

Position Paper with Policy Recommendations on Safeguarding and Protecting the Romani Language

Identity, Education, Harmonisation and Digital Future

By the European Roma Institute for Arts and Culture (ERIAC)
ERIACNET4EU

With contributions by
Ian Hancock
Huseyin Kyuchuk (Hristo Kyuchukov)
Mihaela Zătreanu

Cover: Marina Rosselle, The Tree of the Romani Language, Barvalo, installation view, 2023

ERIAC
EUROPEAN ROMA INSTITUTE FOR ARTS AND CULTURE

ERIAC NET 4EU

Co-funded by the European Union

2026

ERIAC
EUROPEAN ROMA
INSTITUTE FOR ARTS
AND CULTURE

European Roma Institute for Arts and Culture e.V.
Reinhardtstraße 41 - 43. Ground floor, street entrance, DE -10117 Berlin
eriac@eriac.org

EUROPEAN ROMA INSTITUTE FOR ARTS AND CULTURE (ERIAC)

POSITION PAPER
WITH POLICY RECOMMENDATIONS
ON SAFEGUARDING AND PROTECTING THE ROMANI LANGUAGE

Identity, Education, Harmonisation and Digital Future

by the
EUROPEAN ROMA INSTITUTE FOR ARTS AND CULTURE (ERIAC)
ERIACNET4EU

2026

with contributions by
Ian Hancock · Huseyin Kyuchuk (Hristo Kyuchukov) · Mihaela Zătreanu

TABLE OF CONTENTS

1. FOREWORD....3

2. EXECUTIVE SUMMARY....3

3. INTRODUCTION AND RATIONALE....6

3.1 About ERIAC....6
3.2 ERIAC's Work on Romani Language and Culture....7
3.3 About ERIACNET4EU....8
3.4 Objectives of the Position Paper....8

4. METHODOLOGY....9

5. ROMANI LANGUAGE, IDENTITY AND COMMUNITY COHESION....10

5.1 The Historical Development of the Romani Language....10
5.2 Romani Language as a Pillar of Identity....12
5.3 Language, Belonging and Community Cohesion....14
5.4 Romani as European Intangible Cultural Heritage....15
5.5 Challenges to Language Transmission and Use....16

6. EDUCATION, HARMONISATION AND LANGUAGE DEVELOPMENT....18

6.1 The Current State of Romani Language Education in Europe....18
6.2 Teacher Training and Educational Resources....21
6.3 Common Romani and Harmonisation Efforts....24
6.4 Opportunities and Challenges for Language Development....27
6.5 The Role of Institutions in Supporting Romani Language Education....29

7. TECHNOLOGY, AI AND THE FUTURE OF ROMANI....30

7.1 Digital Presence of the Romani Language....30
7.2 Language Technologies and Digital Resources....30
7.3 Artificial Intelligence and Minority Languages....31
7.4 Risks of Digital Exclusion....32
7.5 Opportunities for Language Revitalisation through Technology....32
7.6 AMARI ČHIB: Building Romani Language Infrastructure for the AI Era....32

8. CASE STUDY: ROMANI IN GOOGLE TRANSLATE....33

8.1 Romani Language Revitalisation in Practice....33
8.2 Existing Initiatives and Good Practices....34
8.3 Challenges and Lessons Learned....34
8.4 Implications for European Policy and Practice....34

9. CONCLUSIONS....35

10. POLICY RECOMMENDATIONS....36

10.1 Recommendations for European Institutions....36
10.2 Recommendations for National Governments....36
10.3 Recommendations for Educational Authorities....37
10.4 Recommendations for Roma Civil Society and Cultural Institutions....38
10.5 Recommendations for the Digital and Technology Sector....38

11. CONTRIBUTORS....38

12. BIBLIOGRAPHY....39

1. FOREWORD

For more than a thousand years, the Romani language has travelled with our people – from north-western India, through Persia, Armenia and the Byzantine world, into every corner of Europe and beyond. It has survived slavery, persecution, forced assimilation, bans on its use and the genocide of the Roma during the Second World War. It has survived without a state, without an army, without schools and, for most of its history, without writing. That it is still spoken today by millions of people is a testimony to the resilience of Roma families and communities, who have passed it on from one generation to the next.

Yet survival is not the same as security. In many communities, children no longer acquire Romani at home; in many schools, the language a Roma child brings from the family is still treated as a problem rather than a resource; and in the digital world, where young people increasingly live, learn and communicate, Romani remains under-represented. At the same time, important opportunities are emerging: new CEFR-aligned learning resources are being developed, Romani has entered machine translation, and European institutions increasingly recognise the need to protect, teach and strengthen the language.

The European Roma Institute for Arts and Culture (ERIAC) considers the Romani language to be at the heart of Roma identity and an integral part of Europe's cultural heritage. With this Position Paper, ERIAC brings together the voices of leading Romani linguists and educators to set out where the Romani language stands today, what threatens it, and what must be done – by European institutions, governments, schools, Roma organisations and the technology sector – so that Romani is not only preserved, but used, taught and developed as a living language of the present and the future.

We thank the authors, reviewers and partners who made this paper possible, and we invite all stakeholders to engage with its recommendations.

2. EXECUTIVE SUMMARY

Context

Romani is the language of the Roma, Europe's largest transnational minority, whose population is commonly estimated at 10–12 million in Europe and 12–16 million worldwide[1]. It is a New Indo-Aryan language that left India around the eleventh century and took shape in Anatolia before spreading across Europe from the thirteenth century onwards[2]. There are no reliable census figures for Romani speakers: estimates range from some 3.5 million in Europe[3] to about half of the Roma population worldwide[4]. Romani is protected under the European Charter for Regional or Minority Languages (ECRML) in 16 of the 25 states that have ratified it[5], yet everywhere it remains a minority language without a state, used mostly orally and largely absent from education, administration and digital media.

Developed by the European Roma Institute for Arts and Culture (ERIAC) within the ERIACNET4EU project, this paper is informed by the expertise of ERIAC's Romani Language

[1] Council of Europe; Hancock 2002.
[2] Matras 2002; Elšík & Beníšek 2020.
[3] University of Manchester Romani Project.
[4] Hancock 2002.
[5] Council of Europe 2025.

thematic section and by commissioned contributions from two leading Romani scholars, Ian Hancock and Hristo Kyuchukov.

This Position Paper pursues four objectives: (1) to examine the role of Romani in Roma identity and cohesion; (2) to assess the state of Romani education, language development and harmonisation; (3) to explore the impact of digital technologies and artificial intelligence; and (4) to provide policy recommendations.

Key findings

- History and unity. All Romani groups share a common linguistic history up to their arrival in Europe. Dialect diversity is the result of seven centuries of dispersal, not of separate origins.[6] Knowledge of this shared history is a resource for unity among the many Roma groups (endaya).
- Identity. Romani is a central pillar of Roma identity and the main vehicle of Romani culture (Rromanipen), oral tradition and values. At the same time, many Roma are multilingual and many no longer speak Romani; their Roma identity persists. The language should therefore be promoted as a shared heritage open to all Roma, including those who wish to (re)learn it, and never used as a test of belonging.
- Transmission. Intergenerational transmission is weakened by antigypsyism, internalised stigma, migration, urbanisation and the near absence of Romani from public life. Parents who see the language as a liability may not pass it on.
- Education. Roma children who speak Romani enter school with a fully developed linguistic system. Assessing them only in the majority language confuses language difference with learning difficulty and contributes to placement in special schools and segregation. Provision of Romani in schools remains sporadic and dependent on local initiative, and is hampered by a lack of trained teachers and materials.
- Harmonisation and language development. Romani needs shared resources for education, publishing, cross-border communication and digital use. ERIAC therefore advocates harmonisation rather than the imposition of a single standard dialect: a Roma-led process of developing shared orthographic, grammatical and terminological reference points while respecting, documenting and teaching dialectal diversity. The objective is increased linguistic coherence and mutual intelligibility, not the replacement of living varieties.
- Digital future. Romani entered Google Translate in 2024[7], but its digital presence remains very limited. Machine translation and AI systems risk fixing one variety as “the” Romani, reproducing errors and stereotypes, and leaving most dialects behind. They also offer unprecedented opportunities for learning, documentation and revitalisation, provided that Roma communities take part in shaping them.

Key recommendations

- European institutions: recognise Romani as part of Europe’s linguistic and intangible cultural heritage; support a Roma-led harmonisation process and a Common Romani Language Framework; update the Curriculum Framework for Romani; earmark funding for Romani education, publishing and open language data.
- National governments: ratify and fully implement the ECRML for Romani; guarantee the right to learn Romani at school; end segregation; collect data on Romani use and teaching.

[6] Matras 2005a; Elšík & Beníšek 2020.

[7] Google 2024.

- Educational authorities: introduce bilingual language profiles and linguistically valid assessment; train and employ Romani teachers; develop curricula and textbooks aligned with the Curriculum Framework for Romani; include Roma history in curricula.
- Roma civil society and cultural institutions: promote language pride and family transmission; produce literature, media and cultural content in Romani; take ownership of harmonisation and terminology.
- Digital and technology sector: co-create Romani language technologies with Roma speakers and experts; cover the main dialects; apply community data governance; guard against antigypsyist outputs.

3. INTRODUCTION AND RATIONALE

The premise of this Position Paper is straightforward: language is one of the principal carriers of culture, historical memory and collective continuity. For Roma, a transnational people without a territorial state, Romani has an additional function. It connects communities across borders and preserves concepts, forms of expression and cultural knowledge that cannot be fully transferred into another language. Its weakening therefore concerns not only communication, but also the transmission of a distinct cultural heritage.

The linguistic reality addressed here is nevertheless complex. Romani has developed across centuries of mobility and contact with many languages; it is represented by numerous dialects and written practices; and its speakers live in highly multilingual environments. Any European policy for Romani must therefore avoid a false choice between unity and diversity. The challenge is to create shared resources that allow Romani to function in education, literature, public communication and new technologies while preserving the dialectal realities through which communities continue to live the language.

Because the Romani language has no state to protect it, its future depends on the commitment of Roma communities themselves and on the institutions that represent and serve them at European level. As the only transnational, Roma-led institution dedicated to the arts, culture, history and language of the Roma, ERIAC has both the mandate and the responsibility to articulate a common position on the Romani language and to bring together the expertise needed to safeguard it.

3.1 About ERIAC

The European Roma Institute for Arts and Culture e.V. (ERIAC) is a joint initiative of the Council of Europe, the Roma Foundation for Europe (formerly the Open Society Foundations – Roma Initiatives Office) and the Alliance for the European Roma Institute for Arts and Culture. It was established as an association under German law in April 2017 and launched at a ceremony in Berlin on 8 June 2017. ERIAC has its headquarters in Berlin and regional offices in Serbia and, since 2024, in Romania.

ERIAC exists to increase the self-esteem of Roma and to decrease negative prejudice of the majority population towards the Roma by means of arts, culture, history, and media. It is guided by the principles of respect for Romani dignity and identity, recognition of the diversity of Romani identities and cultures, Romani leadership with the support of non-Roma, political autonomy and high quality standards.

ERIAC is a membership-based organisation and the only transnational cultural institute for Roma arts and culture. It brings together Roma and non-Roma individuals and organisations with demonstrated experience and expertise in arts and culture: in 2026, 280 associate members were approved, 75.3% of whom identify as Roma and 51.6% are women. The associate members are organised in thematic sections – Arts and Culture; History and Commemoration; Research and Publication; Media and Information – and in the Romani Language section, which is cross-sectional and works across all areas. ERIAC is governed through its statutory bodies and supported by the Barvalipe Academy, which brings together distinguished Roma intellectuals, artists and scholars and provides strategic and artistic guidance.

ERIAC cooperates closely with the Council of Europe on the basis of a Memorandum of Understanding signed in 2018. The Council of Europe Strategy for Roma and Traveller

Inclusion (2026–2030) confirms this partnership as the institutional framework for joint work until 2030.[8]

3.2 ERIAC's Work on Romani Language and Culture

The Romani language is a cross-sectional field of ERIAC's work. Since 2020, the Romani Language Initiative (RLI / Romani Čhib) has brought together linguists, teachers, writers, cultural actors and institutions around a long-term objective: to strengthen Romani as a living European language of education, culture, public communication and digital life. The initiative is led by Mihaela Zătreanu and developed in cooperation with ERIAC's international network and the Council of Europe.

RLI works from the principle that the future of Romani requires both protection of dialect diversity and the development of shared resources. ERIAC uses the term harmonisation for this process. Harmonisation does not mean selecting one dialect and declaring it superior, nor does it seek to replace the varieties spoken in families and communities. It means identifying common linguistic ground and developing compatible orthographic, grammatical and terminological reference points so that speakers of different varieties can learn, write, publish and communicate across borders with greater ease. In Romani, the aspiration can be expressed as KHETANI ROMANI ČHIB: a common linguistic space built through convergence, not erasure.

A central practical achievement is the Romani Čhib adult textbook pathway aligned with the Council of Europe's Curriculum Framework for Romani and the Common European Framework of Reference for Languages. The A1 and A2 volumes established a shared pedagogical approach; B1 expanded the linguistic and communicative range; and B2 is being developed for advanced adult learners. The books are produced through collaboration among authors and consultants from different countries and dialect backgrounds. Their value therefore lies not only in the number of learners reached, but in the infrastructure they create: a coherent A1–B2 pathway, reusable terminology, grammatical explanations, a shared writing practice and a body of validated educational language that can support future teacher training and digital tools.

RLI also develops teacher capacity and a transnational academic network involving, among others, the University of Bucharest, Sorbonne, Charles University, the University of Pécs, the University of Novi Sad, Södertörn University, the University of Graz, the University of Manchester and the University of Helsinki. Its annual Safeguarding Our Romani Language conference creates a recurring European space for debate on language rights, teaching, literature, harmonisation and technology. The initiative has also supported the development of a Romani Language Council concept and, in 2025, the formalisation of a Romani PEN framework connecting language development with literary production and cultural prestige.

This work increasingly extends into the digital sphere. ERIAC's AMARI ČHIB initiative, developed in cooperation with Microsoft's innovation and AI-for-good ecosystem, is designed to transform validated educational and cultural resources into structured Romani text and speech data and, progressively, into language technologies. The strategic objective is not merely to make Romani visible online, but to ensure that the language enters emerging technologies with linguistic coherence, documented diversity and Roma participation in decisions about data, terminology and representation.

[8] ERIAC, "About ERIAC", "Membership" and "ERIAC Welcomes the 2026 Thematic Section Coordinators" (eriac.org, accessed 1 October 2026); ERIAC, Minutes of the 8th General Assembly, 28–29 November 2025; Council of Europe, Fifth Evaluation Report on ERIAC (2025); Council of Europe, Strategy for Roma and Traveller Inclusion (2026–2030).

For ERIAC, harmonisation is therefore an enabling framework. It supports mutual understanding and transnational cooperation while allowing regional and local forms to remain visible. The common layer and the dialectal layer are not competing projects: each strengthens the other. A shared framework makes it possible to create textbooks, dictionaries, terminology, teacher training, corpora and AI resources at European scale; systematic documentation of dialects ensures that this common framework remains rooted in the richness of actual Romani-speaking communities.

3.3 About ERIACNET4EU

This paper is published in the frame of the project ERIACNET4EU: ERIAC Network to Advance Roma Inclusion by Combatting and Preventing Antigypsyism in the EU. Initiated in 2022, ERIACNET4EU is a multi-year initiative developed through a framework partnership agreement under the Citizens, Equality, Rights and Values Programme (CERV) of the European Union.

The project aims to ensure quality input to policymaking and policy implementation at local, regional, national and EU levels, through mutual learning and the exchange of good practices among ERIAC members and with other public and private stakeholders; to raise awareness; to increase the capacities and knowledge base of members and stakeholders through training; and to grow the network and community outreach. Within the project, ERIAC releases thematic, evidence-based position papers that gather the expertise of its members to support European institutions and Member States in implementing the EU Roma strategic framework: on education and Roma knowledge production (2023), on Roma remembrance, commemoration and history teaching (2024), and on Romani media representation and production (2025). The present paper, dedicated to the Romani language, continues this series.[9]

3.4 Objectives of the Position Paper

Roma communities across Europe share a linguistic heritage that crosses national borders. The Romani language is one of the most important expressions of their history, culture and sense of belonging, while its many varieties also reflect the different historical experiences of Roma communities in the countries in which they live.

This Position Paper aims to contribute to the recognition, protection, development, and revitalisation of the Romani language as an integral part of Roma identity, cultural continuity, and collective heritage, as well as an essential component of Europe's linguistic and cultural diversity.

Specifically, the Position Paper aims to:

1. Examine the role of the Romani language in strengthening Roma identity, cultural continuity, community cohesion, and transnational connections among Roma communities across Europe.
2. Assess the current opportunities and challenges related to Romani language education, language development, and harmonization initiatives aimed at strengthening communication, knowledge production, and cooperation among Roma communities across Europe.

[9] ERIAC, “ERIACNET4EU” (eriac.org).

3. Explore the impact of digital technologies, artificial intelligence, and emerging innovations on the future development, accessibility, visibility, and sustainability of the Romani language.
4. Provide policy recommendations for European institutions, national governments, educational authorities, Roma organisations, and other relevant stakeholders to strengthen the use, transmission, recognition, and long-term vitality of the Romani language.

4. METHODOLOGY

This Position Paper combines commissioned expert contributions with ERIAC's institutional expertise, policy and document analysis, and desk research. It does not present new empirical data collected specifically for this publication. Instead, it brings established linguistic and educational research into dialogue with ERIAC's practical work on language education, harmonisation and digital development.

Sources and methods

- Historical and comparative linguistic analysis (Ian Hancock): the origins and development of Romani, shared pre-European history, dialect differentiation, identity and the challenges of developing common written resources. Hancock's contribution is retained selectively in this paper and is explicitly attributed where his argument or wording is used.
- Sociolinguistic, psycholinguistic and educational research by Huseyin Kyuchuk (Hristo Kyuchukov): multilingualism, language and identity, the acquisition of Romani by children, bilingual assessment, antigypsyism and educational inequality, and regional approaches to language development. The paper draws on the studies listed in the bibliography where their findings are discussed.
- ERIAC institutional analysis: the experience of the Romani Language Initiative, the A1–B2 educational pathway, teacher development, the emerging European network for Romani language cooperation, harmonisation work, and the development of digital and AI-oriented language resources.
- Policy and document analysis: relevant instruments and programmes of the Council of Europe, European Union and UNESCO, together with European initiatives in Romani language education and digital inclusion.
- Desk research on digital resources and language technologies for Romani, including the 2024 introduction of Romani in Google Translate and ERIAC's AMARI ČHIB concept.
- Roma is used as the umbrella term for the people, in line with European institutional practice. It is acknowledged that some groups (e.g. Sinti, Manouches, Kale, Romanichals) use their own endonyms. Hancock uses "Romanies" and the Romani term endaya (sg. endani) for the different groups.
- Romani (also Romanes, Romani čhib) refers to the language. Dialect groups follow the usual classification: Balkan, Vlax, Central, Northern (including Sinti and North-eastern)[10].
- Standardisation means the codification of one norm. Harmonisation means agreement on shared reference points while respecting dialect diversity (see 3.2). Common Romani refers to a shared written variety or framework.

[10] Matras 2002, 2005a; Boretzky & Igla 2004; Elšík & Beníšek 2020.

Limitations

Most of the empirical evidence comes from Central, Eastern and South-Eastern Europe. Reliable statistics on Romani speakers and on Romani teaching are lacking in most countries. No new data were collected specifically for this paper. The two author contributions reflect different disciplinary perspectives; where they diverge, this paper states ERIAC's position explicitly.

5. ROMANI LANGUAGE, IDENTITY AND COMMUNITY COHESION

5.1 The Historical Development of the Romani Language

Hancock approaches the history of Romani through the language itself: the Indic core, successive layers of contact vocabulary, grammatical development and the later differentiation of dialects. His central policy-relevant point is that today's diversity emerged from a shared linguistic history and should therefore be understood as evidence of historical development, not as evidence that Roma communities have unrelated linguistic origins.

Romani is spoken by populations on every continent except Antarctica, but has never been spoken in the land of its origin. This last fact makes it unique amongst the world's truly diasporic languages.

The story of why this apparently anomalous situation exists is both fascinating and controversial, and rests upon the fact that the details of the provenance and history of its speakers are only now beginning to fall into place. For a while, we remembered where we had come from. In his Chronicon Foroliviense, Hieronymus de Forlivio (the priest Girolamo Fiocchi) described meeting a group of Romanies in Italy in August, 1422, when some of them told him that their ancestors were from India: "aliqui decebant quod erant de India"[11], but over time that fundamental knowledge became lost by us, leaving the way open for often bizarre speculation by outsiders.

HISTORY

> *Kodole kaj mangen te keren tumendar rroburen čivar phenena tumenge o čačipe pala tumari istoriya ("Those who want to make you slaves will never tell you the truth about your history").*

Our ancestors first began to show up in south-eastern Europe during the late Byzantine period, somewhere between 1200 and 1300 AD. They were arriving from Anatolia at the capital, Byzantium (Constantinople, now Istanbul) which straddled both continents, though not as one; indications are that there was a steady, ongoing migration of Romani-speaking groups (as well as of members of other non-Romani populations) leaving Asia Minor and crossing over into Europe.

Based on linguistic data, there seem to have been three major Romani exoduses, the earliest travelling the furthest and reaching Europe's northern and western fringes (Germany, Scandinavia, England, Spain). In Europe, the varieties of Romani spoken fall broadly into one of three or five major dialect groups; by 1500 AD, a Romani presence had been reported from nearly every region.

[11] Muratori 1754.

OUR LANGUAGE

Amari čhib si amari zor ("Our language is our strength").

Ours is an Indian language. This fact was realized by western scholars in the 18th century after its similarities with an Indian language (probably Hindi or Urdu) were fortuitously recognized. The revelation created great institutional interest; spurred by European colonization overseas, the Enlightenment saw an increase in the study of non-western peoples, cultures, religions and languages and the emergence of new academic disciplines. Presumably originally from India, what were Romanies doing in the West? Why did they leave? When? Did they have a country? Why did they occupy so marginal a place everywhere in Europe?

The point of the team's investigative entry was linguistic; not a new idea—in 1870 Alexander Paspati wrote in his monumental work on Romanies in Ottoman Europe that the clues to Romani history were revealed in the language.

A useful teaching metaphor for this is an onion. An onion grows bigger layer upon layer, with the deepest layers being closest to its heart. If we examine the vocabulary of the Romani language, the very first layer, at the centre of the onion, is Indian. The next layers include, one after the other, words from such languages as Phalura, Persian, Kurdish, Armenian, Greek and so on all acquired along the way, reflecting the geographical route taken by the Romanies' ancestors as they came westwards out of India. There is a handful of words from Burushaski, a tiny language spoken in a small area in the Himalayas; there are two or three words from Georgian, spoken in the Caucasus; there is one from Tatar; clues like this help plot a route on the map.

Romani shares its earlier ("pre-Romani") history with other Indo-European languages, which in some of its branches lost the neuter gender category of nouns. As with those branches, we divide them diachronically into Old, Middle and New developmental periods, and just as the Romance languages lost their earlier neuter category, redistributing them into either the masculine or the feminine sets, so many of the Indian languages did so too, during the transition from Middle Indic to New Indic, around the year 1000 AD.

Romani has two grammatical genders, like Hindi and Panjabi and other New Indic languages, having only masculine and feminine nouns and pronouns. Furthermore, the redistribution of the earlier neuter nouns in Romani matches their redistribution in those languages still spoken in India. If it had been taken out of India before the 11th century, it would presumably be a three-gender language today. Assuming, then, that the exodus occurred no earlier than ca. 1000 AD, and knowing that we were already arriving in the West by the 13th century, we are left with a window in time to account for—the years between leaving India and arriving in Europe.

Hancock has advanced a more specific hypothesis linking part of the early history of Romani to the Ghaznavid period and to multilingual military populations in north-western India. He argues that the mixture of Indic forms and the substantial Persian layer may be consistent with the emergence of a levelled lingua franca or koïné before the movement westward[12]. This interpretation remains a scholarly hypothesis and is presented here as Hancock's reconstruction rather than as a settled historical consensus.

Having arrived in Anatolia, the mixed ethnolinguistic group of Indians became settled, and through intermarriage the various native Indian languages became lost as children had been learning the koïné as their mother tongue. Surrounded by Greek, its grammar and vocabulary

[12] Hancock 2000; Marsh 2004, 2008; John 2013.

became extensively hybridized; after its Indian lexical content, the next largest contributor to the Romani lexicon is Greek, amounting to several hundred words. And as the language crystallized, so the Indians, first arriving as a professionally-defined mixed group, gradually became an ethnically-defined one.

The formation of the Romani people and language did not happen quickly, nor did the Indians cross into Europe all at once. The earliest groups to leave Anatolia have fewer Greek words in their dialects, and have Indian words that are missing in others. Having crossed into Europe, the different numbers migrated in different directions, encountering different European languages. Today, Romanies in Spain have been separated from Romanies in Slovakia, for example, for hundreds of years and by hundreds of miles. Their ancestors left Asia at a time when Romani identity was still happening. No wonder, then, that we are not able to present a cohesive account of our overall history.

Nevertheless, all Romani populations share the same history right up until the time of the movement across into Europe; group histories only began to differ once in the West. Paradoxically, as a people whose language, culture and genetic history is ultimately Asian, we have only existed in the West, and this fact underlies the difficulties European governments are confronted with in accommodating their Romani minority populations.

To summarise: the beginnings of the Romani language were in Asia, carried out of north-western India a thousand years ago (O Teljaripen “the exodus”), coalescing in the Middle East and Anatolia over the following two centuries (O Nakhipen “the crossing over”), before being brought across into the West (O Aresipen “the arrival”). Here, it spread out over all of Europe, becoming subject to extensive influence and modification from the dozens of different languages in each territory in which speakers of early Romani stopped. Separated by time and space, and subjected to changes from the coexisting European languages, Romani fragmentation began—and continues to this day.

5.2 Romani Language as a Pillar of Identity

IDENTITY

Hancock’s contribution links language, historical consciousness and the possibility of a broader Romani identity. He argues that identification often remains strongest at the level of the endani or particular Romani group, and that knowledge of a common pre-European history can widen the limits of mutual recognition across endaya.

Hancock also warns against turning language into a test of authenticity. He notes that speakers may judge other dialects as “wrong” or less legitimate, reproducing divisions inside the Romani linguistic space[13]. For a language policy concerned with cohesion, this is a crucial point: a common framework can only succeed if no dialect community is treated as linguistically inferior.

CULTURE

> *Amari čhib si o vurdon kaj ingarel angle kon sam (“Our language is the waggon that transports our identity”).*

The Core of Direct Retention

We have included here the “core of direct retention” chart, which represents what has been lost and what has been gained. It consists of our language, culture and genetics acquired in the

[13] Acton 2021; Toporov 2009.

pre-European period (Indian, Dardic, Iranic, Armenian, Georgian, Greek), until the time of the coalescing of the pre-Romani peoples into the Romanies in Anatolia.

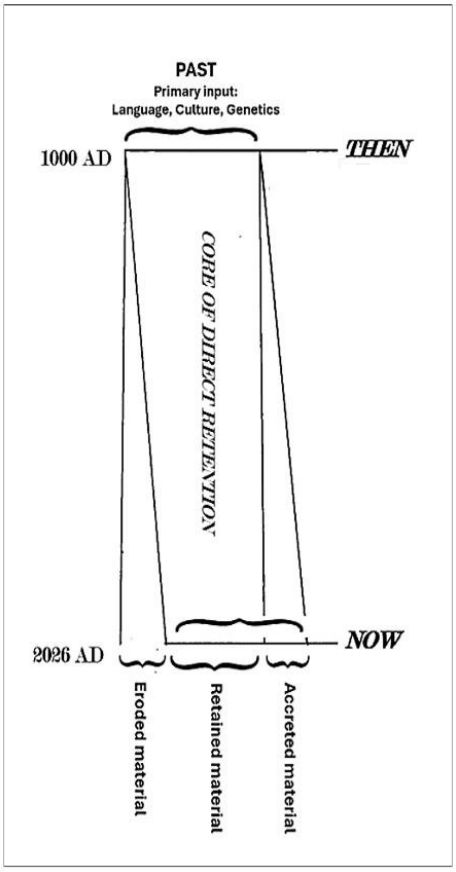


Figure 1. The core of direct retention (Hancock).

Hancock's concept of Rromanipen is important here because it names the cultural world carried through language: values, kinship, social categories, oral knowledge and ways of interpreting belonging. His "core of direct retention" model, reproduced in Figure 1, visualises the relationship between inherited elements and the material accumulated through centuries of contact. The model is not a hierarchy of "pure" and "impure" Romani; rather, it illustrates why all contemporary varieties contain both shared inheritance and historically acquired layers.

Drawing on Huseyin Kyuchuk's (Hristo Kyuchukov's) sociolinguistic research, the picture becomes more explicitly multilingual. Contemporary Roma communities differ by dialect, religion, country, migration history and patterns of language use. Kyuchukov's studies of Roma in Bulgaria, North Macedonia, Cyprus and Germany show that Romani may be central to identity without being the only language through which Roma identity is expressed[14].

Kyuchuk's (Kyuchukov's) central argument is that Romani should be understood as an important resource for Roma identity without assuming that knowledge or use of Romani is a necessary condition of Roma identification. Roma identity and Romani language are closely related historically and culturally, but they are not identical categories.

Identity is not a fixed category that exists independently of social interaction. From a sociolinguistic perspective, identity may be understood as something constructed and negotiated through language practices and participation in different social networks. Speakers

[14] Kyuchukov 2006; Kyuchukov & Kwadrans 2024; Kyuchukov & Samuilov 2019.

may emphasize different aspects of their identities depending on the communicative situation and the social relationships involved.

This perspective is particularly relevant to Roma communities because the category "Roma" encompasses populations that differ with regard to dialect, religion, country of residence, socioeconomic circumstances and patterns of language use. Consequently, the assumption that all Roma speakers share a single linguistic profile risks obscuring significant internal diversity.

This evidence challenges a monolingual conception of ethnic identity. For multilingual speakers, identity may be expressed through a repertoire of languages rather than through a single language. Romani can constitute one component of this repertoire, while Turkish, Bulgarian, German or another language may perform different communicative and social functions.

The distinction between ethnic identity and linguistic identity is therefore fundamental. An individual may identify as Roma without using Romani as their principal language, just as an individual may use Romani without necessarily expressing a single, uniform conception of Roma identity. The findings indicate that language is an important component of identity, but not an exclusive determinant of it.

The two perspectives presented above are complementary rather than contradictory. Hancock reminds us that the Romani language is the most direct link to the shared history of all Romani groups. It is the vehicle of Rromanipen, of kinship terms, values and concepts that exist in no other language: its loss means the gradual loss of a unique culture. Kyuchuk (Kyuchukov) shows that Roma today live in multilingual realities, that language use is shaped by prestige, migration and social context, and that Roma identity persists among those who no longer speak Romani.

ERIAC's position combines both insights. The Romani language is a central pillar and a symbolic core of Roma identity and cultural continuity, and its protection is therefore a matter of collective survival. Roma identity, however, is not conditional on language competence. Roma who do not speak Romani – often because of historical persecution, assimilation policies and stigma – are fully part of the Roma people. They should be supported, not judged, if they wish to (re)learn the language of their ancestors. Romani should be promoted as a shared heritage and a living resource for all Roma, within multilingual repertoires and alongside the national languages Roma speak, never as an instrument of exclusion.

5.3 Language, Belonging and Community Cohesion

COMMUNITY COHESION

Jekhipen si zor ("Unity is strength")

Kyuchuk's (Kyuchukov's) research emphasises that multilingualism is not an exceptional condition in Roma life but a normal feature of many communities. Romani may coexist with Turkish, Bulgarian, German, Macedonian or other languages, each used in different social domains. This matters because language vitality cannot be measured by an ideal of monolingual "purity".

Code-switching and borrowing should therefore not automatically be read as evidence of decline. They can perform communicative and social functions and may reflect long-standing contact. At the same time, Kyuchukov shows that prestige matters: when Romani is associated

with stigma while another language is associated with education, mobility or status, language choices can gradually reshape transmission[15].

Thus, the relevant question is not simply whether Roma speakers use Romani "purely," but how Romani functions within their broader linguistic repertoire.

Migration makes this relationship especially visible. In Kyuchukov and Samuilov's study of Bulgarian Muslim Roma youth in Berlin, changing social networks were associated with changes in language practices and identity. The policy implication is not that migration dissolves Roma identity, but that linguistic belonging is negotiated within new multilingual environments[16].

Research in North Macedonia and Cyprus likewise shows that some respondents identify as Roma while reporting Turkish as their mother tongue. Such findings support a distinction between ethnic continuity and linguistic continuity: Roma identity can persist when Romani use weakens, even though language shift may reduce the transmission of cultural knowledge and intergenerational communication[17].

Roma identity may involve a combination of ancestry, family relations, cultural practices, historical consciousness, social networks, community membership and language. Romani can be an important symbolic and communicative resource within this configuration, but its role varies across communities and individuals.

Consequently, the loss or reduced use of Romani should not automatically be equated with the disappearance of Roma identity. Nevertheless, language shift can have important consequences for the transmission of cultural knowledge, oral traditions and intergenerational forms of communication. This creates a distinction between **ethnic continuity** and **linguistic continuity**: the former may persist under conditions in which the latter becomes weaker.

At the transnational level, Romani nevertheless remains a uniquely powerful connector. Since the First World Romani Congress in London in 1971, language has been linked to the political and cultural articulation of a shared Roma public sphere. Later congresses established language structures and debated common writing. Today, migration, literature, media, cultural events and digital communication create new situations in which speakers of different varieties meet, accommodate and learn from one another.

ERIAC's annual Safeguarding Our Romani Language conference, the Romani PEN framework, the World Day of the Romani Language and the wider RLI network are contemporary examples of this transnational function. They demonstrate why harmonisation matters: not because all Roma should speak identically, but because a shared linguistic layer makes cooperation, education, cultural production and knowledge exchange possible across borders.

5.4 Romani as European Intangible Cultural Heritage

Language is both an element of intangible cultural heritage and the vehicle through which most other elements of that heritage are transmitted. The UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage (2003) explicitly lists "oral traditions and expressions, including language as a vehicle of the intangible cultural heritage" as a domain of

[15] Kyuchukov 2006.

[16] Kyuchukov & Samuilov 2019.

[17] Kyuchukov & Kwadrans 2024.

heritage[18]. For the Roma, whose history has been transmitted orally for most of its existence, this is particularly true. Romani carries oral literature (paramiča – tales, songs, proverbs), kinship and social concepts (e.g. xanamik, the relationship between co-parents-in-law), the values and purity concepts of Rromanipen, the vocabulary of traditional crafts and music, and the collective memory of persecution, including the Holocaust (Samudaripen/Porrajimos)[19].

Romani is also a European heritage in the full sense. It has been spoken in Europe for over seven centuries; it contains layers from Greek, Slavic, Romanian, Hungarian, German and many other European languages; and it is the clearest example of a non-territorial language recognised under the European Charter for Regional or Minority Languages[20]. The Council of Europe Framework Convention on the Value of Cultural Heritage for Society (Faro, 2005)[21] affirms the right of communities to benefit from and contribute to cultural heritage. The Framework Convention for the Protection of National Minorities and Article 22 of the EU Charter of Fundamental Rights (respect for linguistic diversity)[22] provide further anchors.

Recognising Romani as European intangible heritage has practical consequences: documenting dialects and oral traditions before they disappear; including Romani heritage in national inventories and in European heritage programmes; supporting archives (e.g. RomArchive, the Romani Archives and Documentation Center) and museums; and presenting Romani not as a “minority problem” but as part of the cultural wealth of all Europeans.[23]

5.5 Challenges to Language Transmission and Use

The concept of language maintenance provides a useful framework for interpreting the relationship between Romani and identity. Language maintenance occurs when a community continues to transmit and use a language across generations and social domains. Language shift occurs when another language progressively assumes functions previously performed by the community language.

For Roma communities, these processes cannot be examined independently of multilingualism. A speaker may retain Romani within the family while using Bulgarian at school, Turkish in particular community networks and German in professional or public settings. Such a repertoire does not fit easily into a binary model of language maintenance versus language loss.

A more useful approach is to examine domains of use: whether Romani is transmitted to children; used with family and peers; written and read; present in education, literature, media and public life; perceived as useful and prestigious; and able to generate terminology for contemporary realities. These indicators distinguish a language that is merely remembered from a language that continues to function across generations and domains.

Our work on standardization, multilingualism and language use suggests that linguistic vitality is influenced by precisely these factors. The existence of speakers alone is therefore insufficient for understanding the sociolinguistic position of Romani.

[18] Art. 2.2; UNESCO 2003.

[19] Hancock 2002; Matras 2002.

[20] Council of Europe 1992, 2025.

[21] Council of Europe 2005.

[22] Council of Europe 1995; European Union 2012.

[23] UNESCO 2003 Convention, Art. 2; ECRML Factsheet on Romani (CoE); Faro Convention (CETS 199); FCNM (CETS 157); EU Charter of Fundamental Rights, Art. 22.

What we know – and do not know

There are no reliable figures on the number of Romani speakers. Censuses record ethnicity and mother tongue inconsistently, and many Roma do not declare either. The University of Manchester Romani Project estimates "upwards of 3.5 million" speakers in Europe and "upwards of 500,000" elsewhere, noting that the real number may be much higher[24]. Hancock[25] estimates that about half of a Romani-identifying population of 12–16 million speaks the language. The vitality of Romani varies greatly. It is still the first language of most children in many communities in South-Eastern and Central Europe (e.g. Vlax and Balkan varieties in Romania, Serbia, North Macedonia, Kosovo, Bulgaria and Slovakia)[26]. In Western and Northern Europe, by contrast, several varieties have disappeared or survive only as Para-Romani registers (e.g. Angloromani in Britain, Caló in Spain and Portugal); Welsh Romani, documented by Sampson[27], is no longer spoken[28], and others such as Finnish Kalo and Sinti are classified as endangered[29].

Historical drivers

- Centuries of prohibition and forced assimilation: for example the 18th-century Habsburg decrees of Maria Theresa and Joseph II banning the use of Romani[30], the abolition of Roma slavery in the Romanian principalities only in 1855–1856[31], and assimilation policies in many states.
- The genocide of the Roma during the Second World War, which destroyed entire speech communities, particularly in Central Europe[32].
- Communist-era assimilation policies (forced sedentarisation, denial of minority status, the segregated schools called the "Gypsy schools" in Bulgaria and special schools in Czechoslovakia)[33].

Present-day drivers

- Antigypsyism and internalised stigma: parents may withhold the language from their children, seeing it as a liability[34].
- The absence of Romani from schools, media and public life, so that children learn that it has no place in formal knowledge[35].
- Urbanisation, mixed marriages and migration, including intra-EU migration, which change the language ecology of families[36].
- Shift towards the dominant language among young people, increasingly reinforced by digital media in majority languages.

[24] Romani Project, University of Manchester.
[25] Hancock 2002.
[26] Matras 2002, 2013.
[27] Sampson 1926.
[28] Matras 2002.
[29] Moseley 2010.
[30] Council of Europe, Factsheets on Romani History.
[31] Achim 2004.
[32] Hancock 2002; Elšík & Beníšek 2020.
[33] Cf. European Court of Human Rights 2007; Kyuchukov 2015.
[34] Hancock; see 6.3.
[35] Kyuchukov; see 6.1.
[36] Kyuchukov; see 5.3.

Language shift is not inevitable. Where Romani is valued in the family, used with peers, present at school and visible in media, transmission remains strong. The domains listed by Kyuchukov above can serve as indicators for monitoring.[37]

6. EDUCATION, HARMONISATION AND LANGUAGE DEVELOPMENT

6.1 The Current State of Romani Language Education in Europe

Across Europe, the introduction of Romani in mainstream education remains, as noted on the Council of Europe INSCHOOL platform[38], "sporadic, experimental, and dependent on local initiative", while university-level study is more visible. The ECRML Committee of Experts repeatedly identifies the same structural obstacles: a shortage of qualified teachers, inadequate teaching materials and absent or sporadic teacher training[39]. Romani is protected under the Charter in 16 states: under Part III in Bosnia and Herzegovina, Germany, Hungary, Montenegro, Poland, Serbia and Slovakia, and under Part II only in Austria, Czechia, Finland, the Netherlands, Norway, Romania, Slovenia, Sweden and Ukraine.

The following table gives an indicative overview of selected countries, based on Council of Europe monitoring and national sources[40].

Country	ECRML status of Romani	Main form of provision in schools	Notes
Romania	Part II	Romani as mother-tongue subject and "History and Traditions of the Roma", since the early 1990s; some Romani-medium kindergarten/primary classes	Roma inspectors, teachers and school mediators; Romani section at the University of Bucharest (G. Sarău); large body of teaching materials from kindergarten to university
North Macedonia	not a party	Optional subject "Romani Language and Culture", grades 3–9 (since 2008); earlier optional mother-tongue classes	Long codification tradition[41]; limited uptake
Serbia	Part III	Elective "Romani language with elements of national culture" in primary schools	Teacher certification at the University of Belgrade; obstacles: lack of teachers and low parental uptake

[37] University of Manchester Romani Project, "The status of Romani"; Hancock (this paper); Kyuchukov (this paper).
[38] Zătreanu 2022.
[39] Council of Europe 2025.
[40] Council of Europe 2025; Sarău 2016.
[41] Jusuf & Kepeski 1980.

Country	ECRML status of Romani	Main form of provision in schools	Notes
Slovakia	Part III	Romani language and literature in a limited number of schools; experimental programmes	Department of Romani Studies, Constantine the Philosopher University, Nitra; RPPE pilot schools (2022–2025)
Czechia	Part II	Very limited; Romani mostly in extracurricular activities	Romani Studies at Charles University, Prague (since 1991)
Hungary	Part III	Nationality education in Romani and Beás	Romology at the University of Pécs; Beás taught alongside Romani
Austria	Part II	Burgenland-Romani in some schools in Burgenland	Codification of Burgenland-Romani with the University of Graz (Romani Projekt)
Sweden	Part II	Mother-tongue tuition in Romani chib as a national minority language	Romani chib also used in some public services
Finland	Part II	Mother-tongue instruction in Finnish Kalo (limited hours)	Romani language board; Romani studies at the University of Helsinki
Slovenia, Greece	Slovenia: Part II; Greece: not a party	Pilot projects	Romani–Plurilingual Policy Experimentation (CoE, 2022–2025): 4 schools in Slovenia, 3 in Greece

Three lessons emerge. First, legal recognition does not automatically produce teaching: even in Part III countries, provision is often minimal. Second, where provision exists, it usually takes the form of an optional, extra-curricular subject with one or two hours per week, which is not enough for literacy or academic language to develop. Third, the most sustainable systems – Romania being the clearest example[42] – combine political will, a dedicated inspectorate, university-trained Romani teachers, school mediators and a continuous production of materials.[43]

Education, language and equality

Drawing on Huseyin Kyuchuk's (Hristo Kyuchukov's) research, this section treats education as one of the principal institutional domains in which language, identity and equality meet. A child whose home language differs from the language of schooling does not enter school "without language". The child enters with an already developing linguistic system and, in many cases, must acquire the majority language as an additional language.

[42] Sarău 2016.

[43] ECRML Factsheet "Romani and the ECRML" (CoE); Council of Europe, "The importance of Romani language learning and teaching for a more inclusive education for all"; CoE Romani language in education pages; Presidency of North Macedonia (open lesson, Roma Language and Culture); EU Delegation to Serbia, "Romany as optional course".

Kyuchukov's work on teacher-pupil interaction, literacy, bilingualism and assessment repeatedly shows why this distinction matters[44]. When schools evaluate a bilingual Roma child only through the majority language, limited proficiency in that language can be mistaken for limited linguistic, cognitive or educational ability. A formally identical assessment can therefore produce substantively unequal outcomes.

This is especially serious in the context of special education. Kyuchukov and New describe cases in which Roma children were assessed predominantly in the official language and placed in special schools even though bilingualism had not been adequately taken into account[45]. The mechanism is straightforward: a Romani-speaking child has limited proficiency in the school language; testing is conducted through that language; language limitations depress performance; and the result is interpreted as a broader deficit. Where assessment cannot distinguish language difference from learning difficulty, minority-language status can become a pathway to exclusion.

Kyuchuk's (Kyuchukov's) analysis of antigypsyism in education extends the argument beyond individual prejudice. Antigypsyism can be interpersonal, but it can also be institutional: embedded in assessment procedures, segregation, low expectations, the invisibility of Romani in curricula, or the assumption that only the majority language counts as legitimate knowledge[46]. The absence of explicit hostility does not by itself guarantee equal educational opportunity.

Romani competence at school

A major contribution of Kyuchuk's (Kyuchukov's) psycholinguistic work is the empirical demonstration that Romani-speaking children possess measurable and developing grammatical competence in their first language. Studies using Romani-specific assessment instruments show that children acquire complex structures during the preschool and early-school years and that socioeconomic disadvantage should not be equated with weak first-language development[47].

The distinction between oral competence, literacy and academic language is essential. A child may possess sophisticated oral vocabulary, grammar, narrative and pragmatic competence in Romani while having little experience of reading, writing or formal terminology in Romani. This is not evidence of deficient Romani. It reflects the limited institutional domains in which the language has historically been taught and written.

For educational policy, the appropriate model is therefore additive: Romani + the majority language + academic literacy. Recognition of Romani does not imply replacing the language of schooling. It means using the child's complete linguistic repertoire as a foundation for learning and ensuring that the majority language is acquired without delegitimising the home language.

From deficit to linguistic profile

The practical shift proposed by Kyuchuk (Kyuchukov) is from a deficit question - "What does the Roma child lack?" - to a developmental question: "Which languages does the child know, what can the child do in each language, and how can this knowledge support further learning?" This change affects expectations, assessment and pedagogy.

[44] Kyuchukov 1995, 1999, 2014; Kyuchukov, de Villiers & Tabori 2017.

[45] Kyuchukov & New 2018, 2019.

[46] Kyuchukov 2023; Kyuchukov & New 2016.

[47] Kyuchukov & de Villiers 2014; Kyuchukov, Samko & Kopcanova 2017; Kyuchukov et al. 2024.

At school entry, educational authorities should therefore document a child's language history: languages used at home, age of acquisition, frequency and domains of use, and previous educational exposure. Where Romani is a significant home language, assessment should, where feasible, include Romani comprehension, vocabulary, grammar, narrative and pragmatic competence. Reading and writing should be evaluated separately from oral language, and broader cognitive conclusions should be drawn only after language factors have been considered.

This approach also requires sensitivity to dialect variation. A form that differs from the variety used by an assessment instrument is not automatically an error. Assessment must distinguish developmental error, dialectal variation, bilingual transfer and genuine lack of knowledge. Otherwise, a standardised school variety can reproduce the same hierarchy that language policy is trying to overcome.

ERIAC considers these findings directly relevant to the development of Romani education. A harmonised written framework can make curricula, textbooks and assessment tools more transferable across countries, but it must remain capable of representing legitimate dialect forms. Educational coherence and dialect recognition must be designed together.

Domain	Possible linguistic experience
Home communication	Romani
Family interaction	Romani and/or majority language
Community communication	Romani and/or majority language
Preschool	Predominantly majority language
School instruction	Majority language
Literacy	Frequently majority language
Formal Romani literacy	Often limited or absent
Academic vocabulary	Developed primarily in the school language

6.2 Teacher Training and Educational Resources

Teachers' knowledge of Romani

Teachers working with Romani-speaking children do not necessarily need to become fluent Romani speakers. However, they need sufficient linguistic awareness to understand the consequences of bilingual development.

Teacher education should address:

- the difference between first- and second-language acquisition;
- normal bilingual development;
- the grammatical complexity of Romani;
- variation among Romani varieties;
- the difference between oral competence and literacy;
- language-transfer processes;
- appropriate bilingual assessment;
- and the danger of interpreting limited majority-language proficiency as general inability.

Kyuchukov's research provides an empirical foundation for this approach because it demonstrates that Romani-speaking children possess measurable and developing grammatical knowledge in their first language.

Romani language variation

A further methodological issue is linguistic variation. Romani is not a single uniform variety, and Kyuchuk's (Kyuchukov's) work on regional standardisation and child language across several countries reinforces the need to distinguish variation from error[48].

Consequently, language assessment instruments must be sensitive to dialect and regional variation.

A child's use of a form different from that expected in a standardized assessment should not automatically be interpreted as an error.

This is especially important for educational testing because standardized school language can privilege one variety while treating other legitimate varieties as deficient.

A scientifically sound assessment should therefore distinguish:

- developmental error;
- dialectal variation;
- bilingual transfer;
- and genuine lack of knowledge.

A proposed model for school assessment

On the basis of Kyuchuk's (Kyuchukov's) research, a multidimensional assessment model for Romani-speaking pupils can be proposed.

Stage 1: Language history

The school should establish:

- languages spoken at home;
- age of acquisition of each language;
- frequency of use;
- language of interaction with parents;
- language of interaction with peers;
- previous educational exposure.

Stage 2: Romani assessment

Where Romani is the child's home language, assessment should examine appropriate dimensions of:

- comprehension;
- vocabulary;
- grammar;
- narrative;
- sentence processing;
- pragmatic competence.

Stage 3: Majority-language assessment

[48] Kyuchukov 2018.

The child should then be assessed in the language of schooling.

Stage 4: Literacy assessment

Reading and writing should be evaluated separately from oral language.

Stage 5: Cognitive and educational assessment

Only after language factors have been appropriately considered should conclusions about broader educational or cognitive abilities be drawn.

This procedure reduces the risk that second-language limitations will contaminate the assessment of other abilities.

European reference tools

- The Curriculum Framework for Romani (CFR), developed by the Council of Europe Language Policy Unit with the European Roma and Travellers Forum (2008)[49]. It defines learning outcomes at CEFR levels A1–B2 for three age groups (3–6, 7–10, 11–14) across eleven themes and the transversal theme of Romanipe (e.g. “Myself and my family”, “My community”, “At school”). It is accompanied by two European Language Portfolios (ages 6–11 and 11–16) and a Teacher’s Handbook, available in ten languages and varieties, including Lovari and North Central Romani.
- QualiRom (ECML, 2010–2013, co-funded by the EU, Austria and Finland): learning materials based on the CFR were developed, piloted and published online in six varieties (Arlije, East Slovak, Finnish Romani, Gurbet, Lovara and Kalderaš). Since 2015 they have been followed by QualiRom Training and Consultancy[50].
- Romani–Plurilingual Policy Experimentation (Council of Europe, 2022–2025): plurilingual and intercultural approaches using Romani in ten primary schools in Greece, Slovakia and Slovenia[51].
- ERIAC’s Romani Čhib textbook series for adults: A1 and A2, followed by B1 and the B2 level in development. The series is aligned with the CFR and CEFR and is conceived as a continuous A1–B2 learning pathway. Authors and consultants from different dialect backgrounds work within a shared writing and pedagogical framework, making the series both a teaching resource and a practical laboratory of harmonisation.

Teacher education

Romani is taught or researched at university level in a limited but important number of institutions. Through RLI, ERIAC has worked to connect a broader university network including Bucharest, Sorbonne, Charles University, Pécs, Novi Sad, Södertörn, Graz, Manchester and Helsinki. The strategic objective is not only academic study of Romani, but a sustainable ecosystem for teacher education, research, terminology, curriculum development and future digital resources.

Gaps

- Materials for upper-primary and secondary levels, for subject teaching in Romani, and for adult (re)learners (Hancock’s Category B).

[49] Council of Europe 2008.
[50] ECML 2013.
[51] Council of Europe 2022.

- Terminology for school subjects (mathematics, science, grammar), which requires coordinated corpus work (see 6.4).
- Dialect-appropriate materials. Children should be able to recognise their home variety while learning a shared written norm.
- Assessment instruments in Romani, as proposed by Kyuchukov, and training of psychologists and teachers in bilingual assessment.
- Initial teacher education modules on Romani linguistic diversity, bilingual development and antigypsyism for all teachers working with Roma pupils, not only Romani teachers.
- An updated Curriculum Framework for Romani, reflecting digital learning and the harmonisation approach (an update is under preparation with ERIAC).[52]

6.3 Common Romani and Harmonisation Efforts

STANDARDISATION

> *E dobunda le khetanimaske amara čhibako si o angluno paso karing o khetanipe sar jekh narodo, "the achievement of linguistic unity is the first step toward unity as a people."*

Ian Hancock's contribution argues forcefully that a commonly usable written form is necessary if Romani is to expand in education, publishing and transnational communication. His discussion uses the language of "standardisation", reflecting a long tradition in Romani language planning. ERIAC retains the substance of this argument - the need for common resources - while using the term harmonisation for its own institutional approach.

Category A:

Hancock distinguishes between Roma who already speak one or more Romani varieties and Roma who identify as Roma but did not acquire the language. For existing speakers, the challenge is not simply linguistic. It is also connected to prestige, literacy and the experience of being told that Romani is unsuitable for school or public life.

> I wrote in Slovak, and it never occurred to me that you could write in Romani as well. The peasants made fun of us for speaking cant, gibberish, Pharaoh talk; Romani had no place in school, I had never seen a printed word in Romani in my life, and I thought that Slovak was the only language appropriate for literature[53].

As Hancock stresses, language planning cannot succeed without language pride. When stigma is internalised, parents may avoid transmitting Romani and speakers may reject written forms as artificial or unnecessary. Education therefore has a double role: teaching literacy and common resources, while affirming that Romani is capable of expressing contemporary knowledge.

For speakers who already possess a home variety, Hancock's preferred outcome is additive: a common written or transnational form should be learned as an additional resource, not as a replacement for the dialect acquired in the family.

Category B

For Roma who do not speak Romani, the question is different. Hancock identifies three broad possibilities: learning an ancestral local variety where it can still be reconstructed; learning a widely used living variety; or learning a shared form developed for wider

[52] Council of Europe, "Romani language in education"; CoE/ECML, "The QualiRom project"; ERIAC, Romani Čhib B1; University of Manchester REF impact case study.

[53] Lacková 1999, p. 64.

communication. In practice, these choices need not be mutually exclusive. A European learning framework can provide a common entry point while connecting learners to the dialectal history of their families and communities.

Hancock also records recurring objections: that majority languages already serve cross-border communication; that a common variety could privilege one group; or that codification might reduce the spontaneity of a living language. These concerns are legitimate warnings against imposed uniformity. They do not, however, remove the practical need for compatible writing, terminology and teaching resources.

The fact that Romani is already written in literature, education, journalism and digital communication means that the relevant question is no longer whether it should be written, but how written practices can become more mutually compatible without suppressing dialectal identity.

The expense of alternatively producing the same materials—including regularised spelling systems— in each of the many dialects would be considerable, and creating literacy programmes for them would be just as much a problem. And instead of having to expand the lexicon for just one new standard dialect, the other dialects would also have to address this need, not necessarily creating the same new words for, say, “hard drive” or “digital” or “phylum,” thus the problems of miscommunication would simply be perpetuated.

Historically, language planners have debated whether a common form should elaborate one existing variety or combine resources from several. Hancock reviews both approaches and notes that any choice has representational consequences. This is precisely why ERIAC avoids presenting harmonisation as the elevation of a single dialect.

The Romani linguist Jan Kochanowski believed that the standardised variety, which he called Khetani Romani[54], should be based on his own Baltic dialect, which belongs to the Northern group, arguing that it retains more of the original grammar and lexicon than do dialects in the Central, Vlax and Balkan groups which, he claimed, have been more heavily influenced in all areas by the surrounding European languages. He went much further, rashly believing that Romani has “between fifteen and twenty million speakers,” and advocated its adoption as an international auxiliary language by the whole world à la Esperanto, because it is not associated with any existing national government[55]. The non-Romani linguist Marcel Courthiade’s standardised dialect is more eclectic, but it too is based especially on Vlax models, as is Horvat-Muc & Djurić’s proposal[56].

The solution would seem to be a combination of both approaches; while on the basis of its overall number of speakers, and its spread over all of Europe and the Americas, the Vlax dialects would seem to be a good choice, and are in fact the most widely learnt and documented; but at the same time they are heavily influenced phonologically and lexically by other languages: nearly half (40%) of the words in one Kalderash dictionary[57] are Romanian, not original Romani. A Common Romani should avoid dialectal differences in pronunciation, use the forms most widely used and which are most faithful to early Romani, thus /tiro/ ‘your’ rather than [ʧiro], [tjiro], [kiro], &c., /pativ/ ‘honour’ rather than [pakiv] or [paʧiv] and so on.

Kyuchuk (Kyuchukov) approaches the same problem from a sociolinguistic perspective. His work on regional standardisation in the Balkans stresses that codification is not only

[54] Kochanowski 1995.

[55] Kochanowski 1971: 76.

[56] Horvat Muc & Đurić 2012.

[57] Gjerdman & Ljungberg 1963.

linguistic but sociopolitical: it determines which forms become visible in education, publishing and institutions[58].

The question of identity is also connected to the development of Romani as a standardized written language. Our work on regional standardization in the Balkans addresses the practical linguistic problems involved in developing common written forms while taking dialectal diversity into account. The use of Vlax dialects is considered as a basis, the incorporation of lexical resources from other Romani varieties where appropriate, and the formation of new terminology according to Romani grammatical structures.

Standardization is not exclusively a linguistic process. It is also a sociopolitical process because it determines which forms become visible in education, publishing and institutional communication.

The standardization of a language with substantial dialectal diversity inevitably raises questions of representation. A standard variety can facilitate literacy, educational development and wider communication, but it can also create tensions if speakers perceive particular dialects as privileged over others.

For Romani, this problem is especially significant because dialectal variation is an integral characteristic of the language. A successful standardization process therefore needs to recognize variation rather than treating dialectal diversity as evidence of linguistic inadequacy.

Our approach seeks mechanisms for developing common terminology while maintaining the grammatical resources of Romani.

The two expert contributions converge on the practical need for common written resources and on the necessity of taking dialect diversity seriously. ERIAC builds on this common ground but deliberately frames its policy as harmonisation: a process of increased linguistic coherence through shared resources, developed with speakers and institutions, while documenting and respecting different linguistic realities.

Experience so far

- The 1971 and later World Romani Congresses placed standardisation on the agenda. In 1990 (Serock/Warsaw), the International Romani Union adopted the common alphabet developed by Marcel Courthiade, used in Romania, in parts of the Balkans and in many publications[59].
- National and regional norms have emerged: Romania (G. Sarău, based on Courthiade's system), North Macedonia[60], Serbia, Kosovo, Burgenland in Austria (University of Graz), Slovakia and Czechia (Hübschmannová's North Central Romani), Bulgaria (Vlax Romani and Erlija) and Finland.
- Linguists such as Yaron Matras and Milena Hübschmannová have argued for a pluralistic, polycentric codification that accepts several regional written norms, an approach also supported by the European Roma and Travellers Forum[61].
- ERIAC's Romani Čhib A1–B2 pathway uses a consistent shared writing practice and is produced by authors and consultants from different dialect backgrounds. It demonstrates

[58] Kyuchukov 2018.

[59] Matras 1999; Hancock 2002.

[60] Jusuf & Kepeski 1980.

[61] Matras 1999, 2005c, 2015, 2021.

that common educational resources can be built collaboratively without presenting one community's speech as the only legitimate Romani.

ERIAC's position

ERIAC supports harmonisation rather than the imposition of a single standard dialect. In practice, harmonisation means developing shared reference points that can be used across countries while keeping dialect diversity visible and legitimate:

- One common, practical orthography that can be typed on any device (cf. Hancock's recommendations) and that allows each dialect to be written with the same conventions.
- A Common Romani Language Framework for teaching, built around shared grammatical and lexical resources and connected to CEFR/CFR progression, in which major dialect groups are represented and learners are taught to recognise variation.
- Coordinated terminology development, drawing first on Romani's own resources (derivation, compounding, retrieval of words preserved in other dialects), then on shared international vocabulary.
- A Roma-led and transparent governance process, involving linguists, teachers, writers, translators, universities, cultural institutions and community representatives from the main dialect areas. ERIAC's proposed Romani Language Council is intended as a structure for consultation and coordination, not as an authority that declares one dialect "correct".
- Additive learning: the common layer is learned alongside the home dialect, never in place of it. For adult re-learners, it should also provide pathways towards the varieties connected to their families and communities.

6.4 Opportunities and Challenges for Language Development

Orthography

Hancock treats orthography as the first practical infrastructure of a shared written language. A writing system must be learnable, stable and usable across countries and technologies. Earlier efforts, including the alphabet presented by Marcel Courthiade at the Fourth World Romani Congress in 1990, demonstrated the political and cultural importance of common writing, while later practice also revealed the need for technical simplicity and broad accessibility.

The history of Romani writing shows both convergence and plurality. Common systems have been used in books, grammars and dictionaries, while national and regional traditions have also adapted spelling to surrounding languages. For ERIAC, the policy lesson is that harmonisation should seek maximum compatibility across dialects and devices, with as few technical barriers as possible.

Since 1990, there has been a proliferation of Romani-language periodicals from all over Europe. A very useful overview of these has been provided by Dragoljub Acković[62], who reproduces sample pages of several of them. Here, even a cursory examination reveals that there is still a tendency to use orthographies which draw upon the conventions of the coexisting non-Romani language.

[62] Acković 1997.

Recommendations

Hancock therefore argues for one internationally usable set of conventions rather than reproducing the orthographic habits of every majority language. ERIAC shares the objective of compatibility, while recognising that any common orthography must be tested with teachers, writers and speakers from different dialect areas before wider adoption.

Lexicon

Lexical development is equally important. Romani needs terminology for education, administration, science, technology and contemporary cultural life. Hancock proposes drawing first on the resources of Romani itself - including words and derivational patterns preserved in different dialects - and supplementing them where necessary with transparent international vocabulary or carefully developed new forms. ERIAC considers this a field for coordinated terminology work rather than isolated invention.

Grammar

Grammar presents the same challenge. Hancock argues that language development should make maximum use of grammatical resources already present across Romani varieties. The task is therefore not to invent a new grammar, but to document what exists, identify shared structures and make choices that remain intelligible across dialect areas.

A precondition for responsible harmonisation is better documentation. No common framework can represent Romani fairly if major dialects remain insufficiently described. Corpus building, dialect documentation and educational harmonisation must therefore proceed together.

It is unrealistic to regard Romani as being solely an Indian language; while considering its Iranic, Armenian and other ingredients, it must be treated broadly, as an Indo-Hellenic language, its Greek component as fundamental as the Indian, both structurally and lexically.

Summary

Hancock's technical recommendations can be summarised for policy purposes in four principles: prefer historically and cross-dialectally grounded forms where they are viable; recover useful Romani resources preserved in particular varieties; expand terminology transparently when new concepts require it; and keep the writing system as practical and digitally usable as possible. ERIAC treats these principles as inputs to a consultative harmonisation process rather than as rules to be imposed mechanically.

Despite the obstacles described by the authors, Romani is today being written, published and created more than at any time in its history. Since 1990 there has been a proliferation of Romani-language periodicals, literature, poetry and translations, including complete Bible translations into several varieties, classics of world literature, and original works by Romani writers. The new Romani PEN Centre gives Romani authors and translators an international platform.

Further opportunities include:

- University-trained Romani linguists and teachers, many of them Roma, who can carry corpus planning forward.
- Terminology work for school subjects and public administration, building on existing glossaries in Romania, Serbia, North Macedonia and elsewhere.
- European frameworks (CFR, ECRML) that give the language institutional legitimacy.

- The interest of young Roma, including in the diaspora, in (re)learning the language, often through online media.
- Digital tools (see ch. 7), which allow dictionaries, corpora and learning materials to be shared across borders at low cost.

The challenge is to coordinate these energies so that new words, spellings and materials converge rather than multiply. This is precisely the purpose of harmonisation.

6.5 The Role of Institutions in Supporting Romani Language Education

Conclusion

For political reasons, most of the work on the standardisation of Romani should be undertaken by ourselves, ideally native speakers, rather than by non-Romanies; of the twenty-two different people who have published on Romani standardisation over the past forty years, only five have been Romanies (Hancock, Kochanowski, Djuric, Kyuchukov, Jusuf), and this has kept the management of the language in mostly non-Romani hands. The creation of a mixed, but Romanies-dominant, team of specialists would redress a long-standing imbalance of representation, but until educational and social conditions improve drastically, achieving the standardisation and the practical implementation of the language will continue to be the concern of just a handful of individuals, and remain largely in the non-Romani academic domain.

European level

- Council of Europe: the European Charter for Regional or Minority Languages and its Committee of Experts (monitoring of Romani in 16 states); the Framework Convention for the Protection of National Minorities; the Language Policy Unit and the European Centre for Modern Languages (CFR, ELP, QualiRom); the Roma and Travellers Division (Strategy for Roma and Traveller Inclusion 2026–2030, which commits to the harmonisation, protection and teaching of Romani with ERIAC)[63]; Recommendation CM/Rec(2020)2 on the inclusion of the history of Roma and/or Travellers in school curricula and teaching materials.
- European Union: the EU Roma strategic framework for equality, inclusion and participation 2020–2030 and the Council Recommendation of 12 March 2021 on Roma equality, inclusion and participation; funding through Erasmus+, Creative Europe, the European Social Fund+, and digital and research programmes; the Fundamental Rights Agency (data on Roma).
- OSCE/ODIHR (Contact Point for Roma and Sinti Issues) and the OSCE High Commissioner on National Minorities (language rights).
- UNESCO: the 2003 Convention on intangible cultural heritage; multilingual education.

National and local level

Ministries of education and culture, curriculum and examination agencies, school inspectorates (e.g. Roma inspectors in Romania), teacher-training institutions, universities, public broadcasters and local authorities.

Roma and civil society

ERIAC, the International Romani Union and its language commission, Roma NGOs, cultural centres, publishers and media, Roma teachers' and writers' associations, the Romani PEN Centre, religious communities, the World Day of the Romani Language (5 November),

[63] Council of Europe 2026.

initiated in Zagreb in 2008 by the Croatian Romani association Kali Sara and endorsed by the International Romani Union in 2009[64].

A European Romani Language Council

To overcome fragmentation, ERIAC proposes a permanent, Roma-led Romani Language Council or network, coordinated by ERIAC in partnership with relevant European institutions and universities. It should include linguists, teachers, writers, translators and community representatives from the main dialect areas. Its function would be to coordinate terminology, review shared educational resources, document variation, advise on orthographic compatibility and ensure that harmonisation remains participatory. It would not police everyday speech or replace local language authorities; its legitimacy would depend on transparency, representation and the quality of the resources it produces.

7. TECHNOLOGY, AI AND THE FUTURE OF ROMANI

7.1 Digital Presence of the Romani Language

For young Roma, as for all young Europeans, the digital space is where much of daily communication, learning and cultural consumption takes place. Romani is visibly present there, mostly through informal, user-generated content. Romani-language music and video channels, social-media groups, messaging communities, podcasts and online radio connect speakers across countries and continents, and often use Romani spontaneously, in mixed codes and in a variety of spellings. More formal digital content is much scarcer. A Wikipedia exists in Vlax Romani but is small and far less developed than Wikipedias in comparably sized languages; online news in Romani is limited to a few portals; and public-service media offer only occasional Romani programmes.

Two features characterise this presence. First, orthographic diversity: online writers mostly use the spelling conventions of their national language, or English-based spelling on smartphones, rather than the 1990 common alphabet. This confirms Hancock's argument that a practical, diacritic-light orthography is needed. Second, dialect fragmentation: content is produced in many varieties with little cross-linking, which makes it hard to find and hard to use as data for language technologies.

7.2 Language Technologies and Digital Resources

Romani is a low-resource language in terms of language technology. The main existing resources include:

- ROMLEX, the online lexical database developed by the Universities of Graz and Manchester (launched 2005): over 25 Romani varieties, with translations into 14–15 languages[65].
- The Romani Linguistics resources of the University of Manchester (dialect samples, Romani Morpho-Syntax database) and the RomaniNet and QualiRom online materials.
- National online dictionaries and digitised printed dictionaries (e.g. for Kalderash, Lovari, Arli and Burgenland-Romani), and digitised Romani periodicals and literature in archives such as RomArchive.

[64] Kali Sara & IRU 2009.

[65] Romani Project, University of Manchester.

- Machine translation: in June 2024 Google added Romani to Google Translate as one of 110 new languages, using its PaLM 2 large language model. According to Google, the output is closest to Southern Vlax Romani, with elements of Northern Vlax and Balkan Romani.[66]
- Basic infrastructure: Unicode supports all characters of the common alphabet, but keyboards and spell-checkers for Romani are rarely available on standard devices.

What is missing are large, open, well-annotated text and speech corpora covering the main dialects, morphological analysers, speech recognition and synthesis, and a shared repository. Without them, Romani will remain dependent on technology developed by large companies using whatever data they can find, rather than on resources developed together with the community.[67]

7.3 Artificial Intelligence and Minority Languages

Large language models (LLMs) and AI-based translation are transforming how people access information and produce text. For minority languages, they bring both a promise and a risk. The promise is that, for the first time, a language like Romani can be translated, written and learned with tools that were once available only for major languages. The risk is that these systems are trained on very small and unbalanced data. Early observers of Google Translate's Romani noted that it favours Vlax varieties, mixes elements from different dialects inconsistently, struggles with grammatical case and pronoun distinctions, appears to pivot through English, uses inconsistent spelling and diacritics, and does not cover Central, Northern or Para-Romani varieties[68].

Three issues are specific to Romani:

- De facto standardisation: whichever variety an AI system produces may become, in the eyes of users, "the" Romani, bypassing the harmonisation debate described in 6.3. As Jonathan Lee noted, the diversity of Romani dialects risks being "reduced to one single computer-generated language"[69].
- Bias and antigypsyism: AI systems trained on majority-language data can reproduce stereotypes about Roma. The Council of Europe Strategy 2026–2030 commits to evaluating the impact of generative AI on the rights of Roma and Travellers, acknowledging that "AI systems can reinforce harmful narratives"[70].
- Data sovereignty: Romani texts and recordings – including oral heritage – may be collected and used without the consent or benefit of the community. Principles of community data governance, transparency about training data, and involvement of Romani linguists are therefore essential.

The regulatory framework is emerging. The EU Artificial Intelligence Act (Regulation (EU) 2024/1689) sets requirements on transparency, data governance and non-discrimination. European initiatives on multilingual digital infrastructure aim at "digital language equality" for all European languages[71], but Romani has so far received little attention in them.[72]

[66] Google 2024.

[67] University of Manchester Romani Project (ROMLEX, impact case study); Google blog "110 new languages are coming to Google Translate" (June 2024); Romea.cz (30 June 2024).

[68] Elšík & Albert 2024; Lee 2024.

[69] Lee 2024.

[70] Council of Europe 2026.

[71] Rehm & Way 2023.

[72] J. Lee, "Google Translate and the future of the Romani language", The Norwich Radical (30 Sept. 2024); Roma Foundation/rroma.org (July 2024); Council of Europe Strategy 2026–2030; Regulation (EU) 2024/1689.

7.4 Risks of Digital Exclusion

- The digital divide. Many Roma households, especially in segregated settlements, have limited access to broadband, devices and digital skills, although smartphone use is widespread. Digital tools that require computers or paid subscriptions will not reach those who need them most.
- Linguistic exclusion. If only one or two varieties are supported by technology, speakers of other dialects – and the dialects themselves – are pushed further to the margins.
- Quality and trust. Poor machine translations, used without human review in schools, administration or media, may spread errors, damage the prestige of the language and reinforce the old prejudice that Romani is "not a real language".
- Antigypsyism online. Hate speech and stereotypes circulate widely on social media; content moderation rarely covers Romani; and AI systems can amplify stereotypes.
- Absence of Roma in technology development. Very few Roma work in the language-technology sector, so decisions about Romani are taken without Romani speakers.
- Dependence on commercial platforms, whose priorities may change, rather than on open, community-owned resources.

7.5 Opportunities for Language Revitalisation through Technology

Used well, technology can be the strongest ally of Romani revitalisation:

- Learning: apps, online courses, mobile-first lessons aligned with the CFR and ERIAC's Romani Čhib textbooks, and the inclusion of Romani on mainstream language-learning platforms (an idea discussed at ERIAC's 2025 General Assembly). These are especially valuable for adults and young people who did not acquire Romani at home.
- Documentation: digital archives of oral heritage – stories, songs, testimonies of elders, Holocaust survivor accounts – recorded and transcribed with AI assistance before they are lost.
- Media and culture: subtitling of films and series in Romani (including AI-assisted subtitles), podcasts, Romani-language content creators, digital publishing of Romani literature.
- Teaching resources: shared online repositories of materials in several dialects; AI-assisted generation of exercises and texts, reviewed by Romani teachers.
- Harmonisation tools: online dictionaries and terminology databases (e.g. the planned ERIAC English–Romani–English dictionary), spell-checkers and keyboards implementing the common orthography – technology can help a common written norm spread faster than print ever could.
- Community-owned data: open, ethically collected corpora built with Roma communities, which can then be used to improve machine translation and AI on the community's own terms.

7.6 AMARI ČHIB: Building Romani Language Infrastructure for the AI Era

ERIAC's AMARI ČHIB initiative translates the policy principles of this chapter into a practical programme. Developed through the Romani Language Initiative in cooperation with Microsoft's innovation and AI-for-good ecosystem, the concept starts from a simple problem: Romani is a low-resource language in digital terms. Without sufficiently large, validated and diverse text and speech resources, AI systems will continue to produce uneven results and to rely on whatever varieties happen to be most visible online.

The initiative is designed in stages. The first priority is to organise and validate Romani educational and cultural content, including the growing A1–B2 textbook pathway, into reusable language data. This is complemented by the collection and documentation of speech from different speakers and varieties. A second layer concerns digitisation and an accessible archive of Romani-language materials. On this foundation, future tools can include text-to-text assistance, speech-to-text, text-to-speech, learning applications, translation support and a Romani-language AI agent.

The significance of AMARI ČHIB goes beyond one technological product. It creates a bridge between education, harmonisation and digital sovereignty. The same validated terminology and grammatical choices that make textbooks coherent can improve digital corpora; the same dialect documentation that protects linguistic diversity can prevent AI systems from treating one variety as the whole language. Human review by Romani speakers and linguists remains indispensable throughout.

For ERIAC, this is the model that should guide future cooperation with the technology sector: open and reusable resources where appropriate, clear community governance, documentation of dialectal provenance, transparent validation and long-term investment in Roma expertise. AI should not become an external force that standardises Romani by accident. It should become a tool through which Roma communities strengthen the language on their own terms.

8. CASE STUDY: ROMANI IN GOOGLE TRANSLATE

This chapter uses the inclusion of Romani in Google Translate in 2024 as a concrete example of the opportunities and risks discussed above. It is not presented as a complete account of Romani revitalisation, but as a test case for a wider policy question: what happens when a language with substantial dialect diversity enters a global technology before its communities have fully coordinated common digital resources?

8.1 Romani Language Revitalisation in Practice

On 27 June 2024, Google announced the largest expansion in the history of Google Translate: 110 new languages, made possible by its PaLM 2 large language model. Romani was among them. Google stated that the translations produced are closest to Southern Vlax Romani and include elements of Northern Vlax and Balkan Romani – explicitly acknowledging that Romani has “many dialects” and that, where no single standard exists, the system had to prioritise the most commonly used varieties[73].

The announcement was welcomed by many Roma and followed earlier Roma-led campaigns for the inclusion of Romani in Google Translate, such as the 2021 campaign of the Aresel civic platform in Romania[74]. From the outset, Romani linguists also pointed to its limitations[75]. For the first time, anyone with a smartphone could translate between Romani and more than 200 other languages, free of charge. Symbolically, Romani was placed on the same footing as major world languages in one of the most widely used digital tools.

From a revitalisation perspective, the tool potentially supports learners (looking up words and phrases), teachers (preparing materials), Roma in the diaspora (communicating with

[73] Google 2024.
[74] Media-Marketing 2021.
[75] Elšík & Albert 2024.

relatives), and non-Roma professionals (social workers, health staff) communicating with Romani-speaking clients.

8.2 Existing Initiatives and Good Practices

The Google Translate case should be read alongside other initiatives that show what works in Romani revitalisation:

- Romania: the most comprehensive Romani education system in Europe, built since 1990 with Roma inspectors, teachers, mediators, university training and a continuous production of materials, which has helped thousands of pupils study Romani and its history[76].
- Fair assessment: Kyuchukov and colleagues' psycholinguistic tests in Romani, used with 250 children aged 3–10 in several countries, show that children from resource-poor communities have age-appropriate competence in Romani. This is a model for linguistically valid assessment[77].
- European tools: the Curriculum Framework for Romani, the QualiRom materials in six varieties and the Romani–Plurilingual Policy Experimentation.
- ERIAC's Romani Čhib textbooks (A1–B1), written by authors from several dialect backgrounds using one common spelling – a practical example of harmonisation.
- Revitalisation of Romano-Kaló in Spain and Portugal, where a lost inflected Romani is being rebuilt from historical sources and Caló vocabulary (Hancock, Category B).
- Online lexical resources such as ROMLEX, which document many dialects side by side and show that diversity can be made visible rather than erased.

8.3 Challenges and Lessons Learned

Early testing and public commentary identified limitations including a strong orientation towards Vlax forms, inconsistent mixing of dialect material and uneven grammatical quality. These observations should be treated as an invitation to structured cooperation rather than as a rejection of the technology. ERIAC recommends periodic evaluation with native speakers and linguists from several dialect groups, testing everyday, educational, administrative and literary material in both translation directions and assessing adequacy, fluency, dialect consistency and potentially harmful output.

Lessons from the case study

Three lessons follow. First, technology moves faster than language planning: if Roma institutions and experts do not participate, algorithms may create de facto norms through data availability alone. Second, data quality and diversity are decisive; open, validated and dialect-documented resources are essential. Third, digital tools work best when they reinforce human infrastructure - teachers, writers, translators, researchers and cultural institutions - rather than attempting to replace it.

8.4 Implications for European Policy and Practice

- Language planning for Romani can no longer be separated from language technology. The harmonisation process (6.3) should explicitly address machine translation and AI.

[76] Sarău 2016.

[77] See 6.2; Kyuchukov et al. 2024.

- European institutions should treat Romani as a priority language in programmes supporting multilingual digital infrastructure and AI, and fund open Romani data resources.
- A structured dialogue between ERIAC, Romani linguists and technology companies should be established to improve dialect coverage, quality and safeguards against antigypsyist output.
- Schools and public services should be informed of both the usefulness and the limits of machine translation for Romani, and should not use it without human review.
- Good practices in education (Romania, fair assessment, CFR-based materials) show that the human infrastructure – teachers, materials, institutions – remains indispensable. Technology can amplify it, but not replace it.

9. CONCLUSIONS

Te na žanas katar aviljam, sar šay žanas karing kay žas angle? ("If we don't know where we came from, how can we know where we're going?")

Romani is a European language with a thousand-year history and a living presence across the continent. The expert contributions brought together in this paper show both continuity and complexity: a shared linguistic history and cultural core, but also centuries of contact, migration, multilingualism and dialect differentiation. These are not contradictory realities. They are the conditions from which any credible policy for the language must begin.

Ian Hancock's contribution demonstrates why language matters to shared historical consciousness and why common linguistic resources can strengthen cohesion across endaya. Huseyin Kyuchuk's (Hristo Kyuchukov's) research demonstrates, with equal importance, that Roma identity is not conditional on speaking Romani and that multilingualism is a normal part of Roma life. ERIAC's position brings these insights together: Romani is a central pillar of collective heritage, but it must never become a test of belonging.

The educational evidence is equally clear. Roma children who speak Romani arrive at school with a developed linguistic system. Education systems must learn to distinguish language difference from learning difficulty, recognise oral competence even where literacy is limited, and build additive pathways in which Romani and the majority language reinforce rather than exclude one another. The A1–B2 educational pathway, teacher development and a renewed European framework can turn this principle into sustainable practice.

Harmonisation is the bridge between protection and development. It provides the shared orthographic, grammatical and terminological infrastructure required for education, publishing, cross-border communication and digital technologies, while dialect documentation and community participation protect the plurality of the language. The digital transition makes this work urgent: AI systems will increasingly shape which forms are visible, searchable and reproducible.

Responsibility is therefore shared across European institutions, national governments, educational authorities, Roma cultural and civil-society organisations, universities and technology companies. ERIAC's role is to convene expertise, develop resources, connect institutions and demonstrate in practice that linguistic coherence and dialect diversity can be advanced together.

The future of Romani will be decided in the coming generation – in families, in classrooms and on screens. As Hancock's epigraph reminds us: "If we don't know where we came from, how can we know where we're going?"

10. POLICY RECOMMENDATIONS

The following recommendations are addressed to five groups of stakeholders. They are derived from the findings of chapters 5–8 and are intended to translate the paper's analysis into concrete areas for policy and cooperation.

10.1 Recommendations for European Institutions

1. Recognise Romani explicitly as part of Europe's linguistic and intangible cultural heritage in EU and Council of Europe policy documents on multilingualism, culture and Roma inclusion.
2. Support a European, Roma-led harmonisation process for Romani - including compatible orthographic conventions, a Common Romani Language Framework, coordinated terminology and dialect documentation - through cooperation among ERIAC, the Council of Europe, universities and Roma language experts.
3. Update the Curriculum Framework for Romani and related tools (ELP, Teacher's Handbook), with digital components, and fund their translation into further varieties.
4. Earmark funding for Romani language education, teacher training, publishing, media and translation in Erasmus+, Creative Europe, ESF+ and the EU programmes for Roma inclusion.
5. Include Romani as a priority language in EU initiatives on multilingual digital infrastructure, language data and AI, and fund open Romani text and speech corpora under community governance.
6. Monitor Romani systematically: encourage states to extend ECRML protection of Romani to Part III where justified, and collect comparable data on Romani speakers and teaching (e.g. through FRA Roma surveys).
7. Support the World Day of the Romani Language (5 November) as a European observance.

10.2 Recommendations for National Governments

8. Ratify the European Charter for Regional or Minority Languages where this has not yet been done, include Romani in its scope, and implement the recommendations of the Committee of Experts on Romani.
9. Guarantee in law the right of Roma pupils to learn Romani and to receive instruction about Roma history and culture, with adequate hours, qualified teachers and funding.
10. Adopt national Romani language strategies or action plans, developed with Roma communities and linked to national Roma strategic frameworks under the EU framework 2020–2030.
11. End school segregation and the disproportionate placement of Roma children in special education. Monitor placement data, and require linguistically valid assessment before any classification (see 10.3).
12. Include questions on Romani language use in censuses and surveys, with safeguards, to allow evidence-based policy.

13. Support Romani-language media, publishing, theatre and cultural production, and the use of Romani in public services where there is demand.
14. Cooperate across borders on harmonisation, materials and teacher training, including through ERIAC.

10.3 Recommendations for Educational Authorities

15. Desegregation: education systems should monitor and address the disproportionate concentration of Roma children in segregated settings and ensure meaningful access to inclusive mainstream education, rather than directing children towards separate pathways because of linguistic or socio-economic characteristics.
16. Bilingual language profiles: at school entry, children's complete linguistic backgrounds should be documented.
17. Linguistically valid assessment: assessment should distinguish second-language acquisition from learning difficulties. Where Romani is a child's primary or significant language, assessment of language and educational potential should include Romani, and limited knowledge of the school language should never in itself be taken as evidence of cognitive or learning disability.
18. Appropriate assessment instruments: instruments should be linguistically and culturally appropriate and sensitive to dialect variation.
19. Monitoring of placement: schools and education authorities should examine whether language differences contribute disproportionately to special-school placement.
20. Romani-language education and literacy: where communities and families seek it, opportunities to learn Romani, including formal literacy, should be developed, building on children's existing oral competence.
21. Romani in teacher education: teachers should receive training on Romani linguistic diversity and bilingual development.
22. Roma history and the Roma Holocaust: education on Roma history and the Roma Holocaust should be part of the curriculum, to challenge the invisibility of Roma experiences and provide a more complete understanding of European history.
23. Develop curricula and textbooks for Romani at all levels – pre-school, primary, secondary and adult education – aligned with the Curriculum Framework for Romani and the common orthography, with dialect variants represented.
24. Create initial and in-service teacher training for Romani teachers (university programmes, certification), and modules on bilingual development, Romani linguistic diversity and antigypsyism for all teachers.
25. Recruit and train Roma teachers, mediators and inspectors, and give them recognised status and career paths.
26. Implement Recommendation CM/Rec(2020)2: include the history of the Roma, including the Holocaust, in curricula and teaching materials for all pupils.
27. Offer Romani courses for adults and young people who did not acquire Romani at home.
28. Support Romani at university level, including Romani studies and teacher-education programmes.

10.4 Recommendations for Roma Civil Society and Cultural Institutions

29. Campaign for language pride and family transmission, addressing internalised shame (Hancock, Category A): “Amari čhib si amari zor – our language is our strength.”
30. Use Romani actively in cultural production, media, social media, events and organisational communication, and support Romani writers, translators and journalists (e.g. through the Romani PEN Centre).
31. Take ownership of harmonisation: participate in the European process, test and adopt the common orthography, and contribute to terminology work, following Hancock’s call for a Romani-dominant team of specialists and his proposed codification steps (6.4–6.5).
32. Document oral heritage – stories, songs, testimonies, dialects – in cooperation with archives and universities.
33. Promote unity across endaya by valuing all dialects and by making the shared history of the Romani people known (5.1).
34. Engage young people as creators of digital content in Romani, and as future teachers, linguists and technologists.

10.5 Recommendations for the Digital and Technology Sector

35. Co-create Romani language technologies with Romani linguists and community organisations, including ERIAC, from data collection to evaluation.
36. Extend machine translation and AI support to the main dialect groups. Indicate clearly which variety is produced, and allow users to choose.
37. Be transparent about training data, obtain consent, and respect community data governance for Romani texts and recordings, especially oral heritage.
38. Test outputs for antigypsyist bias and stereotypes; include Romani in content-moderation capacity.
39. Provide Romani keyboards, spell-checkers and interface localisation implementing the common orthography, on mainstream devices and platforms.
40. Support open Romani resources (corpora, dictionaries, speech data) and offer training and employment opportunities for young Roma in language technology.

11. CONTRIBUTORS

Ian Hancock – Professor Emeritus of Linguistics, The University of Texas at Austin; Author / Expert Contributor: Historical Linguistics, Romani Language and Identity

Huseyin Kyuchuk (Hristo Kyuchukov) – Professor, Psycholinguist and Romani Linguist; Author / Expert Contributor: Sociolinguistics, Psycholinguistics and Romani Language Education

Mihaela Zătreanu – Director of ERIAC Romania; Leader of ERIAC’s Romani Language Initiative; PhD Candidate in Sociology, Romanian Academy; Textbook Author and Institutional Lead: Harmonisation, Common Romani, Education, Language Policy and Digital Development

12. BIBLIOGRAPHY

Achim, V. (2004). The Roma in Romanian History. Budapest: Central European University Press.

Acković, Dragoljub (1997) Čitajte Ljudi—Ginavnen Romalen—Read People. Belgrade: Rrominterpress.

Acton, T. (2021 [1974]). Gypsy Politics and Social Change. Abingdon: Routledge (Routledge Revivals; first published London: Routledge & Kegan Paul, 1974).

Boretzky, N., & Igla, B. (2004). Kommentierter Dialektatlas des Romani. 2 vols. Wiesbaden: Harrassowitz.

Council of Europe (1992). European Charter for Regional or Minority Languages (ETS No. 148). Strasbourg.

Council of Europe (1995). Framework Convention for the Protection of National Minorities (ETS No. 157). Strasbourg.

Council of Europe (2005). Framework Convention on the Value of Cultural Heritage for Society (Faro Convention, CETS No. 199). Faro.

Council of Europe (2010) Third report submitted by Slovenia Pursuant to article 25, paragraph 1 of the framework convention for the protection of national minorities. Received on 28 April 2010. Strasbourg.

Council of Europe (2008). Curriculum Framework for Romani; European Language Portfolios for Romani; Teacher's Handbook. Strasbourg: Language Policy Division.

Council of Europe (2020). Recommendation CM/Rec(2020)2 on the inclusion of the history of Roma and/or Travellers in school curricula and teaching materials.

Council of Europe (2022). Romani–Plurilingual Policy Experimentation (2022–2025). Strasbourg: Education Department.

Council of Europe (2026). Strategy for Roma and Traveller Inclusion 2026–2030.

Council of Europe (2025). ECRML Secretariat. Factsheet: Romani and the European Charter for Regional or Minority Languages. November 2025.

Council of Europe, Project "Education of Roma Children in Europe". Factsheets on Romani History: The Austro-Hungarian Empire. Strasbourg.

Council of the European Union (2021). Council Recommendation of 12 March 2021 on Roma equality, inclusion and participation (2021/C 93/01).

ECML (2013). QualiRom – Quality Education in Romani for Europe (2010–2013). Graz: European Centre for Modern Languages.

Elšík, V., & Albert, G. (2024). "The current state of the Romani language in the Google Translate program and its prospects." Romea.cz, 1 July 2024.

Elšík, V., & Beníšek, M. (2020). "Romani dialectology." In Y. Matras & A. Tenser (eds.), The Palgrave Handbook of Romani Language and Linguistics (pp. 389–427). Cham: Palgrave Macmillan.

ERIAC Statutes and the 2025 Annual Report.

ERIACNET4EU / CERV. Framework partnership agreement, 2022.

ERIAC (2024). Position Paper with Policy Recommendations on Roma Remembrance, Commemoration and History Teaching. Berlin: ERIAC (ERIACNET4EU).

ERIAC. Romani Language Initiative / Romani Čhib; Romani Čhib A1, A2, B1 and B2 development materials; Safeguarding Our Romani Language conference materials; ERIACNET4EU materials.

ERIAC (2026). AMARI ČHIB – AI-Powered Romani Language Preservation: Project Concept Note for Microsoft's AI for Good Lab, 15 April 2026.

European Commission (2020). EU Roma strategic framework for equality, inclusion and participation 2020–2030. COM(2020) 620.

European Court of Human Rights (2007). D.H. and Others v. the Czech Republic [GC], no. 57325/00, judgment of 13 November 2007.

European Union (2012). Charter of Fundamental Rights of the European Union, Article 22. OJ C 326, 26.10.2012.

European Union (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act).

Friedman, V. (1999) The Romani Language in the Republic of Macedonia: Status, Usage and sociolinguistic perspectives. Acta Linguistica Hungarica, 46 (3-4), 317-339.

Gjerdman, O., & Ljungberg, E. (1963). The Language of the Swedish Coppersmith Gipsy Johan Dimitri Taikon. Uppsala: Lundequistska.

Google (2024). "110 new languages are coming to Google Translate", 27 June 2024.

Halwachs, D. W. (2002). Burgenland-Romani. München: Lincom Europa.

Hancock, I. (2000). "The emergence of Romani as a koïné outside of India." In T. Acton (ed.), Scholarship and the Gypsy Struggle: Commitment in Romani Studies. Hatfield: University of Hertfordshire Press.

Hancock, I. (2002). We Are the Romani People / Ame sam e Rromane džene. Hatfield: University of Hertfordshire Press.

Horvat Muc, J., & Đurić, R. (2012). Standardizacija romskega jezika: Dodatek k učnemu pripomočku. Zveza Romov Slovenije.

John, V. A. (2013). A Reexamination of the Origins of Romani. Master's thesis, The University of Texas at Austin.

Jusuf, Š., & Kepeski, K. (1980). Romani gramatika – Romska gramatika. Skopje: Naša Kniga.

Kali Sara & International Romani Union (2009). Declaration of the World Symposium for the Standardisation and Codification of the Romani Language, Zagreb, 3–5 November 2009.

Kochanowski, J. (1995). "Romani language standardisation." Journal of the Gypsy Lore Society, 5(2), 97–108.

Kochanowski, Jan (1971) "The Future of Romani". In Thomas A. Acton (ed.) Current Changes amongst British Gypsies and their Place in International Patterns of Development. Oxford: Romanestan Publications, pp.76–77

Kyuchukov, H. (1995). Romany Children and Their Preparation for Literacy: A Case Study. Tilburg University Press.

Kyuchukov, H. (1999). "Teacher–pupil interaction in Bulgarian Romani classrooms." European Journal of Intercultural Studies, 10(2), 151–162.

Kyuchukov, H. (2006). "Code-switching among Muslim Roms in Bulgaria." International Journal of the Sociology of Language, 2006(179), 41–51.

Kyuchukov, H. (2014). “Romani language assessment of Roma children.” Journal of Language and Cultural Education, 2(2), 52–64.

Kyuchukov, H. (2018). “On the regional standardization of Romani in the Balkans.” Linguistique Balkanique, 57(2), 167–188.

Kyuchukov, H. (2023). “Romani language, linguistic rights and ‘Antigypsyism’.” Social Communications: Theory and Practice, 15(1), 131–150.

Kyuchukov, H., & de Villiers, J. (2014). “Roma children’s knowledge on Romani.” Journal of Psycholinguistics, 19, 58–65.

Kyuchukov, H. (2015) Romani language education in Bulgaria. In: Kyuchukov, H. Kwadrans, L. and Fizik, L. (eds.) *Romani Studies: Contemporary Trends*. Munich: Lincom, pp. 66-78.

Kyuchukov, H., & Kwadrans, Ł. (2024). “Use of language and identity of the Muslim Romanies on the Balkans.” Problemi na Ustnata Komunikatsia, 13(1), 254–268.

Kyuchukov, H., & New, W. (2016). “Diversity vs. equality: Why the education of Roma children does not work.” Intercultural Education, 27(6), 629–634.

Kyuchukov, H., & New, W. (2018). “Mother tongue of Roma children from special schools.” RUDN Journal of Psychology and Pedagogics, 15(3), 323–333.

Kyuchukov, H., & New, W. (2019). “Roma school mediators and the bilingualism of Roma students.” Edukacja Międzykulturowa, 10(1), 81–100.

Kyuchukov, H., & Samuilov, S. (2019). “Language use and identity among migrant Roma.” East European Journal of Psycholinguistics, 6(1), 47–57.

Kyuchukov, H., de Villiers, J., & Tabori, A. T. (2017). “Why Roma children need language assessments in Romani.” Psychology of Language and Communication, 21(1), 215–243.

Kyuchukov, H., de Villiers, J., Zhu, Y., & Zhong, I. (2024). “Assessing the acquisition of Romani in Roma children.” Language Acquisition, 32(2), 153–188.

Kyuchukov, H., Samko, M., & Kopcanova, D. (2017). “Knowledge of Romani language grammar.” Psychology in Russia: State of the Art, 10(4), 76–85.

Lacková, E. (1999). A False Dawn: My Life as a Gypsy Woman in Slovakia. Hatfield: University of Hertfordshire Press.

Lee, J. (2024). “Google Translate and the future of the Romani language.” The Norwich Radical, 29 September 2024.

Marsh, A. (2004). Ghaznā: Koïnés and Cavalry Charges – The Military Origins of the Romani Language and People. Research monograph, Bilgi University, Istanbul.

Marsh, A. (2008). “No Promised Land”: History, Historiography and the Origins of the Gypsies. Doctoral dissertation, University of Greenwich.

Matras, Y. (1999). “Writing Romani: The pragmatics of codification in a stateless language.” Applied Linguistics, 20(4), 481–502.

Matras, Y. (2002). Romani: A Linguistic Introduction. Cambridge University Press.

Matras, Y. (2005a). “The classification of Romani dialects: A geographic-historical perspective.” In B. Schrammel, D. W. Halwachs & G. Ambrosch (eds.), General and Applied Romani Linguistics. München: Lincom Europa.

Matras, Y. (2005b). The Status of Romani in Europe. Report submitted to the Council of Europe Language Policy Division.

Matras, Y. (2005c). “The future of Romani: Toward a policy of linguistic pluralism.” Roma Rights. Budapest: European Roma Rights Centre.

Matras, Y. (2013). “Mapping the Romani dialects of Romania.” Romani Studies 5, 23(2), 199–243.

Matras, Y. (2015). “Transnational policy and ‘authenticity’ discourses on Romani language and identity.” Language in Society, 44(3), 295–316.

Matras, Y. (2021). “The standardization of a stateless language.” In W. Ayres-Bennett & J. Bellamy (eds.), The Cambridge Handbook of Language Standardization (pp. 645–664). Cambridge: Cambridge University Press.

Media-Marketing (2021). Aresel campaign for the inclusion of Romani in Google Translate. Media-Marketing.com, 10 May 2021.

Moseley, C. (ed.) (2010). Atlas of the World’s Languages in Danger (3rd ed.). Paris: UNESCO Publishing.

Muratori, L. A. (1754). Annali d’Italia. Rome: Barbiellini.

Paspati, A. G. (1870). Études sur les Tchinghianés ou Bohémiens de l’Empire ottoman. Constantinople: Antoine Koroméla.

Rehm, G., & Way, A. (eds.) (2023). European Language Equality: A Strategic Agenda for Digital Language Equality. Cham: Springer.

Sampson, J. (1926). The Dialect of the Gypsies of Wales. Oxford: Clarendon Press.

Sarău, G. (2016). “Planificarea lingvistică pentru limba rromani între anii 1990–2011, în România.” Analele Universității „Ovidius” din Constanța. Seria Filologie, 27(2).

Toporov, V. G. (2009). Crimean Roma: Language and Folklore. Ivanovo: Unona Publishing.

UNESCO (2003). Convention for the Safeguarding of the Intangible Cultural Heritage.

University of Manchester, Romani Project. Resources on the status, dialectology and documentation of Romani; ROMLEX resources with the University of Graz.

Zătreanu, M. (2022). “The importance of Romani language learning and teaching for a more inclusive education for all.” Council of Europe, INSCHOOL – Inclusive Schools: Making a Difference for Roma Children, 11 May 2022.

ERIAC Position Paper
Safeguarding and Protecting the Romani Language
2026

ERIAC
EUROPEAN ROMA INSTITUTE FOR ARTS AND CULTURE

ERIAC NET 4EU

Co-funded by the European Union