

Lori Repetti

## 4 Fieldwork and building corpora for endangered varieties

**Abstract:** Loss of linguistic diversity is viewed by many as one of the great ecological disasters of the twenty-first century, and the Romance language family has not been spared. This chapter deals with unique challenges to the study and documentation of endangered Romance languages. We consider the question of language vs dialect, and the added problems faced by endangered varieties deemed “dialects”. The role played by the highly prestigious and structurally related national languages of the countries in which the endangered varieties are spoken is analysed within the fieldwork context, as speakers may possess a spectrum of linguistic abilities, from the national standard to an archaic local variety. Fieldwork methods and language documentation/description are discussed, along with the types of resources produced and their accessibility.

**Keywords:** endangered Romance languages, “language” vs “dialect”, fieldwork, documentation, revitalization

### 1 Introduction

While linguistic diversity is diminishing across the globe, the situation is especially critical in Europe where education levels are high, and knowledge of the standardized languages is widespread. According to information retrieved from Ethnologue’s list of endangered languages and UNESCO’s *Atlas of the World’s Languages in Danger*, dozens of endangered Romance languages are found throughout Europe, but also in North and South America, Africa, and Asia.<sup>1</sup> The study of these endangered languages and the accessibility of data from them are particularly important and urgent. Unlike artifacts, language cannot be preserved in its natural form once there are no native speakers. It can only be preserved in written form and, more recently, in audio recordings.

---

**1** The question of what counts as an endangered language can vary from context to context, and there are different types of endangerment and death (see Tsunoda 2005, 36–48). For reasons of space, I do not include endangered Romance-based creoles, such as Palenquero (a Spanish-based creole spoken in Colombia) and some dialects of Chavacano (a Spanish-based creole spoken in the Philippines). Other endangered Romance varieties that are not addressed include non-standard spoken varieties of the standard language, such as popular Brazilian Portuguese with its unique grammatical traits (Guy/Zilles 2008, 55). Barbiere (2015) argues that endangered dialects should be included in discussions of endangered languages, and Tsunoda (2005, 5–6) discusses the differences between language death and dialect death. See also section 2.

<https://doi.org/10.1515/9783110365955-005>

This chapter deals with fieldwork and documentation of endangered Romance languages.<sup>2</sup> The following Table provides a list of relevant languages. The information provided must be interpreted with care as it comes from only two sources – Ethnologue and UNESCO – and the calculation of number of speakers varies amongst these. The information is meant to provide an overview and rough estimate of the number of endangered Romance languages, the number of speakers, and the “degrees of endangerment” of each variety.

**Table 1:** Endangered Romance languages

Language	Location	Degree of Endangerment	# of Speakers	Degree of Endangerment	# of Speakers
		<i>Ethnologue</i>	<i>Ethnologue</i>	<i>UNESCO</i>	<i>UNESCO</i>
Alpine-Provençal/ Vivaro-Alpine	France			Definitely Endangered	~200,000
Aragonese	Spain	6b. Threatened	~10,000/ ~20,000 L2	Definitely Endangered	~10,000
Aromanian	Greece, Macedonia, Albania	6b. Threatened	~114,340	Definitely Endangered	~114,340
Asturian/ Astur- Leonese	Spain, Portugal	6b. Threatened	~110,000	Definitely Endangered	~150,000
Burgundian	France			Severely Endangered	
Cajun French	United States (Louisiana)	7. Shifting	~25,600		
Campidanese Sardinian	Italy	6a. Vigorous	~500,000	Definitely Endangered	~900,000
Champenois	France, Belgium			Severely Endangered	
Corsican	France	4. Educational	~31,000	Definitely Endangered	~160,000
Emilian, Romagnol	Italy	9. Dormant		Definitely Endangered	~2,000,000
Extremaduran	Spain	7. Shifting	~201,500		

<sup>2</sup> See Tsunoda (2005) for an introduction to conducting fieldwork among endangered languages.

Language	Location	Degree of Endangerment	# of Speakers	Degree of Endangerment	# of Speakers
		<i>Ethnologue</i>	<i>Ethnologue</i>	<i>UNESCO</i>	<i>UNESCO</i>
Faetar	Italy			Definitely Endangered	~600
Fala	Spain	6a. Vigorous	~10,500		
Franc-Comtois	France, Switzerland			Severely Endangered	
Franco-provençal/Arpitan	France, Switzerland, Italy	8a. Moribund	137,000	Definitely Endangered	~100,000
Friulian	Italy	4. Educational	~300,000	Definitely Endangered	~600,000
Gallo	France			Severely Endangered	~200,000
Gallo-Sicilian	Italy			Definitely Endangered	~65,000
Gallurese Sardinian	Italy	6b. Threatened	~100,000	Definitely Endangered	~100,000
Gascon	France, Spain	(see Occitan)		Definitely Endangered	~250,000
Guernésiais	British Crown dependency	8b. Nearly Extinct	~200	Severely Endangered	~1,327
Istriot	Croatia	7. Shifting	~400/~900 L2	Severely Endangered	~400/~900 L2
Istro-Romanian	Croatia	7. Shifting	~300/~1,100 L2	Severely Endangered	~300/~1,100 L2
Jèrriais/Jersey French	British Crown dependency	8a. Moribund	~1,920	Severely Endangered	~2,000
Judeo-Italian/Corfiot Italkian	Italy, Greece	8a. Moribund	~250	Critically Endangered	
Judezmo/Ladino/Judeo-Spanish	Israel, Turkey, Greece	4. Educational	~112,130	Severely Endangered	
Ladin	Italy	6b. Threatened	~31000	Definitely Endangered	~31,000
Languedocian	France			Severely Endangered	~500,000
Ligurian	Italy, Monaco, France	5. Developing	~505,100	Definitely Endangered	~1,000,000

Language	Location	Degree of Endangerment	# of Speakers	Degree of Endangerment	# of Speakers
		<i>Ethnologue</i>	<i>Ethnologue</i>	<i>UNESCO</i>	<i>UNESCO</i>
Limousin	France			Severely Endangered	~400,000
Logudorese Sardinian	Italy	6b. Threatened	~500,000	Definitely Endangered	~400,000
Lombard	Italy, Switzerland	6a. Vigorous	~3,903,000	Definitely Endangered	~3,500,000
Loreto-Ucayali (Amazonic) Spanish	Peru	6a. Vigorous	~2,800		
Lorrain	France, Belgium			Severely Endangered	
Megleno-Romanian/Meglenitic	Greece, Macedonia	7. Shifting	~5,000	Severely Endangered	~5,000
Minderico	Portugal	8b. Nearly Extinct	~500		
Napoletano-Calabrese/South Italian	Italy	5. Developing	~5,700,000	Vulnerable	~7,500,000
Norman	France			Severely Endangered	
Occitan/Gascon/Auvergnat	France, Italy, Monaco, Spain	6b. Threatened	~218,310	Severely Endangered	
Picard	France, Belgium	5. Developing	~200,000	Severely Endangered	~700,000
Piedmontese	Italy	5. Developing	~1,600,000	Definitely Endangered	~2,000,000
Poitevin-Saintongeais	France			Severely Endangered	
Romansh	Switzerland	4. Educational	~40,039	Definitely Endangered	~35,095
Sassarese Sardinian	Italy	6b. Threatened	~100,000	Definitely Endangered	~120,000
Sicilian	Italy	5. Developing	~4,700,000	Vulnerable	~5,000,000
Venetian/Venetan	Italy, Croatia	5. Developing	~3,852,500	Vulnerable	~4,000,000

Language	Location	Degree of Endangerment	# of Speakers	Degree of Endangerment	# of Speakers
		<i>Ethnologue</i>	<i>Ethnologue</i>	<i>UNESCO</i>	<i>UNESCO</i>
Walloon	Belgium, France	6b. Threatened	~600,000	Definitely Endangered	~600,000

---

#### Degree of Endangerment

---

##### *Ethnologue*

---

4. Educational: The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.

---

5. Developing: The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.

---

6a. Vigorous: The language is unstandardized and in vigorous use among all generations.

---

6b. Threatened: The language is used for face-to-face communication within all generations, but it is losing users.

---

7. Shifting: The child-bearing generation can use the language among themselves, but it is not being transmitted to children.

---

8a. Moribund: The only remaining active users of the language are members of the grandparent generation and older.

---

8b. Nearly Extinct: The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.

---

9. Dormant: The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.

---



---

#### Degree of Endangerment

---

##### **UNESCO**

---

Vulnerable: most children speak the language, but it may be restricted to certain domains (e.g., home).

---

Critically Endangered: the youngest speakers are grandparents and older, and they speak the language partially and infrequently.

---

Definitely Endangered: children no longer learn the language as mother tongue in the home.

---

Severely Endangered: language is spoken by grandparents and older generations; while the parent generation may understand it, they do not speak it to children or among themselves.

---

Romance varieties are increasingly becoming endangered for the same reasons that non-Romance languages become endangered: urbanization, education, increased mobility, social media, increased access and allegiance to standard varieties, language contact, tourism, etc. (Austin/Sallabank 2011). In this article, we will address

problems that present unique challenges to the study and documentation of endangered Romance languages. In particular, we will consider the age-old question of “language” vs “dialect” within this context, and the added challenges faced by endangered varieties deemed “dialects”. Intimately connected to this issue is the role played by the highly prestigious and structurally related national languages of the countries in which the endangered varieties are spoken (section 2). The response to endangerment in the Romance-speaking world (if there is an organized response at all) has been varied, but efforts to promote literacy in the endangered variety have been hampered by the extreme variation among local varieties and, in many cases, the lack of a single variety to identify for standardization (section 3; see also ↗3 Collecting and analysing creole data). The core of this article deals with documentation efforts and challenges to documentation (section 4).

## 2 “Language” vs “dialect”

We often think of endangered languages as varieties spoken in isolated, perhaps inaccessible, and poor communities. This is generally not the situation found within the Romance context. What we do find in the Romance-speaking world, in Europe in particular, is a myriad of Romance varieties which are direct descendants of Latin, but which do not enjoy political and social prestige. This situation is most dramatic in Italy where Romance varieties blend from one to another with few sharp divisions distinguishing one variety from the next. For political and historical reasons these varieties are commonly referred to as “dialects”.<sup>3</sup> For example, Emilian is the direct descendent of the Latin spoken in the Emilia region of northern Italy, which has as many sub-variations as there are towns in the region. It is structurally strikingly different from Italian, and it is not mutually intelligible with Italian. Since it is neither a socially prestigious nor a politically protected variety, most children raised in Emilia do not learn to speak the local variety, resulting in its status as a “dormant” language (according to *Ethnologue*) or a “definitely endangered” language (according to UNESCO).

What is the difference between a “language” and a “dialect”?<sup>4</sup> There are at least two dimensions to the distinction relevant here: one linguistic, and the other socio-politico-historic. The linguistic metric of mutual intelligibility distinguishes “dialects of a language” from “separate languages”: in short, if speakers of different varieties can understand each other, we refer to those varieties as “dialects” of the same language; if speakers of different varieties cannot understand each other, we refer to those varieties as separate “languages”. However, this is complicated by varying

---

<sup>3</sup> See Loporcaro (2013) for an introduction to the linguistic situation of Romance varieties spoken in Italy, and Cravens (2014) for an excellent overview of the distinction between language and dialect in Italy, and the implications for endangered varieties. See also ↗18 The languages and dialects of Italy.

<sup>4</sup> See Hinskens/Auer/Kerswill (2005) for a discussion of the definition of “dialect”.

degrees of mutual intelligibility and by socio-politico-historic considerations. The latter factors are neatly summed up in the well-known quip that “a language is a dialect with an army and a navy”. The result is that the use of the term “language” or “dialect” to refer to a particular variety is the result of many considerations (see footnote 1).

According to the linguistic metric, some so-called “dialects” should be considered separate “languages”, and some Romance varieties recognized as separate “languages” might be considered “dialects” of the same language. For example, we find an imperfect mapping with reference to the “dialects” in Italy, as illustrated in the case of Emilian above: although Emilian is considered a “dialect” of Italian, Italian speakers would not be able to understand Emilian. Similarly, Portuguese encompasses such extreme variation that speakers of Portuguese from different areas might not understand each other: Portuguese speakers from Brazil might have difficulty understanding speakers from Portugal. The categorization of Emilian as a “dialect” of Italian and of Brazilian Portuguese and European Portuguese as Portuguese “dialects” is based on socio-political and historical considerations.

On the other hand, Spanish and Portuguese are clearly considered separate languages despite the fact that speakers of Spanish might be able to understand a good portion of Portuguese, and vice versa. Similarly, Piedmontese (spoken in Italy) is associated with Italian, even though it is grammatically more similar to Occitan (spoken in France), and Corsican (spoken in France) shares many grammatical features with Italian, but most lay people would not associate it with Italian.<sup>5</sup> Here, too, socio-political and historical factors determine the classification.

The distinction between “language” and “dialect” is not simply a terminological one, but has profound implications for language vitality: those varieties recognized as “dialects” could be ineligible for the resources available for endangered “languages”. In particular, “dialects” are not protected by the *Charter for Regional or Minority Languages* (see 721 Revitalization and education). The Charter, created by the Council of Europe in 1992, was designed to protect and promote minority languages (which are largely endangered) that are “traditionally used within a given territory of a state by nationals of that state who form a group numerically smaller than the rest of the state’s population”, and it explicitly excludes “dialects of the official language(s) of

---

5 Another complication arises in the cases of Romance varieties spoken in non-Romance-speaking countries, such as Aromanian varieties spoken in Greece, Bulgaria, Serbia, the Republic of Macedonia, and Albania: are they variants of Romanian, or do they belong to a separate language (Carageani 2002)? This situation is not unique to the Romance world. In Sweden and Norway, for example, the issue arises regarding the Finno-Ugric varieties: are they dialects of Finnish or independent languages? (see Huss 2008).

the state” ([conventions.coe.int/treaty/en/Treaties/html/148.htm](http://conventions.coe.int/treaty/en/Treaties/html/148.htm)).<sup>6</sup> Hence, if a variety is deemed a “dialect” it does not receive the protection afforded other “languages”. The categorization of a variety as a “language” or “dialect” is left up to each member state, resulting, in some cases, in curious situations. For example, Italy, which signed the Charter in 2000 but has not yet ratified it, classifies some indigenous Romance varieties as “languages” and, therefore, covered by the provisions of the Charter, while others are “dialects”. For example, Sardinian is a “language”, while Sicilian is a “dialect”. Linguistically, there is no reason to consider Sicilian more of a dialect of Italian than Sardinian is, or, alternatively, there is no principled reason to consider Sardinian more independent of Italian (hence not a dialect) than Sicilian; the classification was somewhat arbitrary and historico-politically motivated.

The relationship between a variety’s categorization as a “language” or a “dialect” has a profound effect on the responses to its endangered status, and, in particular for the purposes of this chapter, on its documentation.

### 3 Responses to endangerment/revitalization

In all contexts of language endangerment, we usually see three types of responses from interested parties (native speaker community, politicians, intellectuals, lay people, linguists, etc.): non-interference, efforts in maintenance and revitalization, and documentation, the focus of this chapter (Romaine 2008, 8). All three responses have been adopted in various Romance contexts.

#### 3.1 Non-interference

Non-interference is perhaps the dominant response, and some of the most endangered Romance varieties have no overt revitalization efforts underway. For example, Sercquiais (spoken on the island of Sark) has no media presence, no learning/teaching opportunities, and no extra-curricular initiatives aimed at revitalizing the language (Jones 2014). The fact that, for many endangered Romance languages, there is little or no information available on any response to endangerment suggests this to be the situation in most cases.

Some groups, such as administrators, educators, or members of the community, may actually welcome language death, arguing that local languages are not useful in the modern world, that language diversity reduces intercultural communication, and that revitalization efforts are a waste of effort and money (Crowley 2007; Sallabank

---

<sup>6</sup> The Charter also describes the objectives and principles that ratifying states must follow to protect minority languages, and how states can assure the rights of minority languages.



2013; Tsunoda 2005). Typical of this position is the following politician's comments on Guernésiais (spoken on the island of Guernsey), reported by Sallabank (2013, 111): "there are some who don't think it's progressive, why should we go back, I've had one or two 'why should we waste all our time in learning about Guernsey French?'..." These negative language attitudes grow out of negative experiences of Guernésiais children who had "unhappy experiences at school" (Sallabank 2002, 220) and overall felt ashamed of their language, making them "less likely to transmit the language to their children" (Sallabank 2010a, 70). Governments also expedite the demise of endangered languages by not supporting revitalization efforts, often justifying this by characterizing endangered varieties as "dialects" (as opposed to "languages"), thereby rendering them ineligible for certain types of support (see section 2 above).

This type of attitude is portrayed in a series of popular videos produced in 2010 by RAI (Italy's national public broadcasting company) to celebrate the 150<sup>th</sup> anniversary of the unification of Italy. In the fifteen short skits, people speak a "dialect" with another person who does not understand and appears perplexed and confused. The announcer then proclaims, "Se gli italiani fossero quelli di 150 anni fa, probabilmente comunicherebbero ancora così". The reaction to this portrayal of the dialects as incomprehensible was largely negative, and perhaps partly in response to that reaction, the series of advertisements put out recently by Nutella have a very different approach to the dialects, celebrating their diversity and cultural importance, and providing material online of representative varieties.

### 3.2 Revitalization efforts

Many groups in the Romance world are actively involved in language maintenance and revitalization efforts.<sup>7</sup> These efforts range from the creation of resources for learning and teaching endangered languages (for Occitan, see [http://www.crdp-montpellier.fr/languesregionales/occitan/ressources/sceren/dire\\_en\\_oc.html](http://www.crdp-montpellier.fr/languesregionales/occitan/ressources/sceren/dire_en_oc.html); for the use of new technologies for teaching and learning endangered languages, cf. Hugo 2015), to the promotion of literature in the endangered variety (for an Aragonese literary contest, the "Concurso de narratiba e poesía en aragonés", see <https://cultura.unizar.es/concurso-de-narratiba-e-poes%C3%ADa-en-aragon%C3%A9s>), the use of the endangered variety in various media outlets (for example, Belgian state TV and radio provide a few hours of broadcasting in Walloon weekly), and courses and programmes designed to teach about the structure of endangered languages (for example, a course on Piedmontese was offered at the "Languages and Linguistics of the Mediterranean" 2016 summer school: <http://llm.unica.it/events/piedmontese>). An innovative approach to funding revitalization projects – crowdfunding – is described on the "Language Endanger-

---

<sup>7</sup> See Wolfram (2008) for outreach efforts to support language diversity in general.

ment: Revitalizing Minderico” webpage (<https://hubbub.org/p/minderico>). While these efforts are undertaken as proactive steps to language maintenance, some argue that they are nothing more than a requiem for the soon-to-be dead languages (Tamburelli n.d.). These initiatives, argues Tamburelli (n.d.), position the language as a museum piece destined to remain in theaters, poems, and the classrooms of those who study their history, but not to thrive as spoken and living languages.

One of the unique and most exciting characteristics of the Romance context actually turns out to be one of its weaknesses in revitalization efforts. Since there is so much variation among endangered Romance languages, it is not always clear which variety should be the focus for standardization and revitalization. In other words, promoting standardization may actually hurt the vitality of the local varieties, perhaps even creating a prestige hierarchy where one did not exist. This situation is not unique to Romance, and is also found, for example, in Ireland where the creation of a written standard “has also led to a diglossic situation for the varieties of Irish in the Gaeltacht, where spoken Irish takes as its basis the regional dialect, while all forms of written language tend toward the standard, as this is what is to be found in textbooks and in most published material” (Ó hIfearnáin 2008, 125).<sup>8</sup>

The situation with Romansh (also spelled Romantsch, Rumantsch, Romonsch), known for its extremely rich local variation, is illustrative. The need for a single pan-dialectal standard has been felt at least since the nineteenth century (Williamson 1991, 54);<sup>9</sup> however, it was not until the late twentieth century that such a standard, Rumantsch Grischun, was created from the many local varieties. Perhaps because of its artificial and hybrid nature, Rumantsch Grischun has met with “very considerable resistance” (Anderson 2016, 169). Romansh is, therefore, “unusual in being endangered both from without (by German) and from within (by an artificial standard perceived to have minimal relevance or utility)” (Anderson 2016, 169).

### 3.3 Documentation

The third response to endangerment involves documentation of the endangered language.<sup>10</sup> In section 4 we will consider some of the unique challenges to Romance

---

**8** In his discussion of Irish, Ó hIfearnáin (2008, 127) recommends that “creators of a national language policy should seek a compromise that would reinforce intergenerational transmission of the local variety through schooling so as to avoid conflict in the target variety and to encourage community language development”.

**9** The Lia Rumantscha was founded in 1919 to address the growing threat of German as the *lingua franca* in the Romansh-speaking areas of Switzerland (Williamson 1991).

**10** Some online resources to assist in documenting endangered languages include the following: *An Crúbadán* (corpus-building for minority languages) (<http://crubadan.org>, last access 18.02.2018); *Endangered Languages Archive* (preserving and publishing documentation on endangered languages)

language documentation during fieldwork and in working with speakers, and then look at innovative use of modern technological tools in documentation and revitalization of endangered Romance languages.<sup>11</sup>

First, a word on what documentation involves and how documentation differs from description (Austin 2010). Language documentation is “discourse-centered” (Austin 2010, 20) and “aims to record the linguistic practices and traditions of a speech community, along with speakers’ metalinguistic knowledge of those practices and traditions” (Austin 2010, 18). It is multidisciplinary in nature, and crucially involves the active participation of the speech community in all aspects of the documentation process: the collection, analysis, and preservation of many types of language data to be made available for a wide range of users. Description, or the collection and analysis of linguistic data which are made available in grammars, dictionaries, specialized articles, etc., is just one component of documentation. “Documentation” and “Description” differ in their goals, their methodology, and the nature and role of data and metadata.

## 4 Documentation

### 4.1 Fieldwork

Documenting the great linguistic variety found in the Romance-speaking world has a long and illustrious tradition. Perhaps the earliest recorded attempt to register the plethora of Romance varieties can be found in Dante’s early fourteenth-century *De vulgari eloquentia*. This work, as well as those that followed in the next five centuries, had a literary bent: Dante was searching for an “eloquent vernacular”, while subsequent scholars documented local variants of popular stories such as the first novella of the ninth day of the *Decameron*, the Parable of the Prodigal Son, the Lord’s Prayer, the Gospel of Matthew.<sup>12</sup> Documenting the lexicon of these languages through the translation of words and phrases in the national language picked up in the eighteenth century with the compilation and publication of numerous dialect dictionaries. This evolved into a more scientific study of Romance lexical variation and historical

---

(<http://www.elar-archive.org/index.php>, last access 18.02.2018); *Endangered Languages Project* (technology for those working to document, preserve, and teach endangered languages) (<http://www.endangeredlanguages.com>, last access 18.02.2018); *Open Language Archives Community* (network of language archives) (<http://www.language-archives.org>, last access 18.02.2018); *Phoible* (repository of phonological inventories from languages around the world) (<http://phoible.org>, last access 18.02.2018); etc. See also Assini (2014); Jones (2015), etc.

**11** For more information on documenting endangered languages, see the many excellent articles in the journal *Language Documentation and Description*, Austin/Sallabank (2011), Jones (2015), etc.

**12** See Pop (1950) for a thorough review of the history of the study of Romance variation.

phonology in the late nineteenth century with groundbreaking work by Jules Gilliéron with his *Atlas linguistique de la France* (Gilliéron/Edmont 1902–1910), and by Graziadio Isaia Ascoli with his new journal the *Archivio glottologico italiano* (Ascoli 1873) whose goal is to promote “l’esplorazione scientifica dei dialetti italiani ancora superstiti” (Ascoli 1873, xxv). Note that Ascoli was already aware of the endangered status of these varieties (see also 75 Romance dialectology).

These early methodologies (recording oral literature and conducting translation tasks) continue to be used today. Oral literature (myths, stories, nursery rhymes, lullabies, songs, etc.) is argued to be a particularly good source of data since it is “conceived in the language and not the product of translation or interpretation” (Bouquiaux/Thomas 1992, 56), and many community-based and non-professional efforts in language documentation include the publication of songs, poems, sayings, proverbs, etc. in the endangered language. However, this type of data has its limitations as a source of grammatical information (for example, some grammatical structures crystallized in oral literature are no longer productive), of sociolinguistic information (for example, the oral literature may not be widely known, and the information communicated may no longer be relevant in today’s society), and may be viewed as reflecting a stage of the community’s history that speakers no longer want to be associated with (Mosel 2006). Translation tasks continue to be the principal means of eliciting data for lexical and grammatical studies, including investigations of phonology (for example, to establish inventories), morphology (for example, to record paradigms), syntax (for example, to study variations in word order), etc. Problems with this methodology include the structural similarity between the endangered language and the language used during the task (which is usually the national standard, a closely related Romance language) which can lead to responses which conform to the structure of the national standard (see section 4.2).

Other common methodologies within the Romance domain include the use of questionnaires or semi-structured elicitations, natural conversations, and non-structured interviews, which provide more sociolinguistic and pragmatic information. Additional advantages of spontaneous speech data for grammatical investigations include natural variation in pronunciation and syntax, and the avoidance of the problems associated with artificial speech produced in a laboratory setting or controlled context. Drawbacks include variations in speech rate which make comparisons difficult, and the lack of control over the structures produced.

While most early studies focused on lexical variation and historical phonology, in the past few decades emphasis has shifted to other aspects of the language (syntax, sociolinguistics, etc.) and to new methodologies. Endangered Romance languages are particularly important in the field of micro-variation, or the study of minor differences among closely related languages in order to better understand the constraints on language variation and the ways in which a grammar can change over time. The great variation attested among endangered Romance languages provides us with as close to a natural laboratory to study language variation and change as we can ever hope to

find. The study of syntactic micro-variation has found particularly fertile ground in the Romance domain, and investigations of syntactic micro-variation within French varieties illustrate the usefulness of this approach (<http://blogs.univ-tlse2.fr/symila>).

Innovative and experimental approaches to fieldwork on endangered Romance varieties began decades ago and continue today to enrich the field of linguistics. Uguzzoni (1971) was the first to apply the tools of phonetics to carry out an acoustic analysis of Emilian vowels (work which shed light on the evolution of Latin vowels), and now phonetic studies of endangered Romance languages abound. Endangered languages have been the subject of numerous neurolinguistic studies, including studies of the syntax of subject clitic pronouns of speakers with non-fluent aphasia (Chinellato 2004). A very recent innovation in fieldwork involves crowdsourcing language data, for example, for Abruzzese and Molisano data (<http://www.abruzzesemolisano.it>).

Investigation of the sociolinguistic situation of an endangered language community, including language use and attitudes, requires different tools: self-evaluating questionnaires, observations by investigators, grammaticality tests to check sensitivity to language interference, etc. Heritage varieties spoken by the descendants of emigrants (Peyton et al. 2008), such as the varieties of Italo-Romance spoken in the United States (Haller 1993; Tortora 2014), Walloon in Wisconsin (NPR 2015), Friulian in Romania (Iliescu/Melchior 2015), Occitan in North Carolina (Pons 1990), or Veneto in Mexico (Sartor/Ursini 1983; Mackay 1992; Barnes 2009) are often the object of sociolinguistic studies.<sup>13</sup> Intensive fieldwork with speakers of a Veneto variety spoken in Chipilo, Puebla, Mexico (Ursini 1988; Mackay 1992; Barnes 2009) has shown that language maintenance has been successful despite the lack of external support because of the social isolation of the community and because the language has come to symbolize ethnic solidarity and group identity.

The Romance varieties of the Channel Islands of Jersey, Guernsey and Sark (Jèrriais, Guernésiais, and Sercquiais, respectively) are extremely endangered varieties: estimates vary widely, but sources agree that fewer than 3000 people speak Jèrriais, no more than 1500 speak Guernésiais, and there are a mere few dozen speakers of Sercquiais (States of Jersey Statistics Unit 2012; Jones 2014; Warren/Jennings 2015; UNESCO; *Ethnologue*). The sociolinguistic situation of these varieties is particularly well studied (Jones/Bulot 2009), and we know that, for example, Guernésiais is predominantly used in domestic environments, that it is no longer being passed on to children, and that code-mixing with English is frequent (Sallabank 2010b); these are all situations that bode very badly for a language's vitality.

---

**13** Endangered languages are often found in communities of recent immigrants, especially in large urban centres, such as New York City, where an estimated 800 languages are spoken, many of which are endangered. A unique organization in New York City, the Endangered Language Alliance (<http://elalliance.org>, last access 18.02.2018), documents and describes the endangered languages spoken there, including many Romance languages.

While the methods used for gathering data from endangered languages are the same as those recommended for healthy languages, there are some unique issues that arise in the documentation of endangered languages. Some of those issues do not present themselves as problematic for endangered Romance language fieldwork, such as the issue of a common language, but others are particularly problematic within Romance, such as the limited demographics of speakers, which is the topic of the next section.

## 4.2 Speakers

Researchers working on endangered languages try to work with informants of both genders, of different ages, and with high levels of proficiency;<sup>14</sup> however, the demographics of speakers of endangered languages can be quite limited, and the choice of an informant presents unique challenges in the Romance context.<sup>15</sup>

Fluent speakers of endangered Romance languages are usually elderly (a rule of thumb is that those born before WWII are the most fluent, with fluency levels dropping among those born from the mid-1940s onward). This could make it particularly difficult to find informants. For example, a linguist from Stony Brook University, Francisco Ordóñez, had difficulty locating speakers of Gascon in France, but had success in an assisted living facility in Aran (Spain) where he was able to find a number of fluent speakers of Gascon (Ordóñez p.c.). Furthermore, some elderly informants might have speaking difficulties due to missing teeth or cognitive impairment, and the most elderly might have low levels of literacy, rendering certain types of tasks difficult, if not impossible, and further limiting access to information. Given the small number of speakers of some varieties, and the lack of communities of speakers, fieldwork involving observation of speakers can be nearly impossible.

A more uniquely Romance problem involves the language variety the informant uses. Since these endangered varieties are related to the national standards, speakers may have a range of linguistic abilities, from the national standard to an archaic local variety (see also footnote 15). As illustrated in Loporcaro (2013, 6–7) speakers of a southern Italian variety might have a repertoire that includes standard Italian, regional Italian (i.e., a version of the standard language with locally identifiable phonological and lexical characteristics), and the indigenous, local Romance variety (i.e., a variety with unique lexical items, syntactic structures, morphological characteristics,

---

**14** The structural changes found in language obsolescence constitute an interesting and growing area of research (Dorian 1992). In these studies, the grammar of speakers who might not be considered fluent in the most archaic variety of the language is the object of investigation. A common characteristic within the Romance realm is convergence between the local variety and the closely related “roof” language, resulting in a language continuum (Repetti 2014; Cerruti/Regis 2015).

**15** For the typology of speakers of endangered languages, see Grinevald/Bert (2011).

etc.). Because of the close genetic relationship among these varieties, speakers can move seamlessly across the spectrum of their languages. Informants may be reluctant to admit that they know a variety of the local language which is very different from the national language, since the local languages do not enjoy the high social prestige of the standard language. Hence, in different interview contexts, an informant might use different versions of his/her languages. With a researcher from the “inside” (for example, the same country), an informant might use a variety cleansed of its most local traits, so as to not appear too provincial. Alternatively, with a researcher from the “outside” these prejudices might not be felt so acutely. The opposite situation is also attested. The presence of other speakers of the endangered language can similarly have the effect either of encouraging use of the local variety, or of dictating the use of a more regionally neutral variety.

### 4.3 Accessibility of data

The data that result from field research on endangered languages are valuable in all formats, from primary data (recordings, field notes, etc.) to the analysis of those data available in grammars, articles, dictionaries, etc. (Tsunoda 2005, 245–247). Endangered Romance languages have been studied for many decades, and data in many formats are abundant: dictionaries, grammars, descriptions, atlases, collections of texts, monographs, etc. There are gaps in the data, for example, annotated audio and video files are not available for most varieties, but the situation is indeed improving. Much of the progress in this area is made by the speech communities themselves. Audio and video recordings are available on many websites, and groups on social media regularly post information about their language. The issue addressed here is the use of digital resources (websites, databases, etc.) to make the data more widely accessible to two audiences: the community of speakers and the community of linguists.

Tools for community-oriented initiatives need to be user-friendly and contain information for a wide range of interests. Hundreds of such tools exist for communities of speakers of endangered Romance languages. An excellent example can be found for Francoprovençal (*Patois*) (<http://www.patoisvda.org/>). This site contains information accessible to a wide audience on many aspects of culture and language, including resources for promoting and learning Francoprovençal and access to the “guichet linguistique” for assistance in these efforts. Evidence suggests that new resources such as these may indeed be helping the plight of endangered languages (Warren/Jennings 2015, 140).

New technologies have also facilitated the work of professional linguists. Digital resources, though less durable, are more “portable” and accessible than older technologies (printed format). For example, the *Atlante Italo-Svizzero* (Jaberg/Jud 1928–1940), an invaluable but cumbersome eight volume linguistic atlas, is now available in digital format (*NavigAIS*: [www3.pd.istc.cnr.it/navigais](http://www3.pd.istc.cnr.it/navigais)), making access to it much

simpler. The availability of large amounts of data from many languages, in the form of linguistic corpora or databases, has increased our understanding of human language. By accessing large amounts of organized and searchable language data, linguists have been able to discern patterns and make generalizations that were otherwise not noticed. This methodology has spawned a new field of study, Dialectometry, founded by Hans Goebel, which quantitatively analyses data from linguistic atlases to identify and study spatial regularity hidden in the mass of data (*Dialektometrie Projekt*: [www.dialectometry.com](http://www.dialectometry.com)).

Within the Romance realm, we find many innovative approaches to digital data storage and sharing, allowing for broader access to data. These include wikis (such as *Edisyn*, the European Dialect Syntax wiki: <http://www.dialectsyntax.org>), atlases (such as the *Atles interactiu de l'entonació del català*: <http://prosodia.upf.edu/atlesentonacio>), databases (such as the *Clitics of Romance Languages* (CRL): <http://crl.linguistics.stonybrook.edu>), etc. Some of these resources require registration while others do not, and some focus on a particular aspect of grammar, while others are broader in their approach. For example, the *Clitics of Romance Languages* database (Repetti/Ordóñez 2011) provides free access to a large corpus (both audio and text) of a particular grammatical structure (verb + post-verbal pronouns) among endangered Romance languages spoken in France, Italy, and Spain. The corpus consists of utterances containing a verb + pronoun phrase, which, in the languages investigated, have unusual stress patterns. The search function is designed to be easy to use, so that searches with various parameters can be conducted, and we see that the interaction between clitics (syntax) and stress (phonology) can be richer and more complex than had been assumed.

There is a downside to digital formats: they become obsolete when software is no longer supported or different formats are incompatible. This can lead to information being lost when the technology for accessing it is no longer available (Bird/Simons 2003), and the need for “a network of repositories and centers for safeguarding and using this documentation” (Krauss 1992, 8), which is largely lacking.

## 5 Conclusion

The tragic story of the last speaker of Dalmatian is well-known: his death in an explosion in 1898 deprived us of invaluable information on Eastern Romance, language change, language contact, language obsolescence, etc. (Maiden 2016b). Will this be the fate of endangered Romance languages today? I would rather not end on such an ominous note. In fact, there is hope. The languages themselves, with their well-studied history and structures and with easy access to native speakers, however small their numbers might be, could provide clues for ways to avoid language endangerment and prevent language loss. For decades we have heard the tolling of the death knell of the great linguistic variety present in many Romance-speaking contexts, but the variety has



persisted despite the odds. How could Faetar, a Francoprovençal dialect spoken deep in southern Italy, have survived for centuries in linguistic isolation (Nagy 2000)? How did Istro-Romanian, an Eastern Romance language spoken in Croatia, resist the onslaught of Western Romance and Slavic languages (Maiden 2016a)? Why has Veneto persisted among immigrant communities in Mexico, when other Italian emigrants abandoned their language (Sartor/Ursini 1983)? What these communities have experienced can be a clue to help endangered languages thrive in the twenty-first century and beyond. Perhaps it is time to look at these languages as hugely successful, having thrived for generations, centuries, even millennia, despite their numerically inferior status. (See Mufwene 2002 for a shift in focus in studies of endangered languages, and Moriarty 2011 for a shift in the roles of endangered languages.)

**Acknowledgements:** I would like to thank Wendy Ayres-Bennett, Janice Carruthers, José Elías-Ulloa, Lorenzo Filipponio, Olivia Mignone, Chiara Repetti-Ludlow, and Emily Romanello for helpful comments during the preparation of this chapter.

## 6 Bibliography

- Anderson, Stephen R. (2016), *Romansh (Rumantsch)*, in: Adam Ledgeway/Martin Maiden (edd.), *The Oxford Guide to the Romance Languages*, Oxford, Oxford University Press, 169–184.
- Ascoli, Graziadio I. (1873), *Saggi ladini*, *Archivio glottologico italiano* 1, 1–556.
- Assini, Alicia (2014), *New Resources for Endangered Languages*, *MultiLingual* October/November 2014, 28–30, <http://www.multilingual.com> (last access 13.02.2018).
- Austin, Peter K. (2010), *Current Issues in Language Documentation*, in: Peter K. Austin (ed.), *Language Documentation and Description*, vol. 7, London, SOAS, 12–33.
- Austin, Peter K./McGill, Stuart (edd.) (2012), *Endangered Languages*, 4 vol., London, Routledge.
- Austin, Peter K./Sallabank, Julia (2011), *Introduction*, in: Peter K. Austin/Julia Sallabank (edd.), *The Cambridge Handbook of Endangered Languages*, Cambridge, Cambridge University Press, 1–24.
- Barbiers, Sjef (2015), *European Dialect Syntax: Towards an Infrastructure of Documentation and Research of Endangered Dialects*, in: Mari C. Jones (ed.), *Endangered Languages and New Technologies*, Cambridge, Cambridge University Press, 35–48.
- Barnes, Hilary (2009), *A Sociolinguistic Study of Sustained Veneto-Spanish Bilingualism in Chipilo, Mexico*, doctoral thesis, Philadelphia, Pennsylvania State University.
- Bird, Steven/Simons, Gary (2003), *Seven Dimensions of Portability for Language Documentation and Description*, *Language* 79, 557–582.
- Bouquiaux, Luc/Thomas, Jacqueline M. C. (1992), *Studying and Describing Unwritten Languages*, Dallas, SIL.
- Carageani, Gheorghe (2002), *Suggerimenti per un approccio sociolinguistico dell'aromeno*, in: Gheorghe Carageani (ed.), *Studi linguistici sull'aromeno*, Cluj-Napoca, Clusium, 32–46.
- Cerruti, Massimo/Regis, Riccardo (2015), *The Interplay between Dialect and Standard: Evidence from Italo-Romance*, in: Eivind Torgersen et al. (edd.), *Language Variation. European Perspectives V*, Amsterdam/Philadelphia, Benjamins, 55–68.
- Chinellato, Paolo (2004), *Disturbi di sintassi nell'afasia non fluente: Un'analisi linguistica dell'agrammatismo italiano e dialettale*, doctoral thesis, Padova, Università di Padova.

- Cravens, Thomas D. (2014), *Italia linguistica and the “European Charter for Regional or Minority Languages”*, *Forum italicum* 48, 202–218.
- Crowley, Terry (2007), *Field Linguistics: A Beginner’s Guide*, Oxford, Oxford University Press.
- Dorian, Nancy (ed.) (1992), *Investigating Obsolescence*, Cambridge, Cambridge University Press.
- Ethnologue* (n. d.), <https://www.ethnologue.com/endangered-languages> (last access 13.02.2018).
- Gilliéron, Jules/Edmont, Edmond (1902–1910), *Atlas linguistique de la France*, Paris, Champion.
- Grinevald, Colette/Bert, Michel (2011), *Speakers and Communities*, in: Peter K. Austin/Julia Sallabank (edd.), *The Cambridge Handbook of Endangered Languages*, Cambridge, Cambridge University Press, 45–65.
- Guy, Gregory/Zilles, Ana M. S. (2008), *Endangered Language Varieties: Vernacular Speech and Linguistic Standardization in Brazilian Portuguese*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 53–66.
- Haller, Hermann H. (1993), *Una lingua perduta e ritrovata: L’italiano degli Italo-Americani*, Firenze, La Nuova Italia.
- Hinskens, Frans/Auer, Peter/Kerswill, Paul (2005), *The Study of Dialect Convergence and Divergence*, in: Peter Auer/Frans Hinskens/Paul Kerswill (edd.), *Dialect Change: Convergence and Divergence in European Languages*, Cambridge, Cambridge University Press, 1–48.
- Hugo, Russell (2015), *Endangered Languages, Technology and Learning: Immediate Applications and Long-Term Considerations*, in: Mari C. Jones (ed.), *Endangered Languages and New Technologies*, Cambridge, Cambridge University Press, 95–110.
- Huss, Leena (2008), *Scandinavian Minority Language Policies in Transition*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 129–143.
- Iliescu, Maria/Melchior, Luca (2015), *Friulano nel mondo*, in: Sabine Heinemann/Luca Melchior (edd.), *Manuale di linguistica friulana*, Berlin/Boston, De Gruyter, 338–363.
- Jaberg, Karl/Jud, Jakob (1928–1940), *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vol., Zofingen, Ringier.
- Jones, Mari C. (2014), *Jèrriais and Sercquiais Today*, BBC Voices, 23 Sept. <http://www.bbc.co.uk/voices/multilingual/jerriais.shtml> (last access 19.07.2015).
- Jones, Mari C. (ed.) (2015), *Endangered Languages and New Technologies*, Cambridge, Cambridge University Press.
- Jones, Mari C./Bulot, Thierry (2009), *Sociolinguistique de la langue normande (pluralité, normes, représentations)*, Paris, L’Harmattan.
- Krauss, Michael (1992), *The World’s Languages in Crisis*, *Language* 68, 4–10.
- Loporcaro, Michele (2013), *Profilo linguistico dei dialetti italiani*, Roma, Laterza.
- Mackay, Carolyn J. (1992), *Language Maintenance in Chipilo: A Veneto Dialect in Mexico*, *International Journal of the Sociology of Language* 96, 129–145.
- Maiden, Martin (2016a), *Romanian, Istro-Romanian, Megleno-Romanian, and Aromanian*, in: Adam Ledgeway/Martin Maiden (edd.), *The Oxford Guide to the Romance Languages*, Oxford, Oxford University Press, 91–125.
- Maiden, Martin (2016b), *Dalmatian*, in: Adam Ledgeway/Martin Maiden (edd.), *The Oxford Guide to the Romance Languages*, Oxford, Oxford University Press, 126–138.
- Moriarty, Máiréad (2011), *New Roles for Endangered Languages*, in: Peter K. Austin/Julia Sallabank (edd.), *The Cambridge Handbook of Endangered Languages*, Cambridge, Cambridge University Press, 446–458.
- Mosel, Ulrike (2006), *Fieldwork and Community Language Workers*, in: Jost Gippert/Nikolaus P. Himmelmann/Ulrike Mosel (edd.), *Essentials of Language Documentation*, Berlin/Boston, De Gruyter, 67–85.

- Mufwene, Salikoko S. (2002), *Colonisation, Globalisation, and the Future of Languages in the Twenty-First Century*, *International Journal on Multicultural Societies* 4, 162–193.
- Nagy, Naomi (2000), *Faetar*, München, Lincom Europa.
- NPR (2015), *Eau Claire Researchers Working To Save Dying Walloon Language*, <http://www.wpr.org/eau-claire-researchers-working-save-dying-walloon-language> (last access 20.09.2017).
- Ó hlfearnáin, Tadhg (2008), *Endangering Language Vitality through Institutional Development: Ideology, Authority, and Official Standard Irish in the Gaeltacht*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 113–128.
- Peyton, Joy Kreeft, et al. (2008), *Heritage Language Education in the United States: A Need to Reconceptualize and Restructure*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 173–186.
- Pons, Cathy R. (1990), *Language Death among Waldensians of Valdese, North Carolina*, doctoral thesis, Bloomington, Indiana University.
- Pop, Sever (1950), *La Dialectologie. Aperçu historique et méthodes d'enquêtes linguistiques*, Louvain, Université de Louvain.
- Repetti, Lori (2014), *Where Did All the Dialects Go? Aspects of the Influence of Italian on Dialects*, *Forum italicum* 48, 219–226.
- Repetti, Lori/Ordóñez, Francisco (2011), *Clitics of Romance Languages*, <http://crl.linguistics.stonybrook.edu> (last access 13.02.2018).
- Romaine, Suzanne (2008), *Linguistic Diversity, Sustainability, and the Future of the Past*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 7–21.
- Sallabank, Julia (2002), *Writing in an Unwritten Language: The Case of Guernsey French*, *Reading Working Papers in Linguistics* 6, 217–244.
- Sallabank, Julia (2010a), *Language Endangerment: Problems and Solutions*, eSharp (Special Issue: *Communicating Change: Representing Self and Community in a Technological World*), 50–87, <http://www.gla.ac.uk/esharp> (last access 08.07.2015).
- Sallabank, Julia (2010b), *The Role of Social Networks in Endangered Language Maintenance and Revitalization: The case of Guernesiais in the Channel Islands*, *Anthropological Linguistics* 52, <http://www.jstor.org/stable/41330796> (last access 09.06.2015).
- Sallabank, Julia (2013), *Attitudes to Endangered Languages: Identities and Policies*, Cambridge, Cambridge University Press.
- Sartor, Mario/Ursini, Flavia (1983), *Cent'anni di emigrazione: Una comunità veneta sugli altipiani del Messico*, Treviso, Grafiche Antiga.
- States of Jersey Statistics Unit (2012), *Jersey Annual Social Survey 2012*, <https://www.gov.je/SiteCollectionDocuments/Government%20and%20administration/R%20JASS2012%2020121204%20SU.pdf> (last access 19.07.2015).
- Tamburelli, Marco (n. d.), *I dialetti d'Italia: Cronaca di una morte annunciata?*, [http://www.dialettando.com/articoli/detail\\_new.lasso?id=10470](http://www.dialettando.com/articoli/detail_new.lasso?id=10470) (last access 10.07.2015).
- Tortora, Christina (2014), *Heritage Nation vs Heritage Language: Towards a More Nuanced Rhetoric of "Heritage" in Italian Language Pedagogy*, *Forum italicum* 48, 268–291.
- Tsunoda, Tasaku (2005), *Language Endangerment and Language Revitalization*, Berlin/Boston, De Gruyter.
- Uguzzoni, Arianna (1971), *Quantità fonetica e quantità fonemica nell'area dialettale frignanese*, *L'Italia dialettale* 34, 115–136.
- UNESCO (n.d.), <http://webarchive.unesco.org/20160809104708/http://www.unesco.org/new/en/culture/themes/endangered-languages> (last access 13.02.2018).

- Ursini, Flavia (1988), *Veneto e spagnolo a contatto: Il caso Chipilo*, in: *Elementi stranieri nei dialetti italiani. Atti del XIV convegno del CSDI*, Pisa, Pacini, 217–228.
- Warren, Anthony Scott/Jennings, Geraint (2015), *Allant contre vent et mathée: Jèrriais in the Twenty-First Century*, in: Mari C. Jones (ed.), *Endangered Languages and New Technologies*, Cambridge, Cambridge University Press, 127–140.
- Williamson, Robert C. (1991), *Three Romance Languages: Past and Present*, in: Robert C. Williamson (ed.), *Minority Languages and Bilingualism: Case Studies in Maintenance and Shift*, Norwood, Ablex, 50–67.
- Wolfram, Walt (2008), *Language Diversity and the Public Interest*, in: Kendall A. King et al. (edd.), *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, Washington, DC, Georgetown University Press, 187–202.