your personal data.

Right to data portability (GDPR, Article 20)

You have the right to receive your submitted personal data in an organised, generally used and machine-readable format, and also the right to transfer the data to another Data Controller if possible and processing takes place automatically.

Deviating from the rights

In some individual cases, it is possible to deviate from the described rights on the grounds stipulated in the GDPR and the Data Protection Act insofar as the rights would prevent or greatly hinder reaching the goals of scientific or historical research or statistical purposes. The need for deviating from the rights is always assessed case-specifically. It is also possible to deviate from the rights if the data subject cannot, or cannot any longer, be identified.

Archiving

When personal data are processed for archiving purposes serving a public interest, the data subjects do not generally have the above-described rights. The authenticity, reliability, and research value of archived materials would be at risk if the personal data included were changed. Deviation from data subject rights is based on the Data Protection Act, section 32, and the GDPR, Article 17.3.

Profiling and automatised decision-making

In this research, your personal data will not be used for any automatic decision-making. In this research, the purpose of the processing of personal data is not to assess your personal qualities, i.e. profiling, but personal data and qualities are considered from the perspective of broader scientific research.

Incidental results

In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

PERSONAL DATA PROCESSING CONSENT FORM

Name _____

Date of Birth _____

Address _____

☐ I agree to my personal data being stored and processed for scientific and statistical research for the purposes above described (Section A) – NECESSARY for the participation in the research;

☐ I agree to the processing of images representing me and the personal data related to them for the purposes above described (Section A) – NECESSARY for the participation in the research;

☐ I agree to my personal data being processed and disclosed in publications for the research purposes above described (Section A1) – NOT NECESSARY for the participation in the research;

☐ I agree to my personal data being stored and used after the present research has concluded (Section B) – NOT NECESSARY for the participation in the research.

Date Signature

_____ _____

Appendix A.2 – Consent form and privacy notice for the pictures in the ethnographic research

Privacy Notice in compliance with UE GDPR

You are participating in scientific research about the communities of digital nomads in Europe. This privacy notice informs you about the processing of your personal data as part of the research. You have a legal right to receive this information in accordance with the European Union legislation.

Data Controller

- **Data Controller**

Alma Mater Studiorum – Università di Bologna (sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: privacy@unibo.it; PEC: scriviunibo@pec.unibo.it).

- **Data Protection Officer**

Registrated address: via Zamboni 33, 40126 - Bologna, Italia; e-mail: dpo@unibo.it; PEC: scriviunibo@pec.unibo.it.

Responsible person for personal data processing and research team

- **Novella Tedesco** (novella.tedesco2@unibo.it)
- Silvia Bernardini (silvia.bernardini@unibo.it)
- Cristiana Cervini (cristiana.cervini@unibo.it).

DATA PROCESSING

Your personal data will be handled confidentially and without disclosing them to any outsiders. They will be processed and anonymised by the researchers before result publications of any kinds.

(A) PARTICIPATION IN RESEARCH, RELATED OPERATIONS AND ACTIVITIES

If you agree to participate in this research, some of your personal data will be collected during naturally occurring conversations which will be audio- or video-recorded. In the case of self-

recording you may be asked to cancel the data you have recorded after you have given the recordings to the responsible researcher, Novella Tedesco.

The information you provide will be processed by Novella Tedesco, which will replace all identification data with a code. Please note that your data will only be subject to anonymisation after two years as, within that period,
1 - You may be asked to participate in questionnaires, interviews or focus groups, which will also be compared and processed in the light of the answers already provided;
2 - It is in the University's interest to be able to contact you again to discuss certain aspects of the research in greater detail and to keep you informed, if you wish, about the results of the research;
3 - It is in the University's interest to contact you again if you have expressed a willingness to participate in targeted interviews or focus groups.
The data collected and archived for the purpose of the study in question, once anonymised, will be retainedindefinitely and will constitute reusable resources for institutions, individuals and the scientific community.
Age and nationality of participants may be disclosed while maintaining anonymity.

(A1) Information on dissemination

1. Data will only be disseminated in a strictly anonymous form, e.g. through scientific publications, statisticsand scientific conferences.
2. If the picture, in any case without any identification data, proves to be socially useful and usable to show the linguistic behaviour of the population under consideration, it may - with your consent - be the subject of communication and/or dissemination. The communication or dissemination of the data described above will only take place after an assessment of the relevance and not of the excess of processing with respect to the purpose of the collection or if the non-publication of what emerged in the interview could negatively affect the quality of the research/study. The recording of the interview and/or the transcription of the content, in any case without any identification, could be communicated to the project partners and be subject to dissemination/publication (e.g. in scientific journals, Internet, databases accessible to other researchers, institutional or disciplinary data repositories i.e. document management information systems maintained by individual research institutions or associations and consortia of research institutions for a particular scientific community, where data and their meta-information are stored in a secure and reliable manner that guarantees their authenticity and inalterability and distributed to the public in a free and open form, etc.).

3) If you are a well-known person, or if you perform a public function in a certain role or position, it may be in the public interest or socially valuable to let third parties know the information concerning the institution or role you represent (including, for example, your first and last name). You may therefore be asked for consent to your personal data be processed and disseminated for the following purposes

- for the purpose of enabling the processing and dissemination of information that, due to the nature of your role, enables researchers to discover, interpret and revise theories, facts and behaviours relating to the field of knowledge on the above topics;
- deposit of project documents, containing your data, on institutional or disciplinary data repositories;

- so that the information you provide contributes to the formation of public opinion about your views and/or facts relevant to the community.

(B) DATA STORAGE FOR FUTURE RESEARCH ACTIVITIES

Considering that new findings may indicate unprecedented research opportunities for researchers or allow for further study and research on data particular to the study in question, it is possible to allow prolonged storage of the data in non-anonymised form for a period of time of more than two years after the conclusion of this study for possible future research activities and, if necessary, to contact you again to have you express your specific consent for new research, if you so wish.

If, on the other hand, you refuse to consent to the processing described herein, your data will be deleted or anonymised immediately upon expiry of the retention period indicated in (A).

(C) Data storage for administrative purposes

Your data will also be processed for administrative accounting purposes and retained for as long as is strictly necessary for the pursuit of those purposes, with the exception of the ten-year period to ensure the tax, accounting and administrative requirements required by law and, probably, for a longer period, which cannot be determined in advance, as a consequence of various conditions for the lawfulness of the processing (e.g. legal actions making processing necessary for more than ten years).

Legal basis and nature of the data provision

The legal basis for the processing referred to in points (A), (A1) and (B) above lies in consent, pursuant to Article 6, first paragraph, letter a) of Regulation (EU) 2016/679 (General Data Protection Regulation) and, in the case of data falling into special categories, Article 9, second paragraph, letter a) of Regulation (EU) 2016/679. Consent to processing for the purposes described above is optional.

However, it is specified that the provision of data for the purposes set out in points (A) and (A1) is not obligatory but is essential for the realisation of the purposes described. Refusal to grant them will therefore not allow you to participate in the research in question.

The provision of data for the purposes described in (B) is optional, not resulting from a regulatory obligation, but it is necessary to allow the data to be kept for a longer period than the period envisaged for the conclusion of the present study in order to guarantee future research activities and, if necessary, to contact you again so that you may express your specific consent for a new research. Failure to provide data for these purposes will only result in the impossibility of carrying out what has been described above.

The provision of data for the purposes described in point (C) is necessary in order to comply with legal obligations regarding the retention of administrative and accounting documents. In this case, the legal basis can be found in Article 6(1)(c) and (e) of Regulation (EU) 2016/679.

rights of the data subject

Cancellation of consent (GDPR, Article 7)

You have the right to cancel your consent if the processing of personal data is based on consent. Such a cancellation has no impact on the lawfulness of consent-based processing conducted before the cancellation of consent.

Right to access your personal data (GDPR, Article 15)

You have the right to get to know whether and which personal data of yours are processed. If you wish, you can also request a copy of your personal data to be processed.

Right to rectification (GDPR, Article 16)

If there are any inaccuracies or errors in your personal data to be processed, you are entitled to request that these be rectified or supplemented.

Right to erasure (GDPR, Article 17)

You have the right to demand in some cases that your personal data be erased. However, the right of erasure is not applicable if the erasure would prevent or greatly hinder reaching the goal of processing in a scientific research.

Right to restriction of processing (GDPR, Article 18)

You have the right to restrict the processing of your personal data in some cases, like when you challenge the correctness of your personal data.

Right to data portability (GDPR, Article 20)

You have the right to receive your submitted personal data in an organised, generally used and machine-readable format, and also the right to transfer the data to another Data Controller if possible and processing takes place automatically.

Deviating from the rights

In some individual cases, it is possible to deviate from the described rights on the grounds stipulated in the GDPR and the Data Protection Act insofar as the rights would prevent or greatly hinder reaching the goals of scientific or historical research or statistical purposes. The need for deviating from the rights is always assessed case-specifically. It is also possible to deviate from the rights if the data subject cannot, or cannot any longer, be identified.

Archiving

When personal data are processed for archiving purposes serving a public interest, the data subjects do not generally have the above-described rights. The authenticity, reliability, and research value of archived materials would be at risk if the personal data included were changed. Deviation from data subject rights is based on the Data Protection Act, section 32, and the GDPR, Article 17.3.

Profiling and automatised decision-making

In this research, your personal data will not be used for any automatic decision-making. In this research, the purpose of the processing of personal data is not to assess your personal qualities, i.e. profiling, but personal data and qualities are considered from the perspective of broader scientific research.

Incidental results

In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

PERSONAL DATA PROCESSING CONSENT FORM

Name _____

Date of Birth _____

Address _____

☐ I agree to my personal data being stored and processed for scientific and statistical research for the purposes above described (Section A) – NECESSARY for the participation in the research;

☐ I agree to the processing of images representing me and the personal data related to them for the purposes above described (Section A) – NECESSARY for the participation in the research;

☐ I agree to my personal data being processed and disclosed in publications for the research purposes above described (Section A1) –NECESSARY for the participation in the research;

☐ I agree to my personal data being stored and used after the present research has concluded (Section B) – NOT NECESSARY for the participation in the research.

Date Signature

_____ _____

Appendix A.3 – Consent form and privacy notice for the spoken data collection

Privacy Notice in compliance with UE GDPR

You are participating in scientific research about the communities of digital nomads in Europe. This privacy notice informs you about the processing of your personal data as part of the research. You have a legal right to receive this information in accordance with the European Union legislation.

Data Controller

- **Data Controller**

Alma Mater Studiorum – Università di Bologna (sede legale: via Zamboni 33, 40126 - Bologna, Italia; e-mail: privacy@unibo.it; PEC: scriviunibo@pec.unibo.it).

- **Data Protection Officer**

Registrated address: via Zamboni 33, 40126 - Bologna, Italia; e-mail: dpo@unibo.it; PEC: scriviunibo@pec.unibo.it.

Responsible person for personal data processing and research team

- **Novella Tedesco** (novella.tedesco2@unibo.it)
- Silvia Bernardini (silvia.bernardini@unibo.it)
- Cristiana Cervini (cristiana.cervini@unibo.it).

DATA PROCESSING

Your personal data will be handled confidentially and without disclosing them to any outsiders. They will be processed and anonymised by the researchers before result publications of any kinds.

(A) PARTICIPATION IN RESEARCH, RELATED OPERATIONS AND ACTIVITIES

If you agree to participate in this research, some of your personal data will be collected during naturally occurring conversations which will be audio- or video-recorded. In the case of self-

recording you may be asked to cancel the data you have recorded after you have given the recordings to the responsible researcher, Novella Tedesco.

The information you provide will be processed by Novella Tedesco, which will replace all identification data with a code. Please note that your data will only be subject to anonymisation after two years as, within that period,
1 - You may be asked to participate in questionnaires, interviews or focus groups, which will also be compared and processed in the light of the answers already provided;
2 - It is in the University's interest to be able to contact you again to discuss certain aspects of the research in greater detail and to keep you informed, if you wish, about the results of the research;
3 - It is in the University's interest to contact you again if you have expressed a willingness to participate in targeted interviews or focus groups.
The data collected and archived for the purpose of the study in question, once anonymised, will be retainedindefinitely and will constitute reusable resources for institutions, individuals and the scientific community.
Age and nationality of participants may be disclosed while maintaining anonymity.

(A1) Information on dissemination

1. Data will only be disseminated in a strictly anonymous form, e.g. through scientific publications, statisticsand scientific conferences.
2. If the conversation, in any case without any identification data, proves to be socially useful and usable to show the linguistic behaviour of the population under consideration, it may - with your consent - be the subject of communication and/or dissemination. The communication or dissemination of the data described above will only take place after an assessment of the relevance and not of the excess of processing with respect to the purpose of the collection or if the non-publication of what emerged in the interview could negatively affect the quality of the research/study. The recording of the interview and/or the transcription of the content, in any case without any identification, could be communicated to the project partners and be subject to dissemination/publication (e.g. in scientific journals, Internet, databases accessible to other researchers, institutional or disciplinary data repositories i.e. document management information systems maintained by individual research institutions or associations and consortia of research institutions for a particular scientific community, where data and their meta-information are stored in a secure and reliable manner that guarantees their authenticity and inalterability and distributed to the public in a free and open form, etc.).

3) If you are a well-known person, or if you perform a public function in a certain role or position, it may be in the public interest or socially valuable to let third parties know the information from the interview concerning the institution or role you represent (including, for example, your first and last name). You may therefore be asked for consent to your personal data be processed and disseminated for the following purposes

- for the purpose of enabling the processing and dissemination of information that, due to the nature of your role, enables researchers to discover, interpret and revise theories, facts and behaviours relating to the field of knowledge on the above topics;
- deposit of project documents, containing your data, on institutional or disciplinary data repositories;

- so that the information you provide contributes to the formation of public opinion about your views and/or facts relevant to the community.

(B) DATA STORAGE FOR FUTURE RESEARCH ACTIVITIES

Considering that new findings may indicate unprecedented research opportunities for researchers or allow for further study and research on data particular to the study in question, it is possible to allow prolonged storage of the data in non-anonymised form for a period of time of more than two years after the conclusion of this study for possible future research activities and, if necessary, to contact you again to have you express your specific consent for new research, if you so wish.

If, on the other hand, you refuse to consent to the processing described herein, your data will be deleted or anonymised immediately upon expiry of the retention period indicated in (A).

(C) Data storage for administrative purposes

Your data will also be processed for administrative accounting purposes and retained for as long as is strictly necessary for the pursuit of those purposes, with the exception of the ten-year period to ensure the tax, accounting and administrative requirements required by law and, probably, for a longer period, which cannot be determined in advance, as a consequence of various conditions for the lawfulness of the processing (e.g. legal actions making processing necessary for more than ten years).

Legal basis and nature of the data provision

The legal basis for the processing referred to in points (A), (A1) and (B) above lies in consent, pursuant to Article 6, first paragraph, letter a) of Regulation (EU) 2016/679 (General Data Protection Regulation) and, in the case of data falling into special categories, Article 9, second paragraph, letter a) of Regulation (EU) 2016/679. Consent to processing for the purposes described above is optional.

However, it is specified that the provision of data for the purposes set out in points (A) and (A1) is not obligatory but is essential for the realisation of the purposes described. Refusal to grant them will therefore not allow you to participate in the research in question.

The provision of data for the purposes described in (B) is optional, not resulting from a regulatory obligation, but it is necessary to allow the data to be kept for a longer period than the period envisaged for the conclusion of the present study in order to guarantee future research activities and, if necessary, to contact you again so that you may express your specific consent for a new research. Failure to provide data for these purposes will only result in the impossibility of carrying out what has been described above.

The provision of data for the purposes described in point (C) is necessary in order to comply with legal obligations regarding the retention of administrative and accounting documents. In this case, the legal basis can be found in Article 6(1)(c) and (e) of Regulation (EU) 2016/679.

rights of the data subject

Cancellation of consent (GDPR, Article 7)

You have the right to cancel your consent if the processing of personal data is based on consent. Such a cancellation has no impact on the lawfulness of consent-based processing conducted before the cancellation of consent.

Right to access your personal data (GDPR, Article 15)

You have the right to get to know whether and which personal data of yours are processed. If you wish, you can also request a copy of your personal data to be processed.

Right to rectification (GDPR, Article 16)

If there are any inaccuracies or errors in your personal data to be processed, you are entitled to request that these be rectified or supplemented.

Right to erasure (GDPR, Article 17)

You have the right to demand in some cases that your personal data be erased. However, the right of erasure is not applicable if the erasure would prevent or greatly hinder reaching the goal of processing in a scientific research.

Right to restriction of processing (GDPR, Article 18)

You have the right to restrict the processing of your personal data in some cases, like when you challenge the correctness of your personal data.

Right to data portability (GDPR, Article 20)

You have the right to receive your submitted personal data in an organised, generally used and machine-readable format, and also the right to transfer the data to another Data Controller if possible and processing takes place automatically.

Deviating from the rights

In some individual cases, it is possible to deviate from the described rights on the grounds stipulated in the GDPR and the Data Protection Act insofar as the rights would prevent or greatly hinder reaching the goals of scientific or historical research or statistical purposes. The need for deviating from the rights is always assessed case-specifically. It is also possible to deviate from the rights if the data subject cannot, or cannot any longer, be identified.

Archiving

When personal data are processed for archiving purposes serving a public interest, the data subjects do not generally have the above-described rights. The authenticity, reliability, and research value of archived materials would be at risk if the personal data included were changed. Deviation from data subject rights is based on the Data Protection Act, section 32, and the GDPR, Article 17.3.

Profiling and automatised decision-making

In this research, your personal data will not be used for any automatic decision-making. In this research, the purpose of the processing of personal data is not to assess your personal qualities, i.e. profiling, but personal data and qualities are considered from the perspective of broader scientific research.

Incidental results

In the event information relating to dangerous health conditions, committed crimes, behaviour or events in personal life that have harmed or may harm the participant in the study or others is obtained accidentally, this information will be treated as stated in the incidental data management document. To view this document, you may contact Novella Tedesco. Depending on the severity of the incident and the people involved, the researchers will decide whether to contact the person concerned or the relevant authorities. In any case, these results will not be disclosed without the explicit consent of the participant.

PERSONAL DATA PROCESSING CONSENT FORM

Name _____

Date of Birth _____

Address _____

☐ I agree to my personal data being stored and processed for scientific and statistical research for the purposes above described (Section A) – NECESSARY for the participation in the research;

☐ I agree to the processing of images representing me and the personal data related to them for the purposes above described (Section A) – NOT NECESSARY for the participation in the research;

☐ I agree to my personal data being processed and disclosed in publications for the research purposes above described (Section A1) – NOT NECESSARY for the participation in the research;

☐ I agree to my personal data being stored and used after the present research has concluded (Section B) – NOT NECESSARY for the participation in the research.

Date Signature

_____ _____

Appendix B – Contextual notes

Type of event:
Event planning: (planned, semi-spontaneous, spontaneous)
Context: (public, controlled environment, private)
Description:
Number of researchers who took part in the event:
Number of participants:
Number of recordings:
Recording n. 1
Speakers:
Speakers' nationality:
Speakers' languages:
Speakers' travelling habits:

Appendix 3. *Where's your language at? Questionnaire*[61]

LimeSurvey

By participating in this survey, you are helping The Nomad Linguist to learn something more about digital nomads' behaviours and attitudes when it comes to communication practices in intercultural contexts. More information about the research project can be found in the next section and in the Information Sheet, that we ask you to read before you start the questionnaire.

Section A: Personal Information

A1. Full Name

These data will be processed in accordance with the GDPR regulations, as explained in the privacy notice. No information will be used for purposes other than scientific research.

A2. What is your gender?

We are not interested in knowing your biological sex, or your sexual orientation. This is the definition of gender given by the World Health Organization, to which we are broadly referring.

- Female ☐
- Male ☐
- Non-binary ☐
- Prefer not to say ☐

A3. Where were you born?

- Europe ☐
- United Kingdom ☐
- Asia ☐
- North America ☐
- Central America ☐
- South America ☐
- Australia ☐
- New Zeland ☐
- Africa ☐

[61] https://osf.io/m6bdc.

LimeSurvey

A4. What is your nationality?

A5. What is your age?

A6. What is your level of education?

- High school
- Bachelor's Degree
- Master's Degree
- Ph.D.

A7. In which of these fields is your education?

- Languages and linguistics
- Arts & Humanities (other than languages)
- Social Sciences & Journalism
- Business, Administration & Law
- Natural Sciences, Mathematics & Statistics
- Information & Communication Technologies
- Engineering, Manufacturing & Construction
- Agriculture, Forestry, Fisheries & Veterinary medicine
- Health & Welfare
- Services (Transport, Catering & tourism, Nursing)
- Other

Other

A8. What is currently your main occupation?

- work
- study
- both

LimeSurvey

A9. In what field do you work?

- [ ] Education
- [ ] Health and social care
- [ ] Accommodation and food
- [ ] Arts and creative services
- [ ] ICT services
- [ ] Marketing
- [ ] Professional services (no marketing)
- [ ] Construction
- [ ] Finance and insurance (including Crypto)
- [ ] Public sector and defence
- [ ] Sale and retail
- [ ] Agriculture, forestry and fishing
- [ ] Other

Other

A10. How much of your work-related activities are carried out remotely?

(by remote work, we mean that you use technologies to perform work from a distance)

- [ ] about 25% or less
- [ ] about 50%
- [ ] about 75%
- [ ] All of them

A11. Do you travel often?

By 'travel often' we mean that in the last year you have been to 3 or more cities or places in or outside your country.

- [ ] Yes
- [ ] No

A12. Do you work remotely while travelling?

- Yes ☐
- No ☐

A13. Why do you move?

- for pleasure ☐
- for family reasons ☐
- for work ☐
- for study ☐

A14. Do you regard yourself as a digital nomad?

- Yes ☐
- No ☐

Section B: Language Background and Education

B1. Do you consider yourself bilingual (or plurilingual)?

By 'bilingual' or 'plurilingual' we mean that you currently speak two or more languages, you can use them effectively and without particular effort to communicate in your daily life. Note that this definition is very broad: feel free to give the answer you feel most appropriate to describe your language situation.

- Yes ☐
- No ☐

B2. Which languages do you use in your daily life?

Write one language in each box. Only use the necessary boxes.

LimeSurvey

B3. **Which language do you use in your daily life?**

B4. **How many languages can you speak?**

By the expression 'can speak a language' we mean that you are generally able to communicate using this language. You should include in your count both the languages you use on a daily basis and the languages you know but don't use so often.

B5. **Which language(s) did you learn at school?**

- German ☐
- French ☐
- English ☐
- Italian ☐
- Russian ☐
- Dutch ☐
- Spanish ☐
- Romanian ☐
- Greek ☐
- Polish ☐
- Turkish ☐
- Catalan ☐
- Ukrainian ☐
- Serbian ☐
- Danish ☐
- Norwegian ☐
- Finnish ☐
- Portuguese ☐
- Slovenian ☐
- Swedish ☐
- Danish ☐

LimeSurvey

Other ☐

Other

B6. Have you learned any other languages outside school (e.g. private language courses, online, socializing with your friends, travelling)?

Yes ☐

No ☐

B7. Are half or more of your studies in English (e.g. courses in English, exams in English, attending university in a country where English is the natioanl language)?

Yes ☐

No ☐

B8. Have half or more of your studies been in English (e.g. courses in English, exams in English, attending university in a country where English is the natioanl language)?

Yes ☐

No ☐

B9. Are half or more of your work-related activities in English?

Yes ☐

No ☐

B10. How many of your friends are your same nationality?

about 25% or less ☐

about 50% ☐

about 75% ☐

All of them ☐

B11. How strongly do you agree or disagree with the following statements?

(1 – strongly disagree, 2 – disagree, 3 – neither agree nor disagree, 4 – agree, 5 – strongly agree)

	1	2	3	4	5
I am satisfied with the education I got at school in English	☐	☐	☐	☐	☐
I am satisfied with the education I got at school in languages other than English.	☐	☐	☐	☐	☐

	1	2	3	4	5
I am satisfied with the amount of intercultural activities I did at school (e.g. visits in museums, travels, cultural exchanges).	☐	☐	☐	☐	☐

Section C: Linguistic Behaviours and Attitudes

C1. How often do you use English in the following situations?

	Never	Seldom	Sometimes	Often	All the time
... where you live (e.g. coliving, shared flat)	☐	☐	☐	☐	☐
... where you work (e.g. coworking, coworking café)	☐	☐	☐	☐	☐
... when you text your friends or post on social media (e.g. Facebook, Instagram, Whatsapp)	☐	☐	☐	☐	☐

C2. How strongly do you agree or disagree with the following statements?

(1 – strongly disagree, 2 – disagree, 3 – neither agree nor disagree, 4 – agree, 5 – strongly agree)

	1	2	3	4	5
I like engaging with new languages, learning new words and ways of saying from my friends.	☐	☐	☐	☐	☐
I like learning the language of the places where I travel.	☐	☐	☐	☐	☐
In my social life, I would prefer to speak less English and more frequently other languages.	☐	☐	☐	☐	☐
I would like to know English better than I do now.	☐	☐	☐	☐	☐

C3. How strongly do you agree or disagree with the following statements?

(1 – strongly disagree, 2 – disagree, 3 – neither agree nor disagree, 4 – agree, 5 – strongly agree)

I would like to know English better...

	1	2	3	4	5
... to have better job opportunities.	☐	☐	☐	☐	☐
... to be more relaxed in social interactions.	☐	☐	☐	☐	☐
... to participate more in social events and leisure activities (e.g. meet-up events, social meals, yoga classes).	☐	☐	☐	☐	☐
... to get some knowledge about specific subjects (e.g. programming, web development).	☐	☐	☐	☐	☐

C4. How strongly do you agree or disagree with the following statements?

(1 – strongly disagree, 2 – disagree, 3 – neither agree nor disagree, 4 – agree, 5 – strongly agree)

	1	2	3	4	5
In some cases I find it easier to express myself in English than in any other language.	☐	☐	☐	☐	☐
Sometimes I have the feeling I am misunderstood when I speak English with people from other countries.	☐	☐	☐	☐	☐
In international social contexts, I prefer speaking English only.	☐	☐	☐	☐	☐
In international social contexts, I like using a mix of English and other languages.	☐	☐	☐	☐	☐
My high-school teacher would not approve of the English I speak in social contexts.	☐	☐	☐	☐	☐

C5. When you are at a social event where there are people from many different places speaking different languages (such as a community meet-up, a coworking dinner etc.) how often do you.. ?

	Never	Seldom	Sometimes	Often	All the time
... speak only with people who can speak your language(s)	☐	☐	☐	☐	☐
... speak only English, and only with people who can speak English	☐	☐	☐	☐	☐
... speak the language of the country you are in	☐	☐	☐	☐	☐
... speak your language(s) with people who speak (a) different language(s) and you try and understand each other (e.g. Italian and Spanish, Dutch and German)	☐	☐	☐	☐	☐
... mix all the languages you know and use all possible means to communicate effectively (e.g. mimic and gesture, images, technology help, automatic translation)	☐	☐	☐	☐	☐
... use apps for real-time translation (e.g. Google Translate)	☐	☐	☐	☐	☐
... use your smartphone to display images when you don't know a word, or your interlocutor does not understand you	☐	☐	☐	☐	☐

C6. Look at these examples taken from social networks:

'Vamos to the beach!'

'We propose you a real Italian aperitivo'

'Estão readys?!'

How often do you mix different languages like in the examples above... ?

	Never	Seldom	Sometimes	Often	All the time
... while socialising	☐	☐	☐	☐	☐
... at work	☐	☐	☐	☐	☐
... when writing	☐	☐	☐	☐	☐
... when speaking	☐	☐	☐	☐	☐

C7. Think about mixing languages, inventing words or using language in an unconventional way like in the examples above.

Are there any *strange* words or expressions that you use or have heard and would like to share with us?

☐

C8. How often do you use apps such as Google Translator or Chat GPT...

	Never	Seldom	Sometimes	Often	All the time
... while working or studying?	☐	☐	☐	☐	☐
... for social purposes? (e.g., while talking to people from other countries)	☐	☐	☐	☐	☐

C9. How strongly do you agree or disagree with the following statements?

(1 – strongly disagree, 2 – disagree, 3 – neither agree nor disagree, 4 – agree, 5 – strongly agree)

	1	2	3	4	5
Technologies like Google Translate and CHAT GPT are helping me improving my language skills.	☐	☐	☐	☐	☐
My generation communicate too much through technologies.	☐	☐	☐	☐	☐

LimeSurvey

	1	2	3	4	5
Social networks are a great means to stay in contact with friends who live far away, or meet new like-minded people.	☐	☐	☐	☐	☐

C10. Overall, did you like reading and answering our questions in English or would you have preferred another language?

- English was ok ☐
- I would have preferred another language (please tell us why) ☐

C11. You can add any comments about your communication habits or leave us feedback about the questionnaire here.

We really thank you for your engagement! Follow us on Instagram or read our blog to get updates about the project ?

Appendix 4. ***GC2_Circumstantial Annotation***[62]

EVENT 10 – Las Palmas de Gran Canaria. Community Happy-Hour on theBeach

Type of event: Happy-hour on the beach

Description: The event took place at Las Canteras beach, near the volleyball area, where other similar events usually take place. The event took place on the 15th of February 2024, from. 17.30 to 20.00 p.m. approximately. The event was organisedby the researcher and was posted on the Whatsapp group and on the Slack channel of the digital nomad community of Las Palmas de Gran Canaria, as usually done for similar events (see Figure 2), and on the Instagram Profile of the research project. Information about the research project was included in the invitation message. Figure 1 shows the flyer produced by the researcher to share within the digital social network infrastructures.

Figure 1. Flyer for the event

[62] https://osf.io/uk9ny/files/ka87w.

Figure 2. Other community happy-hour events in Las Palmas de Gran Canaria.

The participants were divided into two groups and the recordings were placed in the middle of a beach towel, where food and drinks were. The participants were asked to say their names and some general information about themselves at the beginning.
Despite undertaking a participatory appraoch to research, this time the researcher decided not to join the two groups. The voice of the researcher can be heard in the recordings only at the beginning and end of the activity.
All participants showed large interest in the research and asked to be kept updated about the results. All participants were very happy to meet new people, only two of them already knew each other. All participants were very talkative. At the end of the event, Eva joined the participants in recording n. 2.

Number of researchers who took part in the event: 1

Number of participants: 7

Number of recordings: 2

Recording n. 1 – Novella's phone (Researcher)

Speakers: K*****n, S*****a, E*a

Speakers' nationality: K*****n: Norwegian; S****a: Australian; E*a: Czech.

Speakers' languages: K*****n: Norwegian, English, German; S****a: English, Mandarin, Cantonese; E*a: Czech, English, French, Spanish.

Recording n. 2 – Giorgia's phone (Participant)

Speakers: V****a, P**l, K***a, F****h.

Speakers' nationality: V****a: Italian; P**l: English; K***a: Polish; F****h: Jam aican.

Speakers' languages: V****a: Italian, English, Spanish; P**l: English, Italian;K***a: English, Spanish, German, Polish; F****h: English.

Participant	Pseudonym	Code
K*****n	Anders	NO_C_01
S*****a	Sarah	AU_C_01
E*a	Karla	CZ_C_03
V*****a	Roberta	IT_C_03
P**l	John	GB_C_02
K***a	Natalia	PL_C_02
F****h	Bembe	JM_C_01

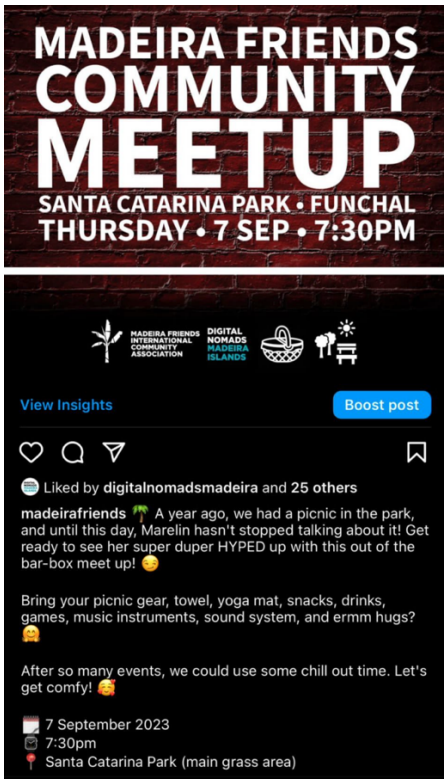
Appendix 5. ***Description of Madeira Islands data collection sessions***[63]

Description of Madeira Islands data collection sessions in the FLO project

M1: Community Picnic at Parque de Santa Catarina (October 23, 2023)

The first Madeiran recording session took place during a picnic social event held at Parque de Santa Catarina in Funchal. Co-organised by the researcher and *Madeira Friends*, a local community organisation supporting digital nomads, the event was modelled on a successful earlier gathering (see Figure 1) and publicly advertised through Instagram, Slack, and WhatsApp channels using the flyer shown in Figure 2. The atmosphere was relaxed and informal, with participants seated together on the grass sharing food, drinks, and early-evening conversation over approximately two and a half hours (19:00–21:30). Six participants attended, representing a highly international constellation: American, Italian, Romanian, German, Canadian, and Iranian backgrounds. Their plurilingual repertoires included combinations of English, Portuguese, Italian, German, Romanian, Farsi, and several other languages, making the event interactionally rich despite its small size.

The researcher's presence was fully integrated into the interactional ecology, as she both organised and participated in the event. Two audio recordings were produced, capturing the same group of speakers.



[63] https://osf.io/2qpdb

Figure 1. Instagram post for a meet-up event at the Santa Catarina Parque in Madeira.

Figure 2. Flyer advertising the picnic event for data collection in Madeira (M1).

This session was included in the first version of NomadLingo.

M2: Birthday Dinner Party at Câmara de Lobos (November 1, 2023)

This recording documents a private social gathering hosted in the home of a community member in Câmara de Lobos. The event, organised by Kami to celebrate her birthday, brought together eight guests including the researcher. Again, group composition reflected the demographic diversity typical of digital nomad communities: Italian, Romanian, Portuguese, British, and Taiwanese participants interacted across three recordings capturing different phases of the evening. The intimate, domestic setting fostered spontaneous translingual exchanges, shaped by high degrees of familiarity among participants.

Participants displayed expansive linguistic repertoires, including Italian, English, Spanish, Romanian, Danish, Portuguese, French, Russian, and Mandarin Chinese, all of which somehow permeated the interactions and provided fertile ground for translanguaging phenomena.

This session was included in the first version of NomadLingo.

M3: Porto Santo Coworkers' Lunch Break at Pingo Doce + old town tour (November 10 2023)

Recorded during a casual lunch break among coworkers at a Porto Santo supermarket café near the Coworking, this session captures a naturally occurring interaction embedded in the researcher's everyday routine during a nomad retreat. The conversation unfolded among six participants, including British, Portuguese, Lebanese, Polish, and Romanian speakers. Two

recordings were produced: the first during lunch, the second while participants informally toured the town centre, guided by a local retreat organiser.

The setting was dynamic and semi-public, with background noise typical of a supermarket environment. Two participants remained mostly silent and eventually left the group and niteractions were dominated by English. The session provides valuable insight into workplace-adjacent nomad interactions, especially the blending of professional and social talk.

This session was not selected to be included in the first version of the NomadLingo corpus, because of the limited number of participants, interaction time and low audio quality.

M4: Porto Santo Hike (November 12, 2023)

This session was recorded during a group hike on Porto Santo island. The recording device was carried in the pocket of a participant, resulting in highly ecologically valid but acoustically challenging data, with periods of silence, wind noise, and variable audibility. The hike involved fourteen participants from diverse backgrounds (British, Portuguese, Lebanese, Polish, Romanian, Estonian, French, Italian).

The mobile, outdoor setting produced fragmented but authentic samples of conversation emerging organically during movement, rest points, and shifting group configurations. Although not all participants spoke frequently, the recording documents how nomads navigate informal interaction while physically engaged in shared activities.

This session was not included in the first version of the NomadLingo corpus, because of its low audio quality.

M5: Porto Santo Community Lunch (November 12, 2023)

Following the hike, participants gathered for a community lunch at a large restaurant table. Fifteen participants were present, though only a subset of them are active speakers. The environment was crowded, producing a recording characterised by overlapping conversations and background voices.

Participants represented a broad range of nationalities – British, Portuguese, Lebanese, Polish, Romanian, Estonian, French, Italian – which shaped the character of the interaction. As with other Porto Santo events, English dominated but coexisted, this time, with Portuguese and Polish.

This session was not included in the first version of the NomadLingo corpus, because of its low audio quality and shortness of the recording.

M6: Community Lunch at Restaurante Gavino, Funchal (November 24, 2023)

This weekly community lunch, organised by a local community manager and promoted through the *MadeiraFriends* WhatsApp group, brought together approximately twenty digital nomads, though the recording focuses on a single subgroup of five participants seated near the researcher. Held in a busy city-centre restaurant, the session contains typical environmental noise, like voices, tableware clutter, and restaurant activity.

The recorded participants came from Russian, Dutch-American, German, Russian, and Italian backgrounds. Their repertoires include English, Russian, Portuguese, Spanish, French, German, Dutch, and Italian. The researcher was seated at the table but not centrally involved in the main conversations, although her voice can occasionally be heard. This session is valuable for documenting how small-group interaction unfolds within larger nomad social structures, however it was not selected to be included in the first version of the NomadLingo corpus because of low audio quality.

M7: Happy-hour at Coworking, Ponta do Sol (November 24, 2023)

This co-organised social event took place at the outdoor area of the coworking space at Centro Cultural John dos Passos. Advertised via Slack and WhatsApp through the flyer shown in Figure 4, the event brought together twelve participants. I prepared a variety of dishes and participants also brought food and drinks to share. Figure 3 shows the main table where the food and drinks were placed, alongside the booklets for collecting informed consent, privacy statements and background information about participants. Three tables allowed stable seating arrangements to improve recording consistency (as shown in Figure 3).

Figure 3. A picture of the initial moments of the data collection event at the Coworking in Ponta do Sol, Madeira.

Three recordings capture different subgroups with varying compositions and linguistic repertoires, including Portuguese, Italian, German, Norwegian, Lebanese, Serbian, Polish, Estonian, and Brazilian backgrounds. The researcher's presence was organisational and introductory but less central than in more intimate settings.

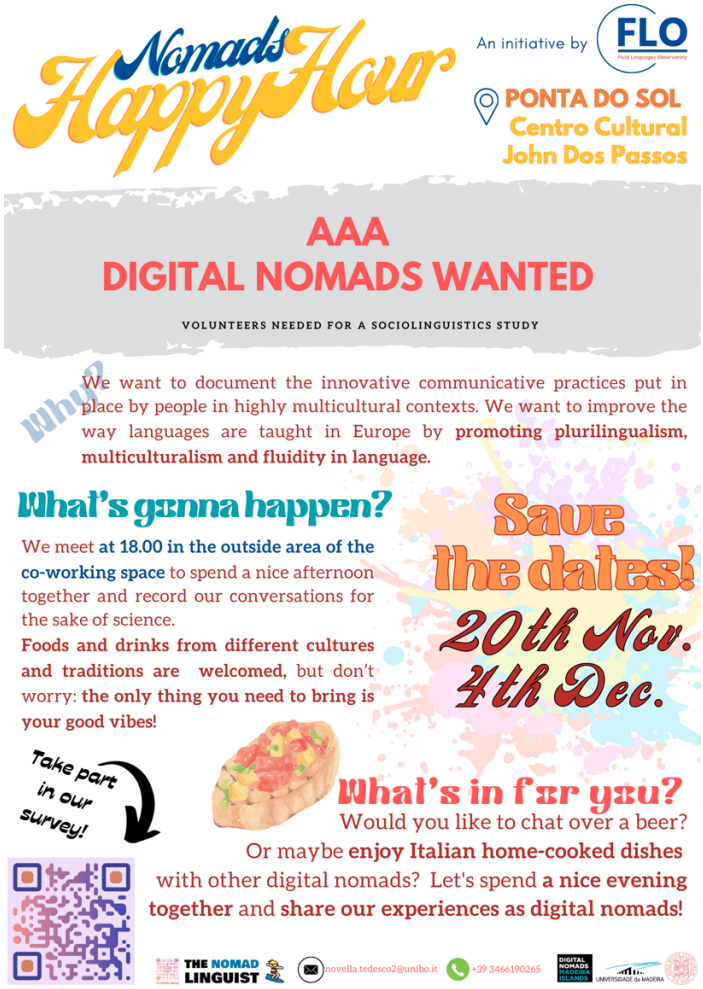


Figure 4. Flyer advertising the two happy-hour events organised at the Coworking in Ponta do Sol for data collection in Madeira (M7 and M8).

This session was included in the first version of the NomadLingo corpus resulting in three .txt files containing transcriptions of the three conversation groups.

M8: Happy-hour at Coworking, Ponta do Sol (December 3, 2023)

This second edition of the Happy-hour event took place at the same coworking venue but unfolded more informally than the November gathering. The event lasted from 18:00 to 21:30, with participants seated at two separate tables, generating two recordings capturing slightly different subgroup dynamics.

Active speakers included Italian, Estonian, Portuguese, Romanian, German, and Hungarian participants, with repertoires spanning Italian, Spanish, English, Estonian, Portuguese,

German, and Romanian. The researcher again co-organised the event, this time she participated naturally in the interactions. For the organisation and recording procedure in this event, the researcher was helped by two linguistics students, part of a group of students in phonetics and phonology who followed a cycle of seminars organised by the researcher at the University of Madeira.

This session is not currently included in NomadLingo but it represents an interesting addition to future versions of the corpus.

Appendix 6. ***Description of Canary Islands data collection sessions***[64]

Description of Canary Islands data collection sessions in the FLO project

GC1 – Carnival dinner at Kite & Surf Nomad House in Gran Canaria (February 9, 2024)

This session was recorded during a Carnival-themed dinner organised jointly by the researcher and the manager of the coliving *Kite & Surf Nomad House* in Las Palmas de Gran Canaria. The event, held on 9 February 2024 from 20:30 to approximately 22:30, took place on the rooftop terrace – an established social hub for residents who routinely gather there for dinners, informal happy-hours, and parties. As with other community events, the invitation was circulated through the coliving WhatsApp group, accompanied by information about the research project. Two tables were arranged to encourage stable seating configurations, although in the final part of the evening several participants gathered around the pool table to watch a game. At this point, one of the two recording devices was placed on one side of the pool table (see Figure 1).

Figure 1. Data collection event in Las Palmas de Gran Canaria.

At the beginning of the event, participants introduced themselves and shared basic information about their linguistic repertoires and nomadic trajectories; this activity was enthusiastically received and contributed to the creation of a relaxed yet engaged interactional environment, besides allowing for the collection of background information about speakers which was later used to include relevant metadata in the corpus (see Section 5.2). The dinner was cooked by the researcher and drinks were brought by the participants. Throughout the dinner, two partially overlapping recordings captured the conversations emerging within two subgroups. The first recording includes participants from Argentina, Sweden, the Czech Republic, Germany, Italy, and Spain, whose repertoires encompass combinations of Spanish, English, Swedish, Czech,

[64] https://osf.io/8tsbv.

German, French, Finnish, and Italian. The second recording documents a smaller group composed of Swedish, Czech, German, and Spanish speakers.

GC2 – Community Happy-Hour on the Beach at Las Cantares, Gran Canaria (February 15, 2024)

This event was a community happy hour held on Las Canteras beach. Organised by the researcher, it took place on February 15th, 2024 from 17:30 to approximately 20:00. The invitation was posted on WhatsApp and Slack community channels, as well as on the Instagram profile of the research project (see Figure 2). The message included details about the study and encouraged voluntary participation.

Figure 2. The advertisement flyer for the community happy-hour organised at the beach in Las Palmas de Gran Canaria.

Participants formed two small groups, each seated on beach towels around shared food and drinks. Recording devices were placed at the centre of the towels to capture the interactions. As in other sessions, the event began with a brief round of introductions, allowing participants to express their linguistic repertoires and personal backgrounds. At this event, the researcher opted not to sit with either group, both to ensure more natural conversational dynamics and to facilitate unobtrusive audio capture.

The first recording includes Norwegian, Australian, and Czech speakers with multilingual repertoires spanning Norwegian, English, German, Mandarin, Cantonese, French, and Spanish. The second recording captures Italian, English, Polish, and Jamaican speakers, whose repertoires include English, Italian, Spanish, German, and Polish. The event was characterised by high levels of engagement: most participants did not know each other beforehand and expressed enthusiasm for meeting new people and for the research aims of the project.

GC3 – Dinner at a nomad house in Las Palmas de Gran Canaria (February 23, 2024)

This session was recorded during a community dinner hosted by the researcher and her flatmates in their shared nomad apartment near Las Canteras. Taking place on 23 February 2024 from 20:30 to approximately midnight, the event was a typical household dinner in the nomad community, with the addition of four external guests selected through a first-come-first-served system on Slack. Such open invitations mirror common community practices in Las Palmas, as illustrated by screenshots included in Figure 3.

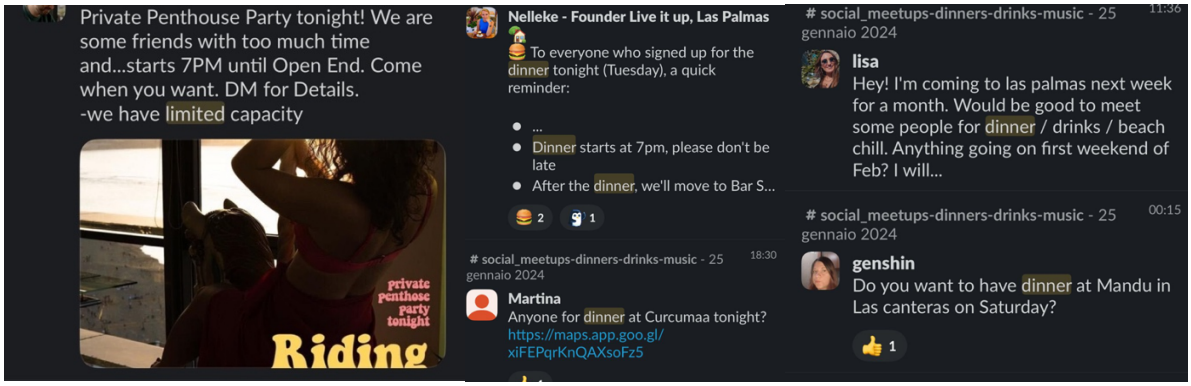


Figure 3. Screenshots of dinner event advertised on the slack channel of the digital nomad community in Gran Canaria.

The dinner unfolded in a small living room, which limited the possibility of creating two stable groups for recording. Although two recording devices were used, the conversations overlap due to spatial proximity. The recordings are clear, with minimal background noise, making them suitable for detailed transcription.

As in the previous events for data collection held in Gran Canaria, the event began with a round of introductions, a practice well received by participants and useful for capturing information about linguistic repertoires and mobility trajectories. Conversations developed dynamically and were rich in translanguaging practices, given the diverse linguistic backgrounds present: Swedish, English, German, Croatian, and Italian, with repertoires including English, Swedish, German, French, Spanish, Dutch, Italian, Hungarian, and more. The researcher and her partner Antonio participated actively, reflecting the interactional intimacy of the setting.

Participants expressed strong appreciation for the event (as shown in Figure 4), which reproduced the conviviality and social cohesion characteristics of nomad shared living spaces.

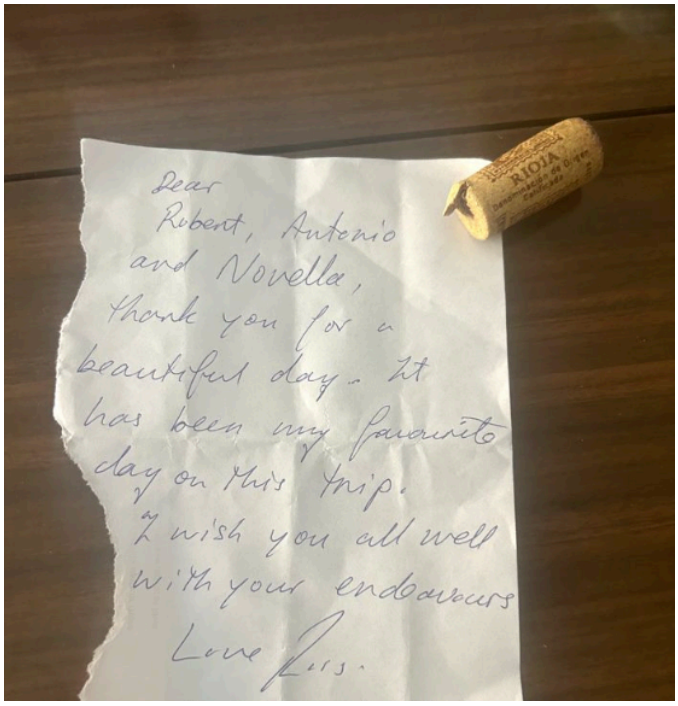


Figure 4. A note left by a participant to the data collection dinner event in Las Palmas de Gran Canaria.

F1 – Community Dinner in Fuerteventura (April 5, 2024)

This event took place in a private flat in the centre of Corralejo on 5 April 2024, from 20:30 to approximately midnight. Organised collaboratively by the researcher and members of the local digital nomad community, the dinner was advertised through the closed WhatsApp group *Cenas y cavernículos*, which includes over 150 members. The setting – a small, open-plan living space (shown in Figure 5) – created a lively interactional environment in which participants frequently moved around, making fixed seating arrangements difficult and occasionally affecting audio quality.

Figure 5. A picture taken at the data collection dinner event in Fuerteventura.

Two recording devices captured the conversations; although some segments overlap, the absence of external noise and the relatively intimate setting ensured that both recordings remain fully interpretable. The researcher and her partner actively participated throughout.

As in previous events, participants introduced themselves and described their linguistic repertoires at the beginning. The session includes a diverse group of Argentinian, Polish, Spanish, English/French, French, and Italian participants, with repertoires spanning Italian, English, Spanish, Catalan, German, French, Mandarin Chinese, Sardo, Portuguese, and Greek. This is the session where less English is spoken in comparison to all the other (an analysis is provided in Chapter 6).

F2 – Nomad chats, brainstorming event at Surfin' Colors Coworking in Corralejo, Fuerteventura (April 8, 2024)

This session was recorded at Surfin' Colors, one of the most active coliving / coworking hubs in northern Fuerteventura. Co-organised by the researcher and the coworking event manager, and advertised on the coworking Whatsapp group (see Figure 6), tit was held on 8 April 2024 from 18:00 to 20:00. The setting – the communal kitchen area – was well suited to informal interaction and allowed for a clear high quality audio recording. Participation was slightly lower than expected due to logistical challenges that required rescheduling the event.

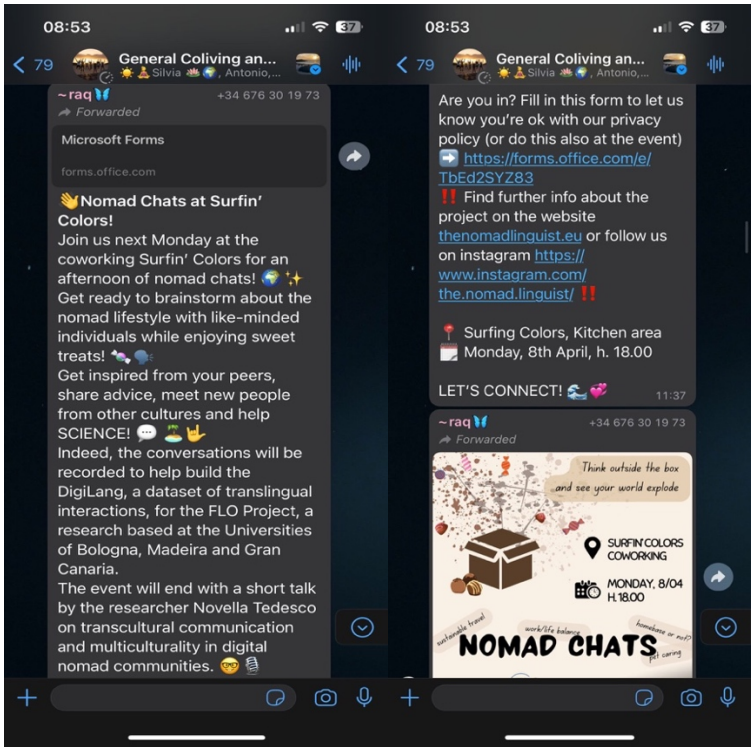


Figure 6. Screenshots of messages advertising the Nomad Chats event for data collection on the Surfin' Colors Whatsapp Group.

After the initial round of self-introductions, participants were given boxes containing sweets with hidden topic cards, each prompting reflections on aspects of nomadic life or themes

identified through prior ethnographic observation (as shown in Figure 7). This structured activity simulated other brainstorming and mastermind events often occurring at digital nomad communities (see, for example, Figure 8)

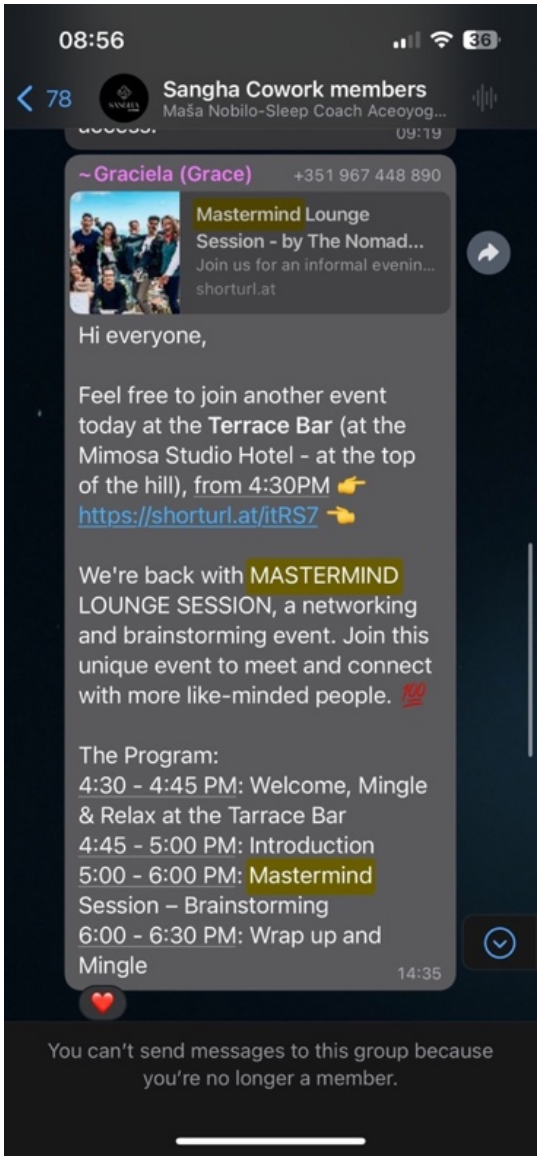


Figure 7. Example of a mastermind event at a coworking in Madeira.

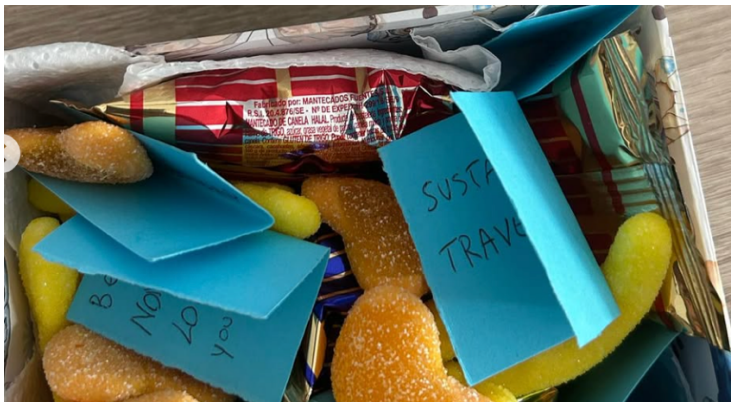


Figure 8. The box used in the Nomad Chats event for data collection in Corralejo, Fuerteventura.

The session involved French, Spanish, Italian, German, and Estonian speakers whose repertoires included a wide range of European languages as well as Portuguese and Russian. The event concluded with a short workshop led by the researcher on inclusive and plural approaches to transcultural communication, as a form of knowledge pay-back to the participants. After the event, the group continued their conversation while going out to dinner, and these final interactions, including the researcher's voice, are captured in the last segment of the recording.

Appendix 7. *NomadLingo Raw Data_Metadata on data collection settings and participants*[65]

nickname	code	languages	gender	LanguagesSpoken	Nationality
Anastasia	RU_M_01	fra|rus|eng|por|spa	Female	5	Russian
Emma	SE_C_01	swe|eng|spa	Female	3	Swedish
Dorin	RO_M_03	ron|eng	Male	2	Romanian
Alexei	GE_M_03	deu|eng|spa|por|rus	Male	5	German
Antea	HR_C_01	eng|hrv|spa|deu|ita|hun	Female	6	Croatian
Adela	CZ_C_01	eng|ces|spa	Female	3	Czech
Adele	IT_C_02	eng|ita|spa	Female	3	Italian
Marco	IT_MC_01	ita|eng|spa	Male	3	Italian
Dan	NL_M_01	eng|nld|deu|ita	Male	4	Dutch / American
Dol	RO_M_04	hun|ron|eng		3	romanian
Chloe	FR_C_01	fra|spa|eng	Female	3	French
Aksel	NO_M_01	nor|eng|por	Male	3	Norwegian
Sophia	SE_C_02	swe|eng|fra|fin	Female	4	Swedish
Louise	GB_C_01	eng|fra|spa|cmn|ita|cat	Female	6	British / French
Martin	CZ_C_02	ces|eng	Male	2	Czech
Finley	GB_M_01	eng|hin	Male	2	British
Viktor	GE_M_02	rus|eng|deu	Male	3	German
Mila	RO_M_02	ron|eng|spa	Female	3	Romanian
Agata	RU_M_02	rus|eng	Female	2	Russian
Karla	CZ_C_03	ces|eng|fra|spa	Female	4	Czech
Francesca	IT_M_02	eng|ita|por	Female	3	Italian
Bembe	JM_C_01	eng	Male	1	Jamaican
Tibor	HU_M_01	hun|eng	Male	2	hungarian
Daniela	PT_M_01	por|eng	Female	2	Portuguese
Ana	BR_M_01	por|eng|spa	Female	3	Brazilian
Luigi	IT_C_01	ita|spa|eng|srd|por|ell	Male	6	Italian
Gloria	IT_M_01	ita|spa|eng	Female	3	Italian
Carlos	ES_C_04	spa|eng|fra	Female	3	Spanish
Mary	AM_M_01	eng|por	Female	2	North American
Paulo	PT_M_02	por|eng|spa|deu	Male	4	Portuguese
Agnes	SE_C_04	eng|swe|deu|spa|fra	Female	5	Swedish
Ana	RO_M_01	eng|dan|ron	Female	3	Romanian
Eva	SE_C_03	swe|eng|spa	Female	3	Swedish
Natalia	PL_C_02	pol|eng|deu|spa	Female	4	polish

[65] https://osf.io/82nzk.

Emilia	GE_C_03	deu|eng|spa	Female	3	German
Elina	EE_C_01	eng|est|spa|fra|ita|por|rus	Female	7	Estonian
Anders	NO_C_01	nor|eng|deu	Male	3	Norwegian
Lea	GE_C_01	deu|eng|spa	Female	3	German
Maria	ES_C_03	spa|fra|eng	Female	3	Spanish
Olly	TW_M_01	cmn|eng	Female	2	Taiwan
Pola	PL_M_01	pol|eng|spa|ita	Female	4	polish
Arnau	ES_C_02	cat|spa|eng|ita|deu|fra	Female	6	Spanish
Amalia	PT_M_04	por|eng|fra	Female	3	Portuguese
Eliise	EE_M_01	est|eng|deu	Female	3	Estonian
Luka	GE_M_01	deu|eng	Male	2	German
Maja	PL_C_01	pol|eng	Female	2	Polish
Adi	GE_M_05	deu|eng	Male	2	german
Samuel	PT_M_03	eng|spa|por	Male	3	Portuguese
Nora	LB_M_01	eng|deu	Female	2	Lebanese
Lorella	RESEARCHER_IT	ita|eng|fra|spa|deu|por|nap	Female	7	Italian
Dinu	RO_M_05	ron|eng	Male	2	Romanian
Felipe	ES_C_01	eng|spa|fra	Male	3	Spanish
John	GB_C_02	eng|ita	Male	2	British
Bogdan	RS_M_01	srp|eng|por|ita	Male	4	Serbian
Alain	FR_C_02	fra|eng|spa	Male	3	French
Dragos	RO_M_06	ron|eng	Male	2	romanian
Gertrud	GE_C_05	deu|eng|ita|spa|fra	Female	5	German
Oliver	SE_C_05	swe|eng|spa	Male	3	Swedish
Rita	PT_M_05	por|eng|fra	Female	3	portuguese
Greg	GB_C_03	eng|fra|deu|spa|nld	Male	5	British
Sarah	AU_C_01	eng|yue|cmn	Female	3	Australian
Ignacio	AR_C_02	spa|eng	Male	2	argentinian
Luisa	IT_M_04	eng|ita|deu|spa|fra	Female	5	Italian
Lotte	GE_C_02	deu|eng|fra	Female	3	german
Marko	GE_C_04	deu|eng|fra|spa	Male	4	German
Mateo	AR_C_01	spa|eng|fra|deu|ita|cat	Male	6	argentinian
Roberta	IT_C_03	ita|eng|spa	Female	3	Italian
Wanda	IT_M_03	ita|eng|rus|spa	Female	4	italian
Khalid	CA_M_01	urd|blx|eng	Female	3	Canadian
Eike	GE_M_04	deu|eng	Male	2	german
Leila	IR_M_01	fas|eng|swe	Female	3	Iranian

Appendix 8. *NomadLingo1.1_Data_Metadata on data collection settings and participants*[66]

Worksheet 1 – Sessions

id	title	languages	Working Languages	genre	subgenre (interaction_type)	date	location	researcher involvement	region	country	continent	planning	context	audio quality	recording devices	recording overlap	participants	description
M1	Picnic in Funchal	TRL	ENG	Conversation	free	22/10/23	Funchal, Madeira	silent observer	Madeira	Portugal	Europe	planned	public	High	2	yes	RESEARCHER_IT|AM_M_01|CA_M_01|GE_M_01|IR_M_01|IT_MC_01|RO_M_01	The event took place at Parco de Santa Caterina in Funchal (Madeira) on the 23rd of October 2023 from 7.00 p.m. to 9.30 p.m. approximately. The event was organised by the researcher for the purposes of the data collection, and by Madeira Friends, an organisation which is active on the territory of Madeira together with NomadX and StartUp Madeira, to enhance services and facilities, organise and sponsor social events and leisure activities for the digital nomad community in different locations around the island. In particular, the idea of the picnic was proposed by one of Madeira Friends managers, who advised the research on a type of event which, according to him, would give the best results in terms of sound quality, participation and organisational effort. The event was organised taking as a model a previous event of the same kind organised by Madeira Friends in September 2023 and was advertised on the Instagram channels of the communities as well as on the Instagram page of the Nomad Linguist. It was mentioned in Funchal weekly schedule on the Slack channel, and it was advertised on the Slack chats and Whatsapp chats where these kinds of activities are usually organised.
M2	Birthday Party in Camera do Lobos	TRL	ENG	Conversation	free	31/10/23	Camera de Lobos, Madeira	participant	Madeira	Portugal	Europe	semi-spontaneous	private	Medium to poor	3	yes	RESEARCHER_IT|IT_MC_01|PT_M_04|PT_M_05|GB_M_01|IT_M_03|IT_M_02|TW_M_01	The event took place in Camara de Lobos (Madeira) on the 1st of November 2023, at the house of PT_M_04, a Madeirian-born member of the community who moved back to the island from Portuguai mainland since she had the possibility to work remotely. The event was organised by RO_M_01, who wanted to celebrate her birthday. Everybody was sitting at the same table, so the two recording devices were used to provide enhanced audio quality of the same conversation. As often happens in these type of events, participants moved a lot across the room, went out on the terrace, left the table, change places, and many times conversations overlap. Nevertheless, the cultural and linguistic richness of the conversation can be well understood from the audio files and, despite some inaudible portion, the file was considered of a good enough quality to be transcribed.
M7	Happy Hour at Coworking in Ponta Do Sol 1	TRL	ENG	Conversation	free	23/11/23	Ponta do Sol, Madeira	no-observer	Madeira	Portugal	Europe	planned	controlled environment	medium	3	no	M7_1: BR_M_01|PT_M_03|LB_M_01|PL_M_01|EE_M_01 M7_2: RESEARCHER_IT|IT_MC_01|GE_M_02|NO_M_01|PT_M_01|PT_M_02|PT_M_03|LB_M_01|RS_M_01|EE_M_01|GE_M_04 M7_3: BR_M_01|RESEARCHER_IT|IT_MC_01|NO_M_01|PT_M_02|PT_M_01|LB_M_01|RS_M_01|EE_M_01|PL_M_01	The event took place at the coworking Centro Cultural John Dos Passos in Ponta do Sol (Madeira) on the 24th of November 2023 from 6.00 p.m. to 9.30 p.m. approximately. The event was organised by the researcher and the community manager together with NomadX and StartUp Madeira. The event was organised taking as a model previous events. It was mentioned in Ponta Do Sol weekly schedule on the Slack channel, and it was advertised on the Slack chats and Whatsapp chats where these kinds of activities are usually organised. A big table was at the centre of the outside area of the coworking, with food and drinks, but participants were asked to sit at one of the three tables which had been previously organised by the community manager, and they were asked to keep their seats to improve audio quality.
GC 1	Dinner at Coliving Kite & Surf in Las Palmas	TRL	ENG	Conversation	free	08/01/24	Las Palmas de Gran Canaria	silent observer	Canary Islands	Spain	Europe	planned	controlled environment	Medium to high	2	no	GC1_1: RESEARCHER_IT|IT_MC_01|AR_C_02|SE_C_01|SE_C_02|CZ_C_01|GE_C_02|ES_C_03|CZ_C_02|GE_C_03|ES_C_04 GC1_2: SE_C_03|RESEARCHER_IT|IT_MC_01|CZ_C_02|GE_C_03|ES_C_05|ES_C_04	The event took place on the rooftop of the coliving Kite & Surf Nomad House, where the people living at the coliving usually gathered for happy-hours, dinners and parties. The event took place on the 9th of February 2024, from. 20.30 to 10.30 p.m. approximately. The event was organised by the researcher and the community manager and was posted on the Whatsapp group of the Coliving, as usually done for similar events. Information about the research were given alongside the invitation message. Two tables were prepared and the participants were asked to choose a table and stick with their own seats. Despite this, towards the end of the event some participants decided to play a pool game, and many moved around the pool table to watch the game. Therefore one of the recording devices was placed at the pool table in the final part of the event.

[66] https://osf.io/4x3s2.

																		The participants were asked to say their names and some general information about themselves at the beginning. This small activity, which is helpful for keeping information about participants related to their linguistic background and their lifestyle as digital nomads, was warmly welcomed by the participants who enjoyed it very much. It should be noted that IT_MC_01, who has been a participant to many events during data collection in Madeira, was not part of the recordings at first and only at the end took part in the conversation after being asked some questions by other participants, the same happened for the researcher. Despite undertaking a participatory approach to research, the researcher and her closest acquaintance decided to eat at a separate table not to overrepresent their voices in the data. However, when interaction was prompted by other participants they both felt it would prevent data naturalness not to answer.
F1	Community dinner in Fuerteventura	TRL	ENG	Conversation	free	04/04/24	Corralejo, Fuerteventura	silent observer	Canary Islands	Spain	Europe	spontaneous	private	medium to poor	2	yes	RESEARCHER_IT|IT_MC_01|IT_C_01|IT_C_02|FR_C_02|AR_C_01|PL_C_01|ES_C_02	The event took place in a flat in the centre of Corralejo rented by one of the members of the community, well acquainted with the researcher. The researcher helped the organisation and cooked for everybody. Because of the room size and architecture, people hardly kept their places , which affected recording quality. The dinner organisation was also posted on one of the community informal Whatsapp groups, called 'Canas y cavernículos', counting more than 150 members. The event took place on the 5th of April 2024, from. 20.30 to 12.00 p.m. approximately. This event has seen the active participation of the researcher, RESEARCHER_IT, and her partner, IT_MC_01, in the conversations. The conversations were recorded by two recording devices but sometimes it was the same conversation. As a matter of fact, due to spatial limitations, it was not possible to divide the participants into two settled groups. Despite this, the absence of external noise and the small location where the event took place make the recording audible and in cases of crossed conversations both wave files seem of a good enough quality to be transcribed. The participants were asked to say their names and some general information about themselves at the beginning. This small activity, which is helpful for keeping information about participants related to their linguistic background and their lifestyle as digital nomads, was warmly welcomed by the participants who enjoyed it much.
F2	Nomad Chats	TRL	ENG	Conversation	task-based	07/04/24	Corralejo, Fuerteventura	participant	Canary Islands	Spain	Europe	planned	controlled environment	Medium to high	2	yes	RESEARCHER_IT|IT_MC_01|GE_C_01|ES_C_01|EE_C_01|FR_C_01	The event took place in the kitchen area of the coliving and coworking Surfin' Colors, in Corralejo, Fuerteventura. It was organised by the researcher and the coworking event manager, inspired by other similar events. The event took place on the 8th of April from 18.00 to 20.00. At the end of the event a 20 minutes workshop about plural and inclusive approaches to transcultural communication was held by the researcher. The participants were asked to say their names and some general information about themselves at the beginning. Then, they were given boxes with some sweet treats with topic cards . In turn, participants picked a card and started talking about one of the topics, which are all related to nomad lifestyle or had been observed during ethnographic fieldwork as key conversational topics in brainstorming and workshops organised at coworkings. The event was particularly successful from the point of view of the participants, but it did not reach as many people as expected because some logistic issues caused event reschedule. At the end of the event, the participants and some other people from the coworking went to eat together (as can be heard in the last bits of the recording, in which also the researcher, RESEARCHER_IT, takes a more active role in the conversation).

Worksheet 2 – Participants

Code	Nickname	Languages	Gender	No. of languages spoken	Nationality	Sessions

AM_M_01	Mary	eng|por	Female	2	North American	M1
AR_C_01	Mateo	spa|eng|fra|deu|ita|cat	Male	6	argentinian	F1
AR_C_02	Ignacio	spa|eng	Male	2	argentinian	GC1_1
BR_M_01	Ana	por|eng|spa	Female	3	Brazilian	M7_1, M7_3
CA_M_01	Khalid	urd|blx|eng	Female	3	Canadian	M1
CZ_C_01	Petra	eng|ces|spa	Female	3	Czech	GC1_1
CZ_C_02	Martin	ces|eng	Male	2	Czech	GC1_1, GC1_2
EE_C_01	Elina	eng|est|spa|fra|ita|por|rus	Female	7	Estonian	F2
EE_M_01	Eliise	est|eng|deu	Female	3	Estonian	M7_1, M7_2, M7_3
ES_C_01	Felipe	eng|spa|fra	Male	3	Spanish	F2
ES_C_02	Arnau	cat|spa|eng|ita|deu|fra	Female	6	Spanish	F1
ES_C_03	Maria	spa|fra|eng	Female	3	Spanish	GC1_1, GC1_2
ES_C_04	Carlos	spa|eng|fra	Female	3	Spanish	GC1_1, GC1_2
FR_C_01	Chloe	fra|spa|eng	Female	3	French	F2
FR_C_02	Alain	fra|eng|spa	Male	3	French	F1
GB_C_01	Louise	eng|fra|spa|cmn|ita|cat	Female	6	British|French	F1
GB_M_01	Finley	eng|hin	Male	2	British	M2
GE_C_01	Lea	deu|eng|spa	Female	3	German	F2
GE_C_02	Lotte	deu|eng|fra	Female	3	german	GC1_1
GE_C_03	Emilia	deu|eng|spa	Female	3	German	GC1_1, GC1_2
GE_M_01	Luka	deu|eng	Male	2	German	M1
GE_M_03	Alexei	deu|eng|spa|por|rus	Male	5	German	M7_2
GE_M_04	Eike	deu|eng	Male	2	german	M7_2
IR_M_01	Leila	fas|eng|swe	Female	3	Iranian	M1
IT_C_01	Luigi	ita|spa|eng|srd|por|ell	Male	6	Italian	F1
IT_C_02	Adele	eng|ita|spa	Female	3	Italian	F1
IT_M_02	Francesca	eng|ita|por	Female	3	Italian	M2
IT_M_03	Wanda	ita|eng|rus|spa	Female	4	italian	M2
IT_MC_01	Marco	ita|eng|spa	Male	3	Italian	M1, M2, M7_2, M7_3, GC1_1, GC1_2, F1, F2
LB_M_01	Nora	eng|deu	Female	2	Lebanese	M7_1, M7_2, M7_3
NO_M_01	Aksel	nor|eng|por	Male	3	Norwegian	M7_3
PL_C_01	Maja	pol|eng	Female	2	Polish	F1
PL_M_01	Pola	pol|eng|spa|ita	Female	4	polish	M7_1, M7_3
PT_M_01	Daniela	por|eng	Female	2	Portuguese	M7_2, M7_3

PT_M_02	Paulo	por|eng|spa|deu	Male	4	Portuguese	M7_2, M7_3
PT_M_03	Samuel	eng|spa|por	Male	3	Portuguese	M7_1, M7_2
PT_M_04	Amalia	por|eng|fra	Female	3	Portuguese	M2
PT_M_05	Rita	por|eng|fra	Female	3	portuguese	M2
RESEARCHER _IT	Lorella	ita|eng|fra|spa|deu|por |nap	Female	7	Italian	M1, M2, M7_2, M7_3, GC1_1, GC1_2, F1, F2
RO_M_01	Ana	eng|dan|ron	Female	3	Romanian	M1, M2
RS_M_01	Bogdan	srp|eng|por|ita	Male	4	Serbian	M7_2, M7_3
SE_C_01	Emma	swe|eng|spa	Female	3	Swedish	GC1_1
SE_C_02	Sophia	swe|eng|fra|fin	Female	4	Swedish	GC1_1
SE_C_03	Eva	swe|eng|spa	Female	3	Swedish	GC1_1, GC1_2
TW_M_01	Olly	cmn|eng	Female	2	Taiwan	M2
VOLUNTEER_1		Female		Honduras		GC1_1
VOLUNTEER_2		Female		Argentinian		GC1_1

Appendix 9. *Annotation scheme for the NomadLingo Corpus*[67]

Level	Annotation type	Element	Attribute	Value	Description
zero	contextual	session	code	*e.g. F1*	identification code for each data collection session
			context	private / controlled_environemnt / public	private means that data were collected in a private house; semi-private means that data were collected in a space that is not private but with restricted access, like a coworking; public means that data were recorded in a public space like a restaurant
			interaction	free / task-based	free means that the conversation is completely free, whereas task-based refer to cases in which the conversation is carried out for a purpose, such as discussing a specific topic, finding a solution to a problem etc.
			languages	*e.g. German,English,Spanish,Italian,Portuguese,French,Neapolitan,Polish,Catalan,Greek,Sardinian,Mandarin*	all languages known by speakers in session
			location	*e.g. Corralejo*	the location where the data were collected
			nationalities	*e.g. 5*	number of nationalities of speakers in session,
			planning	spontaneous / semi-spontaneous / planned	spontaneous means that the data collection session had not been organised by the researcher; semi-spontaneous means that the data collection session had been planned for an event not organised by the researcher; planned means that the event for the data collection session had been organised by the researcher
			region	Canary Islands / Madeira Islands	region where data were collected
			speakers	*e.g. 8*	number of speakers in a session
			speakers_intro	yes / no	whether speakers introduce themselves within the conversation
			title	*e.g. Community dinner in Fuerteventura*	an exeplanatory title

[67] https://osf.io/rb8uw.

one	structural	conversation			contains transcript of the recorded conversation
two	linguistic	misunderstanding	type	*solved / unsolved*	contains a misunderstanding sequence, which can be solved when the involved speakers all seem to agree on a certain meaning, or unsolved, if speakers does not seem to come to an understanding of each other in relation to the linguistic expression which caused the misunderstanding
two/three	structural	turn	ID	*e.g. F1_0181*	identification code for each turn
			length	*e.g. 5*	number of tokens in a turn
three	contextual	intro			contains one or more turns where speakers introduce themselves
three/four	contextual	sound			contains annotation about a non-verbal sound, or inaudible speech portions
three/four/five	contextual	speaker	code	*e.g. PL_C_01*	identification code for speakers
			language	*e.g. French,English,Spanish*	languages spoken by speaker
			nationality	*e.g. French*	speaker nationality
four	linguistic	interactional_move	type	asking / checking / confirming	a portion of a turn or a turn where speakers ask for clarification, or check that their interlocutor(s) understand, or confirm that they understand what their interlocutor(s) is/are saying
four	linguistic	translanguaging	type	codemixing / codeswitching / translation	codemixing identifies a portion of a turn or a turn where words pertaining to two or more languages that are traditionally perceived as distinct; codeswitching identifies a turn or a turn where a language other than English is used; translation identifies a turn where a speaker is translating something that has been said earlier in another language.
four	linguistic	repair	type	reformulation / restart	reformulation identifies a portion of a turn where speakers express the same meaning twice using different phrases; restart identifies a particular type of reformulation happening at the beginning of a sentence (n.b. not at the beginning of a turn)

Appendix 10. ***NomadLingo1.1 open readme.txt***[68]

------Read me file NomadLingo1.1------

Contents

(1)General description

(2)Information about the research project

(3)Information about data collection, privacy regulation and license

(4)Information about the corpus

(5)Transcription procedure and conventions

(6)Encoding

(7)Metadata

(8)Annotation schema

(9)Annotation procedures

(10)Linguistic Annotation

(11)License and use

(12)References

(13)Acknowledgements

1.General description

NomadLingo is the first publicly available corpus documenting multilingual, naturally occurring interactions among European digital nomads, a rapidly growing yet understudied transnational community (Tedesco 2025). The corpus contains transcripts of extracts from naturally-occurring conversations which were audio-recorded between November 2023 and April 2024 at social events organised and promoted within digital nomad communities based in Madeira and Canary Islands. The total time of transcribed recording is 11 hours 38 mins. For further information about the texts in the corpus see Section 4. The version 1.1 open is an updated version of

[68] https://osf.io/83tmy.

the older NomadLingo1.0 open (Tedesco et al. 2025). In this version two annotation layers have been added (see Section 8) and transcripts were further revised.

The corpus aims to represent translingual interactions based on the fluid use of English as a lingua franca, other languages and linguae francae such as Spanish and Portuguese, and strategies of transcultural communication like intercomprehension and peer/self-translation.

The folder NomadLingo1.1 open contains:

- this readme file;

- a folder with the three annotated versions of the corpus (Annotated_versions_NomadLingo1.1), including Trans&repair&misunderstanding&interactionAnnotated_NomadLingo1.1, Trans&repairAnnotated_NomadLingo1.1, and TranslanguagingAnnotated_NomadLingo1.1;

- a folder named Naked_NomadLingo1.1 containing files which only include transcribed conversation, structural and contextual annotation but no linguistic annotation;

- a folder named MetadataNomadLingo1.1 containing information about speakers and sessions in the corpus in .xls and .csv formats;

- a folder named Annotation_schema_NomadLingo1.1 containing the annotation scheme in .xls and .csv formats;

- the corpus in .csv format.

Moreover, another version of the corpus, for which access is restricted for privacy reasons, includes two folders, namely Recordings, containing the .wav files, and Privacy Notice and Informed Consent documents, which contains legal and ethical documentation. To get access to the integral version of the dataset you can contact novella.tedesco2@unibo.it.

You can find the project on GitHub (https://github.com/novella-tedesco/FLO) and OSF (DOI 10.17605/OSF.IO/UK9NY), where codes for data processing and analysis are shared and updated.

2.Information about the research project

This dataset is part of a larger research project named FLO, the European Fluid Languages Observatory, which investigates transcultural communicative strategies characterizing informal communication in the European communities of digital nomads. The data have been collected in the framework of a Ph.D. research at the Department of Interpreting and Translation at the University of Bologna, by Novella Tedesco, supervised by prof. Silvia Bernardini and prof. Cristiana Cervini (https://www.thenomadlinguist.eu/flo/).

The general approach to research undertaken in the FLO project is ethnographic and participatory (Tedesco 2025). For this reason, notable features of NomadLingo1.0 include:

- the presence of the main researcher in the data;

- participants actively involved in the organisation of data collection procedures;

- variable audio quality;

- rich contextual information (see Section 7 and Section 8).

3.Information about data collection, privacy regulation and license

Data were collected with the researcher's personal device and participants' devices. The devices in question are smartphones. The software Mootiv Audio was used to capture audio in .wav format. All participants were asked to fill in a questionnaire, the informed consent and to sign the privacy policy. The transcripts are fully pseudoanonymised. The audio are restricted because they contain sensitive and personal data. To require access to the audio files you can contact Novella Tedesco (novella.tedesco2@unibo.it)

4.Information about the corpus

Number of tokens: 82,897

Number of texts: 9

Number of data collection sessions: 6

Number of transcribed recordings: 9

Number of speakers: 45 DIGITAL NOMADS (+2 HOSTEL VOLUNTEER)

Each text is a transcription of conversations recorded at some of the events organised to collect data within the FLO project. Complete information about all the project sessions can be found in the file NomadLingo_sessions.csv. Each session is marked with a unique identifier (e.g., M1, M2 etc.).

The corpus is available in five different versions:

- Trans&repair&misunderstanding&interactionAnnotated_ NomadLingo1.1 contains texts and annotation about misunderstanding sequences, interactional moves,translanguaging practices and repair strategies (see Sections 8 and 9);

- Trans&repairAnnotated_NomadLingo1.1 features texts and annotation about translanguaging practices and repair strategies (see Sections 8 and 9);

- Trans&repairannotated_NomadLingo1.1 includes texts where only translanguaging phenomena are annotated (see Sections 8 and 9);

- Naked_NomadLingo1.0 contains texts with metadata only related to speakers and sessions, but no annotation regarding linguistic and pragmatic phenomena (see Sections 7 and 8);

- NomadLingo1.1.csv, a single file which contains all metadata and annotation in separated columns from the transcripts. The transcripts without any type of annotation can therefore be extracted by copy-pasting the content of column 10 into a txt file.

The corpus contains 9 texts, with the following characteristics:

- M1:

Original audio length: 1h57mins

Transcribed portion: 59mins

Number of tokens: 9,803

Number of speakers: 7

- M2:

Original audio length: 2h41mins

Transcribed portion: 1h51mins

Number of tokens: 9,543

Number of speakers: 9

- M7_1

Original audio length: 40mins

Transcribed portion: 40mins

Number of tokens: 6,321

Number of speakers: 4

- M7_2

Original audio length: 1h32mins

Transcribed portion: 1h32mins

Number of tokens: 8,817

Number of speakers: 11

- M7_3:

Original audio length: 2h21mins

Transcribed portion: 38mins

Number of tokens: 4,023

Number of speakers: 9

- GC1_1:

Original audio length: 1h47mins

Transcribed portion: 1h47mins

Number of tokens: 13,905

Number of speakers: 14

- GC1_2:

Original audio length: 1h55mins

Transcribed portion: 1h55mins

Number of tokens: 11,341

Number of speakers: 9

- F1

Original audio length: 3h8mins

Transcribed portion: 50mins

Number of tokens: 6,844

Number of speakers: 8

- F2

Original audio length: 1h18mins

Transcribed portion: 1h18mins

Number of tokens: 12,300

Number of speakers: 6

5.Transcription procedure and conventions

The main segmentation criterion is based on discourse units (Mahlberg 2014): the longest possible units that have independent meaning and are understandable even when taken out of context. Specifically, the following criteria are applied flexibly depending on the case:

generally, one conversational turn corresponds to a single segment;

a segment can be interrupted by brief segments spoken by other speakers (in such cases, the segment is not split, and overlapping parts are indicated in square brackets).

e.g. [Yeah it's wow it's just amazing] I have to say.// [Yeah yeah]

The following transcription conventions (inspired by Jefferson 2004) have been applied:

[] = overlapping

: = long sounds

(number) = pause longer than one second

XX = inaudible portion, every X correspondes to one syllable roughly.

? = ascending tone

= hypothesis

- = clipped words

((laughs)) = double round brakets and third person verbs are used to define a non-verbal sound attributed to a specific speaker.

((laughing)) = double round brakets and -ing verbs are used to define a non-verbal sound attributed to more or undefined speakers.

Filler words are transcribed as they are heard, except for hesitations (i.e., filled pauses which could sound like uh, mh, ehm). These are all transcribed 'ehm'.

Transcription followed a double-review semi-automatic procedure, where transcripts generated by WhisperAI, used on a local server to protect the data, were revised by three different revisors. The last revisor also revised annotation (see Sections 8 and 9).

6.Encoding

Texts are encoded in UTF-8 and feature an XML structure.

7.Metadata

The metadata included in the texts relate to the Speaker(s) and communicative context. The communicative context is described in the tag <session>, which contains information related to:

- session code (an identifier attributed to each event, like F2, the second data collection event that took place in Fuerteventura);

- month and year when the data collection took place;

- country (Spain or Portugal), region (Canary Islands or Madeira Islands) and location (such as, Corralejo or Las Palmas de Gran Canaria) where the data were collected;

- planning, which refers to whether the nomads had been gathered for a particular event or if the conversation was grabbed without planning (it should be noted that unplanned data collection suffer from poor data quality, and that even the events referred to as planned are natural events organised by community managers, hostel managers, and only in few cases by the researcher with the help of other community members);

- context, which can be closed, controlled environment or public, depending on whether it's a private or public space;

- number of speakers;

- interaction which can be completely free or task-based;

- whether a self-introduction by the speakers is present within the text;

- number of different nationalities present at the event;

- all the languages known by all the speakers in the session.

The attributes specified for speakers are: "code", "nationality", and "language" (where by language we refer to all the languages that speakers have declared to know). For more information about the Speakers, see the file NomadLingo11_participants.csv.

Indeed, metadata can be accessed also in .csv formats. In the folder called 'Metadata' you can find information related to speakers and sessions in NomadLingo1.1. The last column of the files FLO_sessions.csv and NomadLingo1.0_sessions.cv provides a detailed description of the event.

The text is included into a <conversation> tag and each turn is marked by a unique ID. Turn length in tokens is also specified inside the <turn> tag.

8.Annotation schema

The annotated versions of the corpus include xml tags aimed to highlight contextual information, conversation structure, misunderstanding sequences, translanguaging phenomena and repair strategies.

The xml structure for the files contained in the NomadLingo corpus is here outlined:

```
<session ...>
	<conversation>
		<turn ...>
			<speaker ...>
				[transcription in plain text which may contain linguistic annotation] <translanguaging type="[codemixing|codeswitching]"> </translanguaging> <repair type="[reforlumation|restart]" </repair>
			</speaker>
		</turn>
		<turn ...>
			<intro>
				<speaker ...>
					[transcription in plain text]
				</speaker>
			</intro>
		</turn>
		<misunderstanding type="[solved|unsolved">
			<turn ...>
				<speaker ...>
					[transcription in plain text which may contain linguistic annotation] <interactional_move type="[asking|checking|confirming]"> </interactional_move>
				</speaker>
			</turn>
		</misunderstanding>
		<sound> [description of non-verbal sounds in plain text] </sound>
	</conversation>
```

</session>

The full annotation schema is available in both csv and xls formats in the folder Annotation_schema_NomadLingo1.1. Some examples are provided in the next Section.

9.Annotation procedure

The revision of both transcriptions and annotation involved the participation of Master's students in Specialised Translation at the University of Bologna (Alessandro Mongardini, Eva Zaccariotto, Esther Cocco, Eleonora Castaldo) and, in some cases, study participants who remain anonymous for privacy reasons. For any advice and questions you can contact novella.tedesco2@unibo.it.

10.Linguistic annotation

The corpus features four main linguistic annotation categories.

- Misunderstanding sequences are annotated at turn level. They may contain various turns and other linguistic annotation. The annotation of misunderstanding sequences is inspired by the annotation schema proposed by Cervini & Paone (2025). An example of a misunderstanding sequence is provided below (Example 1).

- Three types of interactional moves are annotated within each turn, at the level of transcription, i.e. containing the transcribed portion to which they refer, namely: a) 'asking', when speakers ask for clarification to their interlocutor (see Example 1); b) 'checking', when speakers check their interlocutor(s) understanding (see Example 2); c) confirming, when a speaker confirms their interlocutor(s) have correctly understood (see Example 3). Examples are provided below:

Example 1)

```
<misunderstanding type="solved">
	<turn ID="M7_3_0097" Length="7">
		<speaker code="PT_M_02" nationality="Portuguese" language="Portuguese,Spanish,English,German">
			Where are you from though in Italy?
		</speaker>
	</turn>
	<turn ID="M7_3_0098" Length="1">
		<speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
			<interactional_move type="asking"> what? </interactional_move>
		</speaker>
	</turn>
	<turn ID="M7_3_0099" Length="7">
		<speaker code="PT_M_02" nationality="Portuguese" language="Portuguese,Spanish,English,German">
			Where are you from though in Italy?
		</speaker>
	</turn>
```

```
        <turn ID="M7_3_0100" Length="3">
                <speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
                        From the north
                </speaker>
        </turn>
</misunderstanding>
```

Example 2)

```
<turn ID="M7_2_0008" Length="1">
        <speaker code="Researcher_IT" nationality="Italian"
language="Italian,English,Neapolitan,Spanish,German,French">
                <interactional_move type="checking"> ok? </interactional_move>
        </speaker>
</turn>
<turn ID="M7_2_0009" Length="2">
        <speaker code="GE_M_04" nationality="German" language="German,English">
                mh mh
        </speaker>
</turn>
```

Example 3)

```
<turn ID="F2_0543" Length="2">
        <speaker code="ES_C_01" nationality="Spanish" language="English,French,Spanish">
                your finance
        </speaker>
</turn>
<turn ID="F2_0544" Length="2">
        <speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
                no ehm
        </speaker>
</turn>
<turn ID="F2_0545" Length="1">
        <speaker code="GE_C_01" nationality="German" language="German,English,Spanish">
                income
        </speaker>
</turn>
<turn ID="F2_0546" Length="29">
        <speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
                <interactional_move type="confirming"> income exactly </interactional_move> depend
of your income it's like if you grow like 30000 you pay seventy percent if you're like fifty of your income
is like twenty-one
        </speaker>
</turn>
```

- Translanguaging phenomena are the third category of linguistic phenomena annotated in the corpus. For the purposes of annotation, translanguaging phenomena are considered as all those instances where translanguaging processes (Garcia and Wei 2014, Wei 2018) become evident in linguistic data. According to translanguaging theories, speakers make use of their semiotic repertoire in a holistic way depending on their communicative aims as well as on the context where communication takes place. For operational reasons, translanguaging phenomena are categorised into: a)codemixing, when a segment contains words in any language or variety other than standard English (see Example 4); b)codeswitching, when a segment is in a language

or variety other than standard English (Bathia et al. 2004) (see Example 5); c) translation (when a segment is a self-translation or peer-interpreting) (Antonini et al. 2017) (see Example 6).

Example 4)

```
<turn ID="M7_2_0082" Length="16">
	<speaker code="RS_M_01" nationality="Serbian" language="English,Portuguese,Serbian,Italian">
		<translanguaging type="codemixing">do you have glasses for the vino? This is called the vino no?</translanguaging>
	</speaker>
</turn>
```

Example 5)

```
<turn ID="GC1_2_0634" Length="2">
	<speaker code="AR_C_02" nationality="Argentinian" language="Spanish,English,Argentinian">
		<translanguaging type="codeswitching">donde queda?</translanguaging>
	</speaker>
</turn>
<turn ID="GC1_2_0635" Length="6">
	<speaker code="ES_C_04" nationality="Spanish" language="Spanish,English,French">
		it's in front of mercadona
	</speaker>
</turn>
```

Example 6)

```
<turn ID="M7_2_0254" Length="2">
	<speaker code="GE_M_04" nationality="German" language="German,English">
		plans change
	</speaker>
</turn>
<turn ID="M7_2_0255" Length="2">
	<speaker code="NO_M_01" nationality="Norwegian" language="Norwegian,English,Portuguese">
		<interactional_move type="asking">  blends strange?</interactional_move>
	</speaker>
</turn>
<turn ID="M7_2_0256" Length="4">
	<speaker code="GE_M_04" nationality="German" language="German,English">
		<translanguaging type="translation">pläne ändern sich</translanguaging>
	</speaker>
</turn>
```

It should be noted that this English-centric representation has been applied for reasons of convenience, given that English is the official declared language of many digital nomad communities (Tedesco 2025) and that the majority of the segments are in English. It should also be noticed that the distinction among different types of translanguaging does not represent any theoretical position but has only been applied for analytical reasons.

- The fourth category of phenomena annotated in the corpus is repairs. The term is generally referred to modifications made by speakers to their own speech while speaking. In this corpus, two types of repair have been annotated, namely reformulation and restart, which both cause some kind of interruption in speech flow. Restart normally occurs at the beginning of a clause, whereas during reformulation speakers modify a part of their speech by rephrasing. However, it should be noticed that the line between these two phenomena is not sharp, and some cases may fall into both categories (Schegloff et al. 1977) (see Examples 7 and 8).

Example 7)

```
<speaker code="GE_M_03" nationality="German" language="German,English,Spanish,Russian,Portuguese">
	I was born in russia but <repair type="restart">i was I descended</repair> of germans and then I went back to germany and so: so it's it's weird to say half German half Russian because-
</speaker>
```

Example 8)

```
<speaker code="IT_MC_01" nationality="Italian" language="Italian,English,Spanish">
	you need to ask for it like in thailand it's like <repair type="reformulation"> everyone every place
</repair>
</speaker>
```

11.License and use

The corpus is available for consultation on NoSketchEngine hosted by the University of Bologna servers (https://bellatrix.sslmit.unibo.it/noske/public/#open).

The transcripts, metadata and annotation scheme are realised under a Creative Commons Licence, but the recordings are closed to the public for privacy reasons. However they can be accessed, as well as the legal documentation, by request. For any queries contact novella.tedesco2@unibo.it.

12.References

Antonini, R., Cirillo, L., Rossato, L. and Torresi, I. (eds.) (2017) Non-professional Interpreting and Translation: State of the Art and Future of an Emerging Field of Research. Amsterdam: John Benjamins Publishing Company.

Bhatia, Tej K. & Ritchie, William C. (Eds.). (2004). The Handbook of Bilingualism. Malden, MA: Blackwell.

Cervini, C.; Paone, E., Annotazione di dati orali in contesti di interazione plurilingue: insegnare l'intercomprensione a studenti di Scienze mediche veterinarie, in: La comunicazione parlata 4 - Spoken communication. I venti anni del GSCP., Roma, Aracne, 2025, pp. 53 - 65 (La comunicazione parlata)

García, Ofelia & Li Wei (2014). Translanguaging: Language, Bilingualism and Education. London: Palgrave Pivot.

Jefferson, Gail (2004) Glossary of transcript symbols with an Introduction. In G. H. Lerner (Ed.) Conversation Analysis: Studies from the first generation (pp. 13-23). Philadelphia: John Benjamins.

Mahlberg, M. (2014). Corpus linguistics and discourse analysis. In K. P. Schneider & A. Barron (Eds.), Pragmatics of Discourse (pp. 215–238). De Gruyter.

Schegloff, E., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organisation of repair in conversation. Language, 53, 361-382.

Tedesco, N. (2025). Translanguaging in the era of digital nomadism: a sociolinguistic perspective on voluntary mobility in Europe. Discov glob soc 3, 50. https://doi.org/10.1007/s44282-025-00191-8.

Tedesco, N., Bernardini S., Cervini, C. (2025). NomadLingo1.0 open [dataset]. http://hdl.handle.net/20.500.11752/OPEN-1042

Wei, L. (2018) Translanguaging as a Practical Theory of Language, Applied Linguistics, Volume 39, Issue 1, Pages 9–30, https://doi.org/10.1093/applin/amx039

13.Acknowledgements

Contributors: Silvia Bernardini, Cristiana Cervini, Alessandro Mongardini, Eva Zaccariotto, Esther Cocco, Eleonora Castaldo, digital nomad communities based in: Porto Santo, Funchal, Ponta do Sol, Las Palmas de Gran Canaria, Maspalomas, Corralejo, Lajares, El Cotillo.