

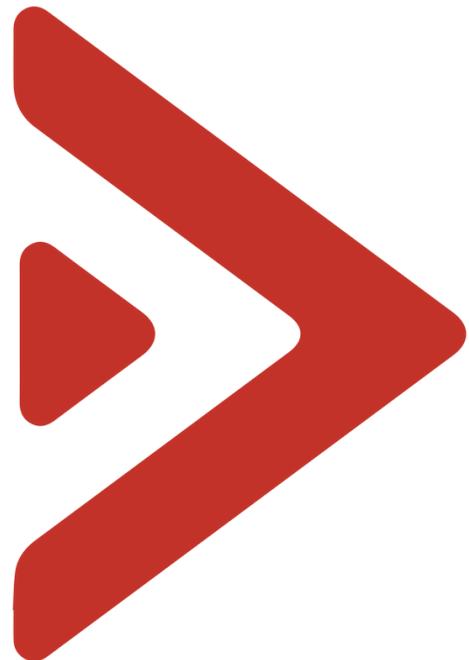
WORKING PAPERS  
OF THE OBSERVATORY OF SPANISH IN EUROPE  
NO. 2 | DECEMBER 2025

DOI: 10.48629/hcias.2025.1.114686

**How to cite:** Álvarez Mella, H., Blattner, C., Gómez-Pavón Durán, A., and Elizondo Romo, M. (2025). (In)visibilization of migration-driven linguistic diversity: What can we know about migratory languages demographics in Europe? *Working Papers of the Observatory of Spanish in Europe* No. 2. Heidelberg Center for Ibero-American Studies, Heidelberg University.

# (In)visibilization of migration-driven linguistic diversity: What can we know about migratory languages demographics in Europe?

**Héctor Álvarez Mella**  
**Charlotte Blattner**  
**Ana Gómez-Pavón Durán**  
**Mariel Elizondo Romo**  
Universität Heidelberg





# **(In)visibilization of migration-driven linguistic diversity: What can we know about migratory languages demographics in Europe?**

Héctor Álvarez Mella<sup>1</sup>, Charlotte Blattner\*<sup>1</sup>, Ana Gómez-Pavón Durán<sup>1</sup>, and Mariel Elizondo Romo<sup>1</sup>

<sup>1</sup> Heidelberg Universität

---

## **Abstract**

This study provides a critical analysis of the data sources used in migration demolinguistics studies, with a focus on Spanish-speaking migration in Europe, and examines both direct and indirect sources of data, including population censuses, registers, surveys, and data on the labor market and educational system. The paper highlights the challenges posed by the lack of data on migration languages, which may be linked to monolingual ideologies, as well as the issue of data comparability across different sources. The authors emphasize the need for standardized data sources and better coordination between different sources to improve the quality and comparability of data on migrant populations. The paper underscores the importance of understanding the social and cultural contexts in which language use occurs to achieve a more nuanced understanding of language use and ability in migrant populations.

## **Keywords**

multilingualism - migration - migratory languages - demolinguistics - demography - linguistic diversity - population statistics

## 1. Introduction

Multilingualism is at the core of European societies. Although countries and supranational institutions are committed to respecting and protecting linguistic diversity, European language regimes mainly focus on *autochthonous multilingualism*, i.e., national and regional languages traditionally tied to territories and communities within nation-states, neglecting migratory languages (Gogolin 2002, Kupisch 2021).

According to Eurostat (2025a), 5.3% (23.8 million people) of the 446.7 million people living in EU-countries were non-EU citizens. These figures, derived from population statistics and spread through social media such as Twitter, can be misleading and underestimate the role of migration in the construction of European societies and language diversity in Europe. Actually, if we zoom through different spatial scales, the contribution of migration to Europe's demographic reality and future becomes more evident. Taking Germany as an example, in 2024 around 30% of the population were immigrants or descendants of immigrants, and nearly half of all families (43%) had a migration background (Statistisches Bundesamt 2025; Federal Ministry for Family Affairs, Senior Citizens, Women and Youth 2024).

Despite the contribution of migration to the linguistic constellation of Europe, migration-driven language diversity is invisibilized at different levels (Gogolin 2002, Extra 2017). First, migratory languages are not explicitly included in the European treaties for protecting language diversity<sup>[1]</sup>, consequently not sufficiently supported by educational systems. Second, institutional demography and population data tends to overlook migratory languages. And third, the absence of accurate quantification of the migrant languages spoken within the borders of European countries can lead to their invisibilization.

The aim of this article is to present a comprehensive compilation, classification and critical analysis of linguistic and demographic sources for the study of migratory languages in Europe. In doing so, we address the central research question of this paper: What can we know about the demography of migratory languages in Europe? Our analysis is based on the assumption that the description and critical commentary on the types of data, the sources, their availability and their comparability is necessary to avoid the statistical oversimplification of linguistic diversity. Indeed, one of the main problems is that the use of demographic data on migration tends to convey a picture of migration groups that is more homogeneous than it actually is (Schwenke 2018). This paper aims to enhance our understanding of migration-induced linguistic diversity in Europe by examining Spanish-speaking migration as a case study to identify the most appropriate sources for demolinguistic research and to develop a model applicable to other migratory language communities. The paper first outlines key aspects of language demography and migration-based linguistic diversity, then discusses approaches to collecting language data, distinguishing between direct and indirect sources. However, the epistemological stance underlying this paper is far from abandoning or postponing the demolinguistic efforts because *“to count, we must be counted”* (Gómez 2020: 134). Linguists as well as politicians need to use demolinguistic data, as population statistics are the only existing source of large-scale information. For both linguistic and political purposes, it is crucial to know the availability, status and comparability of demographic sources and what linguistic information can be obtained from them.

---

[1] European Charter for Regional or Minority Languages.

## 2. Language demography and migration-based linguistic diversity

Language demography studies the links between the social use of languages and the demographic structure of their language communities from the quantitative or statistical analysis of speakers and their spatial distribution (Pauwels 2016; Moreno Fernández 2023). Although it often overlaps with the scientific enterprise of counting speakers, language demography is more concerned with the analysis of existing data mostly created by national or regional institutions through censuses, surveys or registers. In fact, demolinguistics is continuously critically reviewing the validity of statistics on language speakers, including the criteria used to gather the data and the comparability of different sources. Because demolinguistics relies on large datasets that it cannot generate independently, it is crucial for the discipline to ensure the accuracy and reliability of the information it uses.

When applied to the study of migration-driven linguistic diversity, demolinguistics faces conceptual and methodological problems derived from two related epistemological biases that affect both data collection and its analysis. The first bias is *methodological nationalism*, a spread approach in the social science and migration studies "that assume that countries are the natural units for comparative studies, equate society with the nation-state, and conflate national interests with the purposes of social science" (Wimmer and Glick Schiller 2003: 576). In demolinguistics, methodological nationalism has influenced the epistemology of the study of language in different ways, from the definition of particular linguistic practices under categories oriented from national social frames (Busch 2016), to the use of *nation* or *ethnicity* as naturalized notions of community to which one or several languages are assumed and particular linguistic repertoires are expected (Leeman 2018; 2020).

Added and related to methodological nationalism, the second problem for the study of linguistic diversity is the *sedentary bias* (Bakewell 2007), which assumes that mobility is an exceptional state of human societies and understands migration "as harmful and dysfunctional" (Castles 2010: 1567). In line with that, several linguistic regimes associate languages with territories and identify monolingualism as the natural state of communities. Consequently, from this perspective, authentic speakers of a language are understood to be those rooted in the territories and communities where that language holds social dominance, while mobile speakers—generally bilingual or multilingual—tend to be treated as peripheral phenomena within linguistic communities.

In order to avoid *methodological nationalism* and *sedentary bias*, demography of language must shift from space to the individual by focusing on the linguistic repertoires of migrants (Blommaert and Backus 2013). Demolinguistics has to turn its gaze to the main object of demography, the individual life courses, seeking data to examine the processes of linguistic socialization in which speakers acquire linguistic competences during their lives.

Under this lens, the main goal of demolinguistics of migration is to provide an analysis and description of the structure, dynamics, and composition of a migrant community, characterized by the knowledge or usage of a specific language or linguistic variation. Its primary focus is the study of people who can communicate in a language in different migration settings that can be identified using criteria such as "birth in another country, foreign status -which excludes immigrants who have obtained the nationality of the country of destination-, having previously resided in another country or, finally, the fact of having a migratory background, which includes people with direct migratory experience and also those born of migrants in the country of destination" (Loureda et al. 2022: 10).

The concept of migratory background refers to different generations, including both first-generation immigrants and their descendants (i.e., second and third generations). It is worth noting that, under the common definition of international immigrant, the descendants of immigrants are not considered migrants as they have not undergone any spatial displacement. Also, the attribution of migration labels to the descendants of immigrants is highly problematic from the scientific and political perspective (Schwenke 2018). However, as demography of migration is concerned with analyzing cultural phenomena such as language, it has to consider the intergenerational transmission processes that occur within migrant communities.

Taking Spanish as an example, the demolinguistics of Spanish-speaking migration studies the group of people with a migratory background who include Spanish in their linguistic repertoires. To determine who is a **Spanish speaker with a migrant background**, it is necessary to define the two concepts that make up the notion: a) migrant background and b) Spanish speaker.

a) A person with a **migration background** is a person whose biography is or has been decisively influenced by migration, including both migrants who have moved to another country (international migrants<sup>[2]</sup>) and their descendants regardless of their place of birth. The concept of migration background aims to capture the impact of migration on social and cultural dynamics, such as multilingualism or sociolinguistic integration.

b) A **Spanish speaker** is a person whose linguistic repertoire includes communicative competencies in Spanish. Their level of proficiency in Spanish may vary according to the competencies they have acquired throughout their linguistic biography in different spaces of socialization and learning (family, school, community, etc.). According to this criterion, a distinction can be made between a native speaker, capable of using the language or variety "in a way that is appropriate in form and adequate to the communicative context" (Moreno-Fernández 2014: 23), and a speaker of limited competence, i.e. with a sociolinguistically or stylistically limited command and restricted to certain topics or situations (Loureda et al. 2020).

Considering both definitions, the concept of Spanish speakers with a migratory background encompasses a heterogeneous group of different speaker profiles. From the perspective of demolinguistics of migration, there are several groups of people for whom Spanish plays a significant role in their linguistic biography, either as a native speaker or with limited proficiency, regardless of the means of acquisition (e.g. learned through socialization in the Spanish-speaking family or community, as a foreign language, through linguistic immersion in a migratory context, etc.):

- Individuals coming from Spanish-speaking countries who reside in Europe, with or without the nationality of the host country.
- Their descendants, regardless of whether they have been born in a Spanish-speaking country, the host country, or another country.
- Immigrants coming from non-Spanish speaking backgrounds residing in Europe who have acquired Spanish language skills regardless of the context;
- Europeans who are not native Spanish speakers but have acquired Spanish language skills in migratory contexts.

---

[2] The United Nations (UN) define an *international migrant* as "any person who moves to a country other than that of his or her nationality or habitual residence in such a way that the country of destination effectively becomes his or her new country of habitual residence" (United Nations 2022). The **category of migrant** is therefore mainly spatial, as it refers to displacement, and political-administrative, as it depends on the legal frameworks and the migration and nationality regimes of the states that determine which persons are included in the immigrant statistics.

Considering these mobilities, *Spanish-speaking migration* constitutes a group defined by a linguistic attribute: knowing and/or using the Spanish language. Accordingly, the demography of the Spanish-speaking migration deals with a linguistic or communicative reality that is not always, nor comprehensively, reflected in population statistics, as it goes beyond the categories usually used as linguistic indicators for demographics: *mother tongue* and *ethnic origin*. To understand how this affects the demolinguistics of migration, it is necessary to dissect the process of demolinguistic data collection and analysis.

### 3. Collecting information on language demography

The demography of migratory languages must draw upon both migratory and linguistic data. The main obstacle is the lack of adequate data, and even when data is available, comparability can be challenging due to the lack of consensus and clarity in the cultural factors collected in population statistics (Duchêne and Humbert 2018; Leeman 2020; Loureda et al. 2021; Leeman 2023). Another hurdle arises from the fact that demolinguistics often rely on data created from "administrative actions and procedures, whether at the local, national or international level" that has not been collected using sociolinguistic methods and even less for linguistic purposes (Moreno Fernández 2023). Beyond the problems of the sources, the aforementioned bias makes more complex the study of migratory populations in Europe. Statistics do not fully reflect the reality of the European linguistic landscape because they only collect information on individuals registered in the country's legal system for at least one year and frequently do not include second and third generation migrants, who generally receive the linguistic heritage of their parents.

Despite their inherent limitations, demolinguistic studies strive to reconstruct an accurate depiction of reality by utilizing all available data. In the context of language research, language demography plays a vital role in documenting and monitoring the linguistic composition of countries and regions, as well as the socio-linguistic integration of communities, while also assessing the status and vitality of minority or minoritized languages, including those associated with migration. Considering language policy and governance frameworks, quantifications and demolinguistic analyses contribute greatly to supporting language policies and planning aimed at managing aspects such as language usage spaces, education, and rights, among others.

This quantitative approach is the distinguishing feature of demolinguistics with respect to other disciplines dedicated to the study of linguistic populations such as ethnolinguistics (Extra 2010). Although demographers usually refer to "population counts", demographic data are closer to probability than to accounting. Indeed, demographic data, which are probabilistic in nature, enable estimation of the size of a population group in a given space, its evolution over time, as well as the likelihood that their members possess certain sociodemographic and cultural characteristics, including linguistic competence.

From the spatial point of view, demolinguistics use linguistic and population data contained in population statistics at the local, national and international levels. In Europe, this type of sources usually have several problems of a) availability and b) comparability:

a) First, there is no *linguistic census* in most European countries and linguistic information is not always available in demographic sources. However, language is one of the human attributes that has been of interest to demography. Indeed, population censuses and other official demographic sources include information related to language, such as the ethno-linguistic identity of citizens or their language uses and skills.

b) Secondly, existing demolinguistic data are not always comparable because it depends on the definition of linguistic categories used by institutions, and the techniques and tools selected for data collection, and census and documentation practices, which are generally very varied and uncoordinated.

Both limited availability and comparability force demolinguistics to construct databases by combining several statistical sources, which must be test for accuracy and comparability. Thus, the processes of demolinguistic documentation attempts to use direct (or linguistic) sources, which contain information on the languages of citizens, such as population censuses and civil registers, and combine it with indirect (or demographic) sources that make it possible to quantify groups of speakers on the basis of sociodemographic indicators that suggest more or less probable linguistic profiles.

## 4. Direct sources

Direct sources in demolinguistics "provide first-hand information on the number of speakers of a language in a given territory" (Moreno Fernández 2023: 94). This information is gathered through linguistic questions (e.g., mother tongue, home language, degree of proficiency in the local language, etc.) that are included in censuses, population registers, or surveys (Extra 2017).

### 4.1 Language in the censuses

Censuses provide periodic counts (usually conducted every 10 years) of the population and dwellings of a given country or region, and collect information on its main characteristics (geographic, demographic, social and economic, as well as household and family characteristics). Their purpose is to provide a statistical basis for governance and management of social aspects of a country or region. Unlike civil registers (see § 4.2), censuses collect their data from a sample that allows a more or less adjusted statistical extrapolation of their results to the population as a whole.

Population censuses usually include information relevant to the demolinguistics of migration in sections dedicated to languages, and also to ethnic identity, given that certain ethnic groups are usually defined by the shared use of a language or linguistic variety. In the case of certain migratory groups, such as Spanish-speaking migration, however, ethnicity is not relevant since the Spanish-speaking linguistic community does not constitute an established ethnic entity as such in the demographic statistics on any of the European countries (Leeman 2023).

The most valuable census information for demolinguistics of migration is derived from linguistic questions in combination with other sociodemographic variables of the citizens. While there are international recommendations for unifying questionnaires, topics, and questions, Europe lacks a common population census, resulting in a highly heterogeneous picture across the region. An analysis of the census questionnaires of all European countries over the last twenty years allows us to classify them into three categories according to their "interest" in language (**Table 1**):<sup>[3]</sup>

a) *Censuses without linguistic questions*. According to the available data, 34.6% of the European countries does not include linguistic questions in their population censuses,

---

[3] Countries not listed in the table are those for which no information is available: Andorra and San Marino.

showing that the collection of statistics on the linguistic repertoires and usage of the population is not a priority.

b) *Censuses with only one linguistic question.* 22.4% of the European countries includes at least one language-related question in their census, which usually refers to the mother tongue of the citizens. This question is used to estimating the number of people who belong to the group of native speakers of a language. Germany and Austria place their focus on the home language.

c) *Censuses with several linguistic questions.* 42.8% of the European countries includes more than one linguistic question in their population census. Most countries include the question of mother tongue or native language combined with other linguistic information. In some countries, such as Ireland, Spain, or the United Kingdom, a distinction is made between the regional and the majority languages, and level of proficiency in the latter is also recorded. Luxembourg and Switzerland, on the other hand, make a distinction between main language (“the language you think in and know best”) and language used in the different domains (home, school, work).

Type	Subtype	Countries
<b>a) Census without a linguistic question</b>		Belgium, Denmark, France, Greece, Iceland, Italy, Latvia, Liechtenstein, Monaco, Netherlands, Norway, Poland, Portugal, Sweden, Turkey
<b>b) Census with a single question</b>	Mother tongue	Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Finland, Romania, Serbia, Slovakia
	Home language	Austria, Germany
<b>c) Census with several questions</b>	Language spoken at home/with family and friends	Belarus, Hungary, Ireland, Luxembourg, North Macedonia, Slovenia, Spain, Switzerland
	Main language	Luxembourg, Malta, Moldova, United Kingdom, Switzerland
	Mother tongue / Native language	Armenia, Azerbaijan, Belarus, Cyprus, Estonia, Georgia, Hungary, Kazakhstan, Lithuania, Moldova, Montenegro, North Macedonia, Slovenia, Spain, Ukraine
	Language spoken at work/educational institution	Luxembourg, Switzerland
	Other languages spoken fluently	Armenia, Azerbaijan, Belarus, Cyprus, Estonia, Lithuania, Hungary, Montenegro, Russia, Ukraine
	Regional language	Azerbaijan, Estonia, Georgia, Ireland, Spain, United Kingdom (Wales, Scotland, Northern Ireland)
	Proficiency level in the country's majority language	Ireland, Kazakhstan, Spain United Kingdom

**Table 1:** Classification of European censuses according to linguistic information. Source: United Nations Statistics Division and censuses from the countries in the EU.

Census information presents certain limitations for demolinguistic documentation and analysis. Firstly, the heterogeneity of questions and methods used to document linguistic information makes it challenging to compare the results of national censuses. This fact has an obvious effect on demolinguistics: they do not allow an aggregate quantification of the number of speakers based on data from different censuses using different language questions. Secondly, the sampling nature of censuses and the ways in which information is tabulated and reported may not capture minority language groups due to the lack of representability: National statistical offices only report information on the most common non-national languages in the country due to small samples of language groups that are insufficient for accurate extrapolation to the whole population.

Generally, censuses in Europe are not very effective in capturing linguistic diversity. This is due to the predominance of monolingual ideologies that are not sensitive to heterogeneity and are rooted in methodological nationalism and sedentary bias. The EU censuses that include one or more linguistic questions (see **Annex 1** for a detailed description) collect information about:

- Mother tongue with only a single answer possible, either as an open-ended question (e.g. Romania, Slovakia) or as a dual choice between national language and other (e.g. Albania).
- Language(s) most commonly spoken at home (Germany and Austria)
- Mother tongue and main language or language spoken fluently (e.g. Estonia, Lithuania)
- Mother tongue and knowledge of the regional language (e.g. Ireland, Spain)
- Main languages and languages used in different domains (e.g. Switzerland, Luxembourg)

Taking Spanish as an example, when one linguistic question is asked, a Spanish-speaker will show up on the data only when they choose Spanish as either their mother tongue or a language spoken fluently. In the latter case, no information is given of whether the language was acquired as a home language or as a foreign language, only that it is part of the linguistic repertoire of the speaker; in the former case, the time of acquisition is known but not the domains where the language is spoken. Only the censuses with the most linguistic questions (Switzerland, Luxembourg, and Spain) allow for a more complex profile of multilingual speaker to be adequately represented; on the contrary, the linguistic questions in most censuses are oriented towards a more static notion of a bilingual speaker who is imagined as having the national language as one of the default languages in their repertoire and a migratory or a regional language as the second.

The usefulness of censuses for analyzing the demolinguistics of migration is highly dependent on the nature of the linguistic questions included, and they are generally only applicable for country-specific analyses. Furthermore, it is worth noting that the European Union and other international organizations are advocating for a methodological transformation of demographic documentation tools, whereby registers are assuming a greater responsibility, in response mainly to the growing recognition of regional cultures and languages. This new approach, which has already been implemented in some countries for their 2021 censuses, involves merging sample data from digital questionnaires with sociodemographic information gathered from registers like the census or civil registry. These registers have expanded the scope of information required from citizens during their administrative processes.

For quantifying speakers with a migratory background in Europe, linguistic information from censuses can be used to contrastively adjust demolinguistic and population data

collected from other sources. What makes censuses particularly valuable is that the linguistic information can be examined in relation to sociodemographic variables, such as sex, age, country of birth of the respondent and sometimes also variables related to their parents, length of stay in the country, marital status, type of marriage (endogamous/exogamous), etc. (Clyne 2003; Pauwels 2016). National statistical institutes of the European countries usually make available selected reports and statistics that cross-reference the linguistic question(s) with one of the above-mentioned variables. But a more in-depth study of correlations requires access to census microdata that are usually made available to interested researchers through an official request form.

#### 4.2 Language in the population registers

Civil registers often include information on realities, events and processes that citizens report to local, regional or national administrations, such as names, births and deaths, civil status or nationalizations. Capturing citizens' linguistic knowledge and other aspects of their cultural repertoires can be challenging to record accurately. There are, nonetheless, administrative institutions in Europe that gather linguistic data.

Finland stands out as a paradigmatic country that registers linguistic information about its citizens. The Finnish Population Information Systems is a national registry with basic information on Finnish citizens and foreign citizens residing in Finland permanently or temporarily. Among the personal information registered is the mother tongue (native language). The Finnish national registry is particularly relevant for the study of migration demolinguistics, given that the registration form for foreign populations includes a question specifically asking for their native language. As shown in **Figure 1**, the form only allows the entry of a single language, excluding the possibility of representing the linguistic profiles of speakers with several mother tongues.

12. Country of birth		13. Place of birth	
14. Native language (Only one can be registered)			
15. Contact language		If your native language is neither Finnish nor Swedish, please indicate the language in which you would like to receive official messages that are not available in your native language.	
<input type="radio"/> Finnish	<input type="radio"/> Swedish		
16. Citzenships			

**Figure 1:** Linguistic question contained in the Finnish population register. Source: Finnish Population Information Systems.

Registers have an obvious advantage over censuses: they allow for a highly accurate quantification of a given linguistic group. However, whether the advantage of digital data collection and processing and the universal character of the register can be exploited depends on the type and quality of the linguistic information collected. If, as in the Finnish registry, individuals are only asked about their mother tongue and are limited to providing a single response, the registry may fail to capture the complex multilingual repertoires that are especially common among immigrant populations, rendering them invisible.

### 4.3 Language in surveys and linguistic surveys

Surveys can be carried out either by administrative authorities or by other entities or initiatives in a given area. The main goal of surveys is often to gain a deeper understanding of a particular linguistic reality, and as such, they typically include more targeted questions than a census. In addition to quantifying the presence of different languages, surveys also seek to gather information about language use, proficiency, frequency, attitudes, and other related factors. Surveys can be complemented with interviews or focus group meetings.

Demolinguistics finds relevant information for the study of migratory languages in surveys carried out at the international European level, such as the Eurobarometer or the Labor Force Survey, and at the national and regional level in the different countries where they are available.

The **Eurobarometer** stands out as the preeminent language survey within the European Union. Although its main objective is to study the foreign language proficiency of European citizens, in its three versions of 2000, 2005, 2012, and 2023 it includes a question on mother tongue. In its first two versions, the questionnaire asked only about mother tongue; in 2012 and 2023 the wording of the question changed to "Thinking about the languages that you speak, which language is your mother tongue?". Since the survey has samples for each country, its results allow us to estimate how many people have Spanish as their mother tongue in each of the non-Spanish-speaking countries of Europe.

The **Labour Force Survey** is the largest European sample survey of private households, which has been conducted since 1983. Through quarterly results, the participation in the labor market of people over 15 years old is measured, as well as those who are not part of the active population. Once the surveys have been conducted using questionnaires and interviews, and samples have been selected, the national statistics institutes of various countries (currently 35 countries participate) transmit the results to Eurostat. Eurostat then harmonizes the data at the European level. The survey contains a dedicated module focused on the labor market situation of migrants and their immediate descendants ("Labour market situation of migrants and their immediate descendants"), which has been included in 2008, 2014, and 2021. The module aims to assess whether a migrant's country of training has any impact on their integration into the labor market. It includes responses from foreign citizens and their descendants over 15 years old to three specific questions:

1. Proficiency in the main language of the host country prior to migration
2. Current proficiency in the main language of the host country
3. Participation in a language program in the main language of the host country

The second question enables the identification of immigrants' proficiency in the main language of their host country. This is useful when studying migration in Spain, since it allows for an assessment of the level of competence in the Spanish language among this group. When combined with other variables, it provides a comprehensive understanding of other aspects such as generation, age, sex, and origin.

In addition to surveys at the European level, there are also initiatives by national and regional institutions to carry out linguistic surveys. They are carried out with the aim of finding out the uses and linguistic knowledge of the population, both of national and regional languages, as well as of foreign languages.

An instance of this would be the survey on **Citizens and their free time ("I Cittadini e il Tempo Libero" (CTL))** which has been conducted in Italy since 1995 and is part of an integrated system of social surveys. It includes questions on the use of Italian and its dialects, the use of the mother tongue of foreign citizens in their households or within their closest social circles, and the foreign language skills of the population (Istituto Nazionale di Statistica 2017). Similarly, the German Socio-Economic Panel Study (SOEP) of the **German Institute for Economic Research (DIW)**, which annually collects representative information on socioeconomic aspects of private households in Germany, contains questions related to language uses in the household, mother tongue and foreign language learning (Adler 2019).

In some multilingual European countries, surveys are organized to find out about the practices and uses related to the various co-official languages. In Switzerland, a five-yearly survey on language, culture and religion of the population has been conducted since 2014. This survey includes questions on the languages most used in different stages and contexts (in childhood, with descendants, with partners, with friends, at home, in the workplace, etc.); as well as on the languages that the population wishes to learn (Bundesamt für Statistik 2022). In Spain, there are also surveys on the knowledge and linguistic uses of Galician (Instituto Galego de Estadística), Catalan (e.g. Survey of linguistic uses of the population of the Generalitat de Catalunya) or Basque (Instituto Vasco de Estadística).

Given the changes that migration generates in the linguistic reality of a country, the migrant population in some countries receives special attention from national institutions. In the German SOEP surveys, participants of foreign origin must answer questions related to their knowledge of the German language, their mother tongue and other foreign languages, and their use of languages in different areas (family, work, friendships). Some states even consider it necessary to design surveys exclusively for this population group, such as the Spanish **National Immigrant Survey** (2007) conducted by the National Institute of Statistics (INE), which collects information on the mother tongues of the participants.

## 5. Indirect sources

As observed in the analysis of direct sources, it is typical for demographic data to lack details regarding the languages spoken by citizens. Moreover, when available, such information may be incomplete, fragmented, and subject to conceptual and methodological constraints, which hinder demolinguistic analysis. For this reason, the study of languages and their populations usually resorts to indirect demographic sources that enable the estimation of linguistic information from primary data, generally of a sociodemographic type (Moreno Fernández 2014; Moreno Fernández 2023). The idea behind the estimates is that specific data pertaining to citizens can offer insights into their linguistic socialization processes in which they are or would have been immersed.

The following section presents the indirect sources used to gather information on the following speaker profiles:

- Individuals with migration background who reside in Europe (Individuals with their own migration experience with or without the nationality of the host country and descendants of immigrants born in a Spanish-speaking country, the host country or another country);
- Europeans who are not native Spanish speakers but have acquired Spanish language skills in migratory contexts.

Indirect demographic sources do not allow for reliable estimates of the population of immigrants from non-Spanish-speaking backgrounds residing in Europe who have acquired Spanish language skills, regardless of the learning context. If a country's censuses or registries do not include linguistic questions, this information can only be inferred from reports on Spanish learners compiled by the Instituto Cervantes in its accredited centers. However, such data do not provide an accurate picture of the actual number of migrants who have learned Spanish as a foreign language.

### 5.1 Social indicators for linguistic competence

The main demographic variable that makes it possible to identify all these groups in population statistics is provenance, which is usually approached by demolinguistics on the basis of two social indicators: nationality and country of birth.

Eurostat offers a definition of the two categories that can be used for any European demographic source:

- *Nationality* denotes the legal bond between an individual and his or her state, which is acquired by birth or following a naturalization process based on declaration, election, marriage or other avenues considered in the naturalization laws in each country.
- The *country of birth* is the country in which the birth takes place, which usually coincides with the country of residence (considering its current borders) of the mother at the time of birth.

The national statistical services of each country gather information on nationality and country of birth that is usually collected in different civil registers, such as the residents' register or the foreigners' register. Both demographic data can be used as linguistic indicators, but certain considerations must be taken into account. For instance, it should be noted that having the nationality of a country or being born there does not necessarily indicate proficiency in one or more of the languages spoken in that country. However, its usefulness for studies in linguistic demography is evident because it allows for the objective identification of a group and comparison of data between countries and regions. In contrast to country of birth, which is a circumstantial and invariable fact that does not reflect sociolinguistic integration dynamics, nationality combined with naturalization statistics allows for the grouping of citizens who may share the language associated with their nationality (Extra 2017; Leeman 2023). In the case of the Spanish-speaking population from Latin America, for example, the combined comparison of both indicators is important if it is considered that some citizens "may have the nationality of a European country, as is the case with those from Argentina, Uruguay and Venezuela with Italian nationality" (Leeman 2023).

The analysis of Spanish-speaking migration and its characteristics relies on various data sources available from the two main regions under consideration. The majority of data on migration in Spain comes from the National Institute of Statistics (INE), while data on other European countries is largely sourced from Eurostat (**Table 2**).

Territory	Source and time frame	Data type
Spain	INE Population figures (1998-2021)	Resident population of foreign origin (naturalized or not) according to origin, sex and age
	INE Continuous Household Survey (2013-2021)	Descendants of foreign nationals

<b>Europe</b>	Eurostat Population figures (1998-2020)	Resident population of foreign origin (not naturalized) according to origin, sex and age
	Eurostat Acquisition of citizenship (1998-2020)	Foreign population that has acquired the citizenship of a European country according to year and country of naturalization
	Eurostat Fertility statistics (2011-2020)	Birth rate indicators by age and country (allows estimation of second generation group)

**Table 2:** Sources for the demographic study of migration.

The sources mentioned above provide us with valuable information for studying various migratory profiles. As mentioned earlier, the Spanish-speaking population with a migratory background comprises several subgroups that can be distinguished with varying levels of accuracy using population statistics. **Table 3** presents the primary profiles based on the sources used for their quantification:

<b>Source</b>	<b>Migration profile</b>
<b>Eurostat</b>	Spanish-speakers with a migrant background residing in Europe (excluding Spain)
<b>INE</b>	Spanish-speakers with a migrant background residing in Spain and European migrants returning to their countries of origin

**Table 3:** Sources for the quantification of migration profiles in Europe.

The two demographic sources are presented below, based on the consideration of the demography of Spanish as a migratory language in Europe.

## 5.2. European Statistical Office (Eurostat)

The group of Spanish speakers with a migratory background residing in Europe (excluding Spain) can be reconstructed from the data provided by the European Statistical Office (Eurostat), which carries out the largest compilation and unification of population data in Europe. In its database (<https://ec.europa.eu/eurostat/data/database>), Eurostat compiles population statistics provided by member states, reflecting demographic counts as of December 31st of the reference year. These statistics are published in a central database with reference to January 1st of the following year.

The unit of observation for population statistics is the usual resident, i.e. the person who "normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment, or religious pilgrimage" (Eurostat 2025b). To be included in these statistics, individuals must meet one of two criteria: either they have been a usual resident in the country for at least twelve years prior to the count, or they have arrived in the country within the last twelve months with the intention of residing there for at least one year. This criterion excludes individuals who are only in the country temporarily for short-term mobility purposes, such as tourists, exchange students, seasonal workers, or employees sent by companies to their foreign delegations for short stays.

Among its population statistics, Eurostat publishes two series of data that provide information on the main social indicators of migration: country of origin (Series: "Population on 1st January by age group, sex and country of birth") and citizenship

(Series: "Population on 1st January by age group, sex and citizenship"). When analyzing Spanish-speaking migration, the population database by citizenship can help us identify the foreign population group with citizenship from a country where Spanish is an official language. The sociodemographic variables in the database make it possible to describe the group and its evolution since 1998:

- Sociodemographic variables: gender and age. Age is particularly relevant for demolingistic studies as it provides information on the aging of the group and its generational trends.
- Spatial distribution. Since the data is recorded in each country, information about the place of residence is also available. This allows for analysis of spatial distribution. The Eurostat data portal does not offer statistics on the regional distribution of the population within each country, although according to the methodological description, the data should be available. To obtain this information, one must consult national sources. The place of residence is relevant because it allows for the description of linguistic spaces where the group is found, the languages with which the speakers come into contact, and also allows for the identification of spatial patterns such as some priority relationships between origin and destination.

Thus, the main advantage of the *citizenship country* criterion is that it can be combined with naturalization statistics. Immigrants who have obtained citizenship in the host country are no longer included in subsequent statistics as residents with their previous nationality, even in cases where dual nationality allows them to maintain it. Eurostat's database ("Acquisition and loss of citizenship" [migr\_acqn]) collects national statistics on citizenship acquisitions that occur each year in European countries. Analyzing this source allows for the identification of a group of people from Spanish-speaking countries who do not appear in the population database with their previous nationality from a Spanish-speaking country.

Eurostat does not include statistics on immigrant offspring, so it is not possible to obtain a direct quantification from its databases. However, birth statistics combined with data on the foreign population by origin allow us to estimate a group of persons with at least one immigrant parent, taking into account their origin.

### 5.3. Spanish National Institute of Statistics (INE)

In Spain, population statistics provide several databases that are useful for the demolingistic study of Spanish-speaking migration. The National Institute of Statistics (INE) offers demographic information that allows for the quantification and description of the group of people with a migratory background who speak Spanish in Spain. This group consists of two main categories: migrants from Spanish-speaking countries and regions, and migrants from non-Spanish-speaking countries who have acquired Spanish language skills through their migratory experience.

The data series of the National Statistics Institute (INE) that allows for a homogeneous grouping of migratory origins began in 2002 and is available in the section dedicated to population figures. The resident population statistics make it possible to quantify citizens of foreign origin (naturalized or not) according to origin, sex, and age. The group of citizens from a Spanish-speaking country can be assumed to have native Spanish language skills, with little margin of error. To better estimate the language proficiency of the remaining population, the statistics can be supplemented with the data from the language section of the Active Population Survey-Module (*Encuesta de Población Activa-*

*Módulo*<sup>[4]</sup>, national version of the Labour Force Survey, discussed in **section 4.3** of this paper), which provides the percentage breakdown of migrants by origin who consider themselves proficient in Spanish at different levels: native, advanced, intermediate, or beginner.

The Residential Variation Statistics (*Estadísticas de Variaciones Residenciales*<sup>[5]</sup>) is a source of great value for estimating the group of people with a migratory background who have returned to their countries of origin after a stay in Spain. This group of returnees is relevant for the demolinguistics of Spanish-speaking migration because they are people of European origin who have acquired Spanish language skills during their migratory experience.

Finally, the Continuous Household Survey (*Encuesta Continua de Hogares*)<sup>[6]</sup> offers a basic description of the Spanish-speaking immigrant descendants group. The data is categorized by country of origin, but only includes descendants of foreign mothers, excluding cases of descendants of foreign men married to Spanish nationals. Furthermore, third-generation descendants, who are often overlooked in European population statistics, are not included in the survey.

## 6. Conclusions

This paper provides a critical review of the sources available for the demolinguistic study of migration in Europe. The combination of linguistic, demographic, and migratory statistics with population data is essential for establishing the empirical basis needed for a quantitative analysis of language speakers and to highlight linguistic diversity driven by migration. The sources discussed in this paper allow for a quantitative approach to the main questions of migration demolinguistics, including: the number of individuals with a migratory background who are proficient in a given language, their linguistic and socio-demographic profiles, the factors influencing their spatial distribution and language use, and their future demographic scenarios.

The analysis of the demolinguistic sources presented in this study allows for several conclusions to be drawn. First, direct sources show notable limitations for describing migratory languages and migration-based linguistic diversity in Europe. A total of 34.6% of countries do not include any linguistic question in their population censuses, and 22.4% include only one. This question is typically about mother tongue or the main home language and is designed for a single response, meaning that bilinguals are forced to choose one of their languages, leaving the others invisible. The remaining half of European countries include multiple questions that make it possible to identify citizens' multilingual profiles and to describe speakers of migratory languages. The challenge here is that there is a relatively wide variety of questions and combinations of questions, which hinders cross-country comparability and limits the construction of a synthetic perspective on migration-based linguistic diversity.

---

[4] [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736195124](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736195124)

[5] [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736177013&menu=resultados&idp=1254734710990](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177013&menu=resultados&idp=1254734710990)

[6] [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176952&menu=resultados&idp=1254735572981](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176952&menu=resultados&idp=1254735572981)

Given these issues, linguistic information from censuses can still be useful for the demolinguistics of migration in Europe when producing estimates or estimation ranges for particular groups of speakers and specific demolinguistic phenomena. One example is the German microcensus, which makes it possible to analyze responses to the main home language question according to the speakers' country of origin or that of their parents. This information allows the vitality of a language in private use to be examined across generations. In this regard, census and registry data can be used in targeted ways to address specific issues within each country or region. In the case of transnational surveys such as the Eurobarometer, the data do provide homogenized and comparable information for many countries, making it possible to establish an expected range for the demolinguistic quantification of mobile speakers based on the mother-tongue question. To summarize, direct sources can be used to address specific demolinguistic questions related to the particular territories covered by each census; meanwhile, transnational direct sources can be used to generate useful estimates, even though they do not allow for micro-demolinguistic analyzes of speaker profiles.

The quantification of the profiles of speakers of a language with a migratory background in Europe must rely on indirect sources in order to produce estimates based on criteria that allow linguistic profiles to be inferred from sociodemographic data. As shown above, four types of speakers with a migratory background can be analyzed using a combination of national and international sources.

With regard to **migration from Spanish-speaking countries**, the quantification of individuals originating from these countries and residing in Europe—whether or not they hold the nationality of the host country—can be carried out using national or European statistics (such as Eurostat) on the foreign population and on naturalised foreign nationals. For the latter, data are often unavailable, which makes it necessary to produce estimates based on annual naturalisation statistics. The analysis of **their descendants**, regardless of whether they were born in a Spanish-speaking country, in the host country, or elsewhere, is more complex due to the limited availability of data on parents' country of origin in European population statistics such as Eurostat. Only certain national sources provide partial quantitative approximations of their volume and sociodemographic characteristics.

Turning to **European Spanish speakers with a migratory background**, these groups form part of a floating demolinguistic population whose linguistic profiles are shaped by mobility and non-linear migration trajectories. For **immigrants from non-Spanish-speaking backgrounds who have acquired Spanish in Europe**, quantification relies on the same types of sources used for migrants from Spanish-speaking countries. Relevant data come from national statistical institutes—particularly the Spanish INE, which provides detailed information on country of birth, nationality, and migratory flows—and from the statistical offices of Spanish-speaking countries recording Europeans residing in their territories. Together, these sources allow for indirect estimates of Europeans who acquire Spanish through international mobility.

For **Europeans who are not native Spanish speakers but acquire Spanish in migratory contexts**, including return migration and circular mobility, the main available data also come from the INE of Spain, which records outflows, returns, and circular movements. However, these data are limited to the period 2008–2021 and offer only a partial picture of mobility dynamics, thereby restricting the level of detail achievable in demolinguistic profiling.

In all cases, the available sources provide useful yet incomplete insights into multilingual populations whose linguistic trajectories are shaped by increasingly diverse, fluid, and multidirectional forms of mobility across Europe.

The compilation of sources presented in this article makes it possible to determine which source allows for the estimation—albeit with varying degrees of accuracy—of each of the migrant profiles described in Section 1. In addition to demonstrating the potential of available linguistic and demographic sources, their critical evaluation underscores the lack of sensitivity of national accounts in European countries towards migratory languages. The insufficient data on the language repertoires of citizens with migratory backgrounds is a consequence of the prevalence of monolingual linguistic regimes (Gogolin 1997) in most regions of Europe, which are rooted in methodological nationalism and sedentarism bias. Even in countries that recognize themselves as migration destinations, such as Germany, the statistics only collect linguistic data aimed at identifying migrant population groups that may face difficulties communicating in German, posing a potential obstacle to their social integration (Adler 2019).

The linguistic information provided by the sources analyzed in this paper alone is not sufficient for conducting demolinguistic studies of the increasingly diverse and multilingual European societies. To address this issue, sources of information must evolve from static national-ethnic notions to reflect the mobile and complex nature of these dynamic societies. Additionally, when studying migratory languages from the perspective of their populations, the diversity of sources and the complexity of existing linguistic profiles must be considered, in order to avoid oversimplification of statements based solely on demographic statistics.

To achieve a more comprehensive demolinguistics of complexity, it is crucial to connect existing linguistic and socio-demographic data with the life trajectories of individuals. This involves analyzing the processes of acquiring linguistic skills starting from childhood through various centers of linguistic socialization such as family, peer group, and education. By doing so, we can gain a deeper understanding of linguistic diversity and its intricate relationship with individuals' experiences and backgrounds. In 2023, a citizen initiative was launched in Finland requesting that the Population Information System allow multiple languages to be recorded. As in many population registries and censuses across European countries, the Finnish system has so far permitted only one mother tongue to be entered in response to the linguistic question. This illustrates not only a mismatch between the diverse linguistic reality of European societies and the instruments used to collect information about language repertoires, but also a growing public dissatisfaction with the way linguistic diversity is rendered invisible—an invisibility inherited from monolingual habits and ideologies.

## Bibliographic references

- ADLER, A. (2019), Sprachstatistik in Deutschland. *Deutsche Sprache* 3(47). Berlin: Erich Schmidt Press, pp. 197-219.
- BAKEWELL, O. (2007), Keeping Them in Their Place: the ambivalent relationship between development and migration in Africa, *IMI Working Paper 8*. Oxford: International Migration Institute.
- BLOMMAERT, J. & BACKUS, A. (2013), Superdiverse Repertoires and the Individual. In: Saint-Georges, I.d. & Weber, J.J. (Eds.) *Multilingualism and Multimodality. The Future of Education Research*. Rotterdam: SensePublishers. [https://doi.org/10.1007/978-94-6209-266-2\\_2](https://doi.org/10.1007/978-94-6209-266-2_2)
- BUNDESAMT FÜR STATISTIK (2022), *Erhebung zur Sprache, Religion und Kultur*, <https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/erhebungen/esrk.html>
- BUNDESMINISTERIUM FÜR FAMILIE, SENIOREN, FRAUEN UND JUGEND (2024), *Familienreport 2024*. <https://www.bmbfsfj.bund.de/resource/blob/239468/a09d21ecd295be59a9aced5b10d7c5b7/familienreport-2024-data.pdf>
- BUSCH, B. (2016), Categorizing languages and speakers: why linguists should mistrust census data and statistics, *Working Papers in Urban Language & Literacies* 189, pp. 1-18. [WP189 Busch 2016. Categorizing\\_languages.pdf](https://www.urbansociology.com/wp-content/uploads/2016/08/WP189-Busch-2016-Categorizing-languages.pdf)
- ECONOMIC AND SOCIAL COUNCIL OF SPAIN (2019), *Immigration in Spain: effects and opportunities. Report 02/2019*. <https://www.ces.es/documents/10180/5209150/Inf0219.pdf>
- CASTLES, S. (2010), Understanding global migration: A social transformation perspective, *Journal of Ethnic and Migration Studies* 36(10), pp. 1565-1586. <https://doi.org/10.1080/1369183X.2010.489381>
- CLYNE, M. G. (2003), *Dynamics of Language Contact*. Cambridge: Cambridge University Press.
- DEUTSCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG (n.d.). *Socio-Economic Panel*. <https://www.diw.de/de/soep>
- DUCHÊNE, A. & HUMBERT, P. (2018), Surveying languages: the art of governing. Speakers by numbers, *The International Journal of the Sociology of Language* 2018(252), pp. 1-20. <https://doi.org/10.1515/ijsl-2018-0012>
- EUROBAROMETER (2001), *Europeans and languages (EBS147)*. <https://europa.eu/eurobarometer/surveys/detail/202>
- EUROBAROMETER (2006), *Europeans and languages (SP237)*. <https://europa.eu/eurobarometer/surveys/detail/1562>
- EUROBAROMETER (2012), *Europeans and their Languages (SP386)*. <https://europa.eu/eurobarometer/surveys/detail/1049>
- EUROBAROMETER (2024), *Europeans and their languages (SP540)*. <https://europa.eu/eurobarometer/surveys/detail/2979>
- EUROSTAT (2025a), <https://ec.europa.eu/eurostat/web/main/home>
- EUROSTAT (2025b), *Migration and asylum*. <https://ec.europa.eu/eurostat/web/migration-asylum/international-migration-citizenship/methodology>
- EXTRA, G. (2010), Mapping linguistic diversity in multicultural contexts: Demolinguistic perspectives. In: Fishman, J. & García, O. (Eds.) *Handbook of language and ethnic identity*. Oxford, Oxford University Press, pp.107-122.
- EXTRA, G. (2017), The Constellation of Languages in Europe: Comparative Perspectives on Regional and Immigrant Minority Languages. In Kagan, O., Carreira, M.M., & Hitchens Chik, C. (Eds.), *The Routledge Handbook of Heritage Language Education: From Innovation to Program Building* (Routledge Handbooks in Linguistics), New York: Routledge, pp. 11-21.
- EXTRA, G. & GORTER, D. (2008), The constellation of languages in Europe: an inclusive approach. In: Extra, G. & Gorter, D. (Eds.), *Multilingual Europe: Facts and Policies*. Berlin, New York: De Gruyter Mouton, pp. 3-62. <https://doi.org/10.1515/9783110208351.1.3>

FINNISH POPULATION INFORMATION SYSTEMS (n.d.). <https://dvv.fi/en/population-information-system>

GOGOLIN, I. (1997), The 'monolingual habitus' as the common feature in teaching in the language of the majority in different countries, *Per Linguam*, 13(2), 38-49.

GOGOLIN, I. (2002), *Linguistic Diversity and New Minorities in Europe. Guide for the Development of Language Education Policies in Europe: From Linguistic Diversity to Plurilingual Education*. Reference Study, Strasbourg: Council of Europe.

GÓMEZ, L. E. (2022), *Inventing Latinos: A new story of American racism*. The New Press.

ISTITUTO NAZIONALE DI STATISTICA (2015), *I cittadini e il tempo libero (CTL)*. <https://www.istat.it/it/archivio/227593>

KUPISCH, T. (2021), Heritage Languages in Europe. In Montrul, S. & Polinsky, M. (Eds.), *The Cambridge Handbook of Heritage Languages and Linguistics* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press, pp. 45-68.

LEEMAN, J. (2018), Becoming Hispanic: The negotiation of ethnoracial identity in US census interviews. *Latino Studies*, 16, 432-460.

LEEMAN, J. (2020), Los datos censales en el estudio del multilingüismo y la migración: Cuestiones ideológicas y consecuencias epistémicas. *Iberoromania*, 2020(91), 77-92.

LEEMAN, J. (2023), (In)visible identities and inequities: The construction of Latinidad in European censuses. In: Márquez Reiter, R. & Patiño-Santos, A. (Eds.), *Language practices and processes among Latin Americans in Europe*. London: Routledge.

LOUREDA, Ó., ÁLVAREZ MELLA, H., & BLATTNER, C. (2021), La Europa hispanohablante: una mirada demolingüística, In: García Delgado, J.L. (Dir.), *El español lengua internacional: proyección y economía*. Pamplona: Universidad Nebrija, Civitas and Thompson Reuters, pp. 65-83.

LOUREDA, Ó., GUTIÉRREZ PALACIOS, R., ÁLVAREZ MELLA, H., & BLATTNER, C. (2022), La migración hispanohablante en Europa: perfiles demolingüísticos e integración social. *Working Paper, Observatorio Nebrija del Español*, Universidad de Nebrija.

LOUREDA, Ó., MORENO FERNÁNDEZ, F., ÁLVAREZ MELLA, H., & SCHEFFLER, D. (2020), Alemania y sus hablantes de español, In: VV. AA., *El español en el mundo 2020. Anuario del Instituto Cervantes*. Madrid: Instituto Cervantes and Bala Perdida Editorial S. L., pp. 321-354.

MORENO FERNÁNDEZ, F. (2014), Fundamentos de demografía lingüística a propósito de la lengua española. *Revista Internacional de Lingüística Iberoamericana*, Vol. 12, No. 2 (24), pp. 19-38.

MORENO FERNÁNDEZ, F. (2023), *Language Demography*. London: Routledge.

PAUWELS, A. (2016), *Language Maintenance and Shift*. Cambridge: Cambridge University Press.

POULAIN, M. (2008), European migration statistics: definitions, data and challenges. In: Barni, M. & Extra, G. (Eds.), *Mapping Linguistic Diversity in Multicultural Contexts*. Berlin, New York: Mouton de Gruyter, pp. 43-66.

SCHWENKE, H. (2018), *Globale Migration zur Einführung*. Hamburg: Junius.

STATISTISCHES BUNDESAMT (2025), Bevölkerung nach Migrationshintergrund, Alter, Geschlecht  
Ersterggebnisse  
Tabelle  
12211-01.  
<https://www.destatis.de/DE/Themen/GesellschaftUmwelt/Bevoelkerung/Migration-Integration/Tabellen/liste-migrationshintergrundgeschlecht.html#116670>.

UNITED NATIONS (2017), Principles and Recommendations for Population and Housing Censuses. Online: [https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles\\_and\\_Recommendations/Population-and-Housing-Censuses/Series\\_M67rev3-E.pdf](https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf) [last accessed 30/09/2025]

UNITED NATIONS (2022), Migrant definition. Online: <https://refugeesmigrants.un.org/definitions> [last accessed 30/09/2025].

UNITED NATIONS STATISTIC DIVISION (n.d.), *World Population and Housing Census Programme*. <https://unstats.un.org/unsd/demographic-social/census/document-resources/>

WIMMER, A. & GLICK SCHILLER, N. (2003), Methodological Nationalism, the Social Sciences, and the Study of Migration: An Essay in Historical Epistemology. *International Migration Review*, 37(3), 576-610.

Annex 1. Operationalization of language questions in 32 EU Countries (expansion based on the table by Extra & Gorter 2008, 20)

EU countries	Mother tongue	(Other) language(s) spoken (frequently)	Language(s) (most frequently) spoken at home	Language(s) spoken with family and friends/ at work/ at school	Speak well/ average/ a little	Understand/ Speak/ Read/ Write
Albania (2011)	+	-	-	-	-	-
Bosnia and Herzegovina (2013)	+	-	-	-	-	-
Bulgaria (2011)	+	-	-	-	-	-
Croatia (2021)	+	-	-	-	-	-
Czech Republic (2021)	(1)	-	-	-	-	-
Finland (2021)	+	-	-	-	-	-
Romania (2021)	+	-	-	-	-	-
Serbia (2021)	+	-	-	-	-	-
Slovakia (2021)	+	-	-	-	-	-
Austria (2011)	-	+	-	-	-	-
Germany (2021)	-	+	-	-	-	-
Armenia (2011)	+	+	-	-	-	-
Azerbaijan (2009)	+	+	-	-	-	-
Cyprus (2021)	+	+	-	-	-	-
Estonia (2011)	+	+	-	-	-	-
Georgia (2014)	+	+	-	-	-	-
Lithuania (2020)	+	+	-	-	-	-
Ukraine (2012)	+	+	-	-	-	-
Moldova (2014)	+	+	-	-	-	-
North Macedonia (2021)	+	-	+	-	-	-

Annex 1. (continued)

EU countries	Mother tongue	(Other) language(s) spoken (frequently)	Language(s) (most frequently) spoken at home	Language(s) spoken with family and friends/ at work/ at school	Speak well/ average/ a little	Understand/ Speak/ Read/ Write
Slovenia (2002)	+	-	+	-	-	-
United Kingdom (2021)	+	-	-	-	+	(2)
Kazakhstan (2009)	+	-	-	-	-	+
Malta (2021)	+	-	-	-	-	(3)
Ireland (2022)	-	(4)	+	-	-	-
Russia (2011)	-	+	-	-	-	+
Belarus (2009)	+	(5)	+	-	-	-
Hungary (2022)	(1)*	(6)	-	(1)*	-	-
Montenegro (2023)	+	+	-	-	-	+
Luxembourg (2011)	(7)	-	+	+	-	-
Switzerland (2022)	(7)	-	+	+	-	-
Spain (2021)	(8)	+	-	+	-	+

(1) You can state two

(2) For official languages other than English in Northern Ireland, Scotland, and Wales

(3) Are you able to read and write a short, simple sentence in at least one language?

(4) Only Irish; if yes, daily within/outside the educational system/weekly/less often/never

(5) Which other languages do you use freely

(6) What languages do you speak or understand? In what languages can you make yourself understood?

(7) In which language do you think and know best

(8) Do you remember which language you spoke in the first place, at home, as a child. Mark all that apply

\* Optional