THE CHALLENGES OF LANGUAGE CODIFICATION IN A TRANS-NATIONAL CONTEXT

YARON MATRAS

1. BACKGROUND

Romani is now the established reference term for the language referred to by its speakers as *řomanes* ('in a Rom way') or *řomani čhib* ('the Romani language'). It is a language of Indic origin, and the only New Indo-Aryan language spoken exclusively in Europe since medieval times. In the absence of written documentation on the language prior to the sixteenth century, historical discussions of Romani have relied on a comparison with documented varieties of medieval Indo-Aryan languages (e.g. already Pott 1844, Turner 1926, Sampson 1926) for the early period, and, for later periods, on a comparison of the various dialects of Romani itself (cf. Matras 2002, Elšík & Matras 2006). The consensus view in academic discussions of the language[1] is that Romani developed out of an idiom that shared its most ancient isoglosses with the group of Central Indo-Aryan languages (such as Hindi), but maintained certain archaisms, alongside innovations that are typical of the Northwestern Indo-Aryan languages (such as Kashmiri). This is interpreted as evidence of a migration to this region, possibly around the middle of the first millennium. Romani then moved to the Byzantine Empire, where it was spoken, in close contact with Greek, possibly from around the 10th century onwards. In earlier discussions, layers of Armenian, Iranian, and some Caucasian loans in Romani had been interpreted as evidence of a gradual migration westwards, interrupted by periods of settlement in the respective regions (i.e. Iran, Armenia; cf. Miklosich 1872-1880). Current approaches, however, point out the possibility that elements from these languages were acquired alongside the heavy Greek component in what was at the time a multilingual region in Anatolia (cf. Matras 2002).

---

[1] Non-academic publications have played a major role in recent years in disseminating a politically constructed view of the language as reflecting, allegedly, the origin of its speakers in a population of 'Rajput' warriors who were assembled in medieval India in order to resist Islamic invasions, were captured and brought to Europe against their will, and turned into marginalised and often itinerant, service-providing economies (see Hancock 2002 as an example).

The Balkans were the home of Romani throughout the Byzantine period, and it is in this region that we still find the highest density of Romani-speaking populations. With the decline of the Byzantine Empire, Romani groups began to migrate westwards and northwards. The period of migrations in the 14th and 15th centuries was succeeded by a period of settlement during which the major dialect differences are believed to have been formed. The nature of internal innovations in Romani dialects and their geographical distribution patterns tend to point to a rather recent differentiation based on a more or less uniform inventory of structures, termed Early Romani. This stage preceded the dispersion of the groups. Innovations thus tend to be local or regional, but in any event geographically contained. This makes it highly plausible that they emerged *in situ*, after settlement in the regions. The major difference between the dialects therefore do not represent ancient divisions within the Romani-speaking population prior to the migration from the Balkans westwards. The earliest written records of Romani are fragmented sources from the 16th and 17th centuries, and are followed by a plethora of wordlists from the 18th century. These sources already represent a picture of dialectal diversity that is very close to the one we find today. This indicates that the major changes leading to dialect differences evolved in the period between the 15th-17th centuries, when the groups first became separated from one another through settlement within strict political borders. In part, the differences between the dialects can be traced to the influences of distinct contact languages, such as Turkish, Slavonic languages, Hungarian and German.

Despite their diversity, Romani varieties form a continuum of dialects that are mutually intelligible, at least in direct relation to their geographical distance from one another, and to some extent also to the source of borrowed lexical and morpho-lexical material. The most pronounced division between the dialects is a dense bundle of isoglosses that stretch roughly between the northern Adriatic coast toward the northeast, cutting through Slovenia and Hungary, and ultimately reaching transcarpathian Ukraine. This line represents the Habsburg-Ottoman frontier zone in the period during which structural innovations left their mark on both northern and southern dialects of Romani; the political border prevented their diffusion in either direction across the line. Mutual intelligibility between the dialects that emerged to the west of this frontier zone (German, Welsh, and Scandinavian Romani), and those to the east, is limited. On the other hand, the dialects of Romani that we encounter on the south-to-north axis in the east, from

southern Greece and up to the Baltics, represent a very gradual continuum of varieties that are fully comprehensible to other speakers within a distance of some 800 kilometres or even more.

2.   THE SOCIOLINGUISTICS OF ROMANI

Use of Romani has traditionally been limited to oral communication within the family and extended family, as well as with other families within the local Romani community. Encounters with other Roms have usually been much less frequent, and social barriers among local communities or clusters of extended families often meant that Romani was rarely used among individuals who did not know one another personally, or at least by family affiliation. In other words, the traditional domain in which Romani was used can be said to have been limited to a large extent to situations in which, in response to the question "who is your father?" a speaker can provide the interlocutor with an answer that is meaningful to both. This means that Romani was traditionally absent from most domains involving business, from communication with institutions or local services, from education and the media. In recent generations, Romani has even experienced a retreat from some of its more traditional domains. For many younger community members, it is no longer the default choice when speaking to neighbours who are not family members, even if they are Roms. Moreover, Romani is often used alongside the majority (or another minority) language even within the home, and in some families, in some regions, intense interaction with non-Roms leads to a preference in favour of the non-Romani language among the younger generations even within the family domain.

By contrast, the rise of an intellectual Romani movement from the late 1960s onwards, and the proliferation of organised Romani cultural and political activities especially since the political transition in eastern Europe in 1990, have enabled the language to enter into new domains of use. There are several hundred book publications in Romani, some of them translations of literary works, others original texts, as well as magazines and educational materials. The latter have not yet entered the curriculum in any systematic way, except in Romania, where Romani is widespread as a subject in elementary and secondary schools. But classes are held on an experimental basis throughout Europe. Romani-language broadcasting enjoys the support of state media in sev-

eral countries, including Finland, Sweden, Hungary, Germany, Serbia, and Macedonia, and some broadcasts are accessible online to an international audience. A number of film and theatre productions in Romani has also helped the language increase its presence in the public domain.

Alongside this expansion into domains of communication from which it was previously absent, it is possible to recognise several domains of interaction that constitute new communicative functions, and which have been created in order to promote Romani within the public sphere (see Halwachs 2005). These include the production of Romani folklore, political and religious manifestations and institutions, as well as efforts to safeguard and codify the language itself. While these activities contribute to the presence of Romani in the public sphere, they do not, strictly speaking, show that the language is taking on communicative functions that were relevant to the speaker community but were so far handled in the majority language, but rather, that a sector of language activists within the community is engaged in constructing new communicative functions and interactional domains in order to promote the use of Romani. One of the most prolific uses of the language outside the traditional domains is its inclusion on websites devoted to Romani culture and public life, and its use in email communication among language, culture and political activists. These are also the areas in which Romani is most widely encountered in written form. Rather than constitute a case of top-to-bottom codification, Romani literacy is thus largely a bottom-up process, with individuals tailoring norms and conventions to their communicative needs and the expectations of their immediate addressees.

3. ISSUES OF STATUS

No verifiable figures about the number of Romani speakers are available. Estimates vary from the most conservative figure of around 3.5 million to a maximum estimate of up to 8 million. Either way, Romani is one of the larger minority languages in Europe, and, since its enlargement in 2004, also within the European Union. The largest speaker populations are found in southeastern Europe, especially in Romania, Bulgaria, Turkey, Macedonia, and Serbia and Montenegro, as well as in Greece, Slovakia, Moldavia, and Hungary. Sizeable Romani-speaking populations exist in most other countries of central and east-

ern Europe. Romani communities that settled in the westernmost 'fringe' countries, such as Portugal and Spain, the United Kingdom, and the Scandinavian countries (with the exception of Finland), some 5-6 centuries ago, have abandoned Romani and have adopted the majority language (albeit retaining some Romani vocabulary in group-internal conversation). Speaker communities in these regions consist predominantly of later immigrants from central or eastern Europe.

Romani language policy has been on the agenda for several decades now, as the subject of a series of international recommendations as well as state legislation. The Council of Europe has assumed a leading role in the process of granting Romani international recognition. In 1981, the Standing Conference of Local and Regional Authorities called on member states to recognise Romanies as ethnic minorities and to grant them "the same status and advantages as other minorities enjoy, in particular concerning respect and support for their own culture and language" (Resolution 125 (1981) on the role and responsibility of local and regional authorities in regard to the cultural and social problems of populations of nomadic origin). In 1983, the Council of Cultural Co-operation recommended that "the Romany language and culture be used and accorded the same respect as regional languages and cultures and those of other minorities". In 1989, the Council of Ministers of Education declared as its aim to promote teaching methods and teaching materials that "give consideration for the history, culture and language of Gypsies and Travellers" and to encourage research on those topics (Resolution 89/C, 153/025). In 1993, the Council of Europe's Parliamentary Assembly called for the establishment of "a European programme for the study of Romanes and a translation bureau specialising in the language", and recommended that "the provisions for non-territorial languages as set out in the European Charter for Regional or Minority Languages should be applied to Gypsy minorities" (Rec. 1203/1993). In February 2000, the Committee of Ministers recommended to the member states that "in countries where the Romani language is spoken, opportunities to learn in the mother tongue should be offered at school to Roma/Gypsy children", and that "the participation of representatives of the Roma/Gypsy community should be encouraged in the development of teaching material on the history, culture or language of the Roma/Gypsies" (Res. 2000/4).

Romani has a recognised status deriving either directly (through explicit reference) or indirectly (through general reference to minority languages) from the constitutions of several European countries,

including Macedonia, Austria, Finland, and Hungary. In the European Charter for Minority and Regional Languages of 1992, Romani (Romany) is mentioned explicitly as an example of a non-territorial language, to which some parts of the Charter may be applied. Of the eighteen states that have ratified the European Charter for Regional or Minority Languages, ten apply it to Romani: Austria, Croatia, Finland, Germany, Hungary, the Netherlands, Norway, the Slovak Republic, Slovenia, and Sweden. In many of these countries, however, concrete implementation of the Charter in respect of Romani is yet to be achieved. In some countries, concrete measures have been implemented, in collaboration with community representatives and experts, such as weekly or even daily Romani-language broadcasting, provision of translation and interpreting facilities, financial support for the creation of teaching materials and language documentation, the appointment of coordinators for a Romani curriculum, and experimental introduction of Romani into the teaching curriculum (usually, however, on an irregular basis). On the whole, however, most initiatives to promote Romani in culture, media, and education, still originate in, and are run by, NGOs. Romani is still largely absent from the state-run education system. The strongest practical support for Romani within the education system is found in Romania, where a national Romani language curriculum was adopted in 1999, and has since received widescale implementation, through all levels, from pre-school to higher education. Initiatives to introduce a centralised curriculum are underway in a number of countries.

4. PROBLEMS OF CODIFICATION

The earliest efforts to codify Romani were made in the Soviet Union in the 1920s, coupled with the production of numerous literary, political, and educational texts, all in the North Russian dialect of Romani, in an adapted form of the Cyrillic script (see documentation on http://www.rombiblio.ru). Since the early 1970s, the creation of a single Standard Romani has been on the agenda of a small, international circle of Romani activists (cf. Hancock 1993, Courthiade 1989). None of the suggestions made, however, succeeded in gaining any wide basis of supporters. In 1990, some 40 activists who assembled at a meeting of the International Romani Union in Warsaw voted to confirm, for a trial period, an alphabet drafted by French language enthusiast Marcel

Cortiade (cf. Kenrick 1996). In the aftermath of the Warsaw meeting, much effort was put into further developing and disseminating this alphabet, along with a standardisation concept that included the creation of technical and political vocabulary, and a strategy that included the creation of a dictionary, a Romani encyclopaedia, and educational materials. Within a relatively short period of time, however, the concept was abandoned by most of those who had signed the Warsaw resolution. Its main success has been its imposition as a Standard by the Romanian Ministry of Education. Outside Romania, the concept is used by several individual writers in Albania and Macedonia. Its popular rejection is connected in part to its reliance on a set of abstract, meta-phonemic symbols, such as {θ} to represent both /d/ and /t/ in the Romani nominal ending *-te/-de* (both written *-θe* in this system), {ʒ} to represent inter-dialectal variation between sounds, and more. Noteworthy is also an attempt to create an artificial form of Romani for written purposes in Spain; this variety, called *Romanó-Kaló* by its authors, is a blend of Romani lexical roots and selected grammatical features, taken from various dialects and adapted to Spanish orthography and sometimes even grammatical rules. It has been in use since the early 1990s in a series of political publications by a Spanish Romani NGO.

The bulk of Romani language publishing has a regional or local orientation, and lays no claim to becoming an international Standard (for an overview see also Matras 1999). This kind of process faces firstly the challenge to design a writing system for the language. The choice of most symbols is straightforward, since almost all codification attempts employ the Roman alphabet. Problems arise with those sounds that are historically absent in Latin, and for which modern European languages tend to find diverse solutions, most notably the sounds /ʃ, tʃ, ʒ, dʒ, x, j/, and palatalisation. In addition, Romani has aspirated sounds, which are generally represented as {ph, th, kh, čh}. Another challenge is to identify the variety or dialect that is to be written. The choice here is, for almost all writers, also straightforward: they employ their own variety, which is usually the variety of the local or regional Romani community that constitutes their immediate target audience. Finally, there is the challenge of finding the resources needed to produce, print and disseminate written work in Romani. So far, the bulk of literacy activities has been carried out within the Romani NGO sector, financed by grants from foundations, international organisations, and sometimes local authorities and governments. Although there are exceptions, most products have been distributed free of charge within a rather closed network of contacts.

One of the earliest examples of regional codification of Romani was
the production of texts in Czechoslovakia in 1968, in the East Slovak
variety of Romani, using an alphabet based on that of Czech, featuring
the diacritics {š, č, ž, d', l', ň, t'}. The concept was revived in the early
1990s and is now used in numerous Romani-language publications, in
print and on the internet, including media, educational and literary pro-
ductions, in both the Czech and Slovak Republics (cf. Hübschmannová
1995, Hübschmannová & Neustupný 1996). In Macedonia, a writing
system based on the Roman script, using either the Arli or Džambazi
(Gurbet) variety of Romani, was propagated in the 1970s in a manu-
script by Jusuf & Kepeski, later published as a book in 1980. A related
concept was adopted by the 1992 Romani Language Standardisation
Conference sponsored by the Macedonian government, and is in use
today (cf. Friedman 1999). In neighbouring Serbia, a closely related sys-
tem is used, employing the Roman script with diacritics {š, č, ž, ć, đ},
and based largely on the Gurbet dialect. In Hungary, Romani-language
publications use the Lovari dialect and a writing system which, unlike
Hungarian, uses the combinations {sh ch zh} and avoids indication of
vowel length, while on the other hand Hungarian symbols for palatals
{ny gy ty} are included. In Finland, the Finnish or Kaale dialect of
Romani is used as a written language, based on an alphabet designed in
the early 1970s, and inspired primarily by the use of diacritics {š, č, ž,}
in international academic transliteration conventions. The same idea is
behind the choice of a somewhat similar graphemic representation for
texts in the Kelderash-dialect, produced in Sweden in the 1980s-1990s,
and for materials in the Lovari dialect, produced in Norway in the mid-
1990s. In Bulgaria, a similar system, using the Roman rather than
Cyrillic script, has been in use by some authors, though the majority
now seem to be moving toward a system that avoids diacritics, using
instead the grapheme combinations {sh, ch, zh}. The characteristic fea-
ture of the Bulgarian writing system for Romani is the presence of {w},
representing a central vowel. Writers in Bulgaria choose a variety of
dialects, which is characteristic of the considerable dialectal variation of
Romani in this country, though the most common choices are the Erli
dialect of the Sofia region, and the so-called 'Drindari'-type dialect spo-
ken in the central regions of Bulgaria.

In Austria, the Burgenland Romani variety was codified by a team of
linguists from Graz University in cooperation with community repre-
sentatives, using a German writing system, thus {sch, tsch, dsch} and so
on, while other varieties, such as those of the Lovara, Kelderash, and

Arlija, tend to use the international academic conventions (cf. Halwachs 2005). Apart from a handful of writers in Albania and Macedonia, only the Ministry of Education in Romania has adopted the Cortiade (Warsaw 1990) writing system, which it uses consistently for Romani in numerous school textbooks, readers, and other educational material, including at university level. Independent authors writing Romani in Romania, however, use their own systems, often a blend between the Romanian writing system and features of the international academic transliteration conventions. The most commonly used dialect in Romani publication in Romania is related to the Kelderash and other so-called 'Vlax' dialects of Romani.

On the whole, then, it is possible to make the following generalisations about the writing systems used for Romani: They tend to constitute a mixture of international writing conventions, sometimes combined with selected features of the alphabets of the respective national languages. As far as the international conventions are concerned, there are two main tendencies. The first is to use the diacritics {š, č, ž,} along with other conventions, such as values of {x, j, ph, th, kh}. These derive, after all, from a consistent and more or less uniform system of transliteration which is commonly used for the various dialects of Romani in Romani linguistics, as well as in numerous applied publications such as grammars and dictionaries, including the multi-dialectal online Romani lexicon Romlex (http://romani.kfunigraz.ac.at/romlex/). The second option is to avoid diacritics, and to use the grapheme combinations {sh, ch, zh} instead. This is the preferred option in most email communication, due to difficulties in the transmission of diacritics in some applications. It is widespread in Hungary and recently also in Bulgaria, as well as in the individual use of authors and translators (and appears to be the preferred option for translators working for international organisations, such as the Council of Europe or OSCE, or international NGOs, such as ERRC).

5. BETWEEN STANDARD AND PLURALISM

The concept of language standardisation, embraced by the International Romani Union in its early days, regarded written Romani language as a symbol of political unity, but also as a token of loyalty to a centralised policy of cultural emancipation. This concept has failed to win popular support among Romani language and culture activists and

writers in the communities. This rejection was inevitable, for several reasons. Romani society has traditionally lacked any rigid hierarchical social structure, and there is therefore no dominant class, and so also no dominant form of speech that is acceptable to all. The geographical dispersion of Romani communities means that the propagation of a standard cannot be controlled by imposing sanctions or measures of exclusion on those who do not use it, and the absence of resources means that few incentives can be given in order to attract users. On the other hand, Romani language activists and educators value in the first instance the support of their immediate communities, and it is they who constitute their target audience. It is noteworthy however that although political unity (in the sense of a joint pursuit of issues of common concern) is on the agenda of most associations and initiatives, most do not regard linguistic diversity as an obstacle to unity.

Pluralism thus represents the overwhelming trend on the ground, with written Romani showing regional codification with some international orientation. A new generation of Romani intellectuals is exposed to various forms of the language, both oral and written, through encounters with other Roma at international conferences, through internships and training seminars in the NGO domain, and through regular email communication and text messaging. It is especially via the latter two media that writing, including trans-national correspondence, has acquired a new position in the daily communication patterns of individuals. This generation can accept, comprehend and make creative use of different forms and varieties of the Romani language. For the younger generation especially, the future of any Romani-language movement, flexibility in respect of the written language is supported by their practical experience as composers of text-messages and users of email and chat-rooms, all of which are media that stress that the practical reward in communication is being understood, rather than adhering to a fixed norm. New technologies support this trend: it is easy to search and replace one spelling convention by another in electronic texts in order to adapt a text to a different audience or setting. The Romlex online translator even incorporates an approximation algorithm that allows the user to look up words in a spelling form that is different from the one displayed by the authors of the facility – an impossibility in any printed dictionary.

6. OUTLOOK

In some respects, Romani is in a unique position: it is a non-territorial language; its speakers do not constitute the majority in any one region, and are not concentrated in any particular region, either. There are no community institutions that are responsible for the population of speakers as a whole, and even very few institutions at the local level. There is no widespread tradition of literacy in the language, and in fact most Romani communities are characterised by high illiteracy rates altogether. There is no social sector within the Romani community that has power or social prestige over others, and no obvious choice for a prestige variety. There is, finally, no recognised authority that could plan or implement a language policy. Instead, Romani codification efforts, as well as efforts to safeguard the language through status planning, constitute an unusual bottom-up process. This process is characterised by pluralism of forms and solutions to corpus planning, and by local initiatives engaged in status planning. Nevertheless, the proliferation of Romani language activities over the past few years, and the emergence of a growing lobbying power for Romani affairs around multilateral institutions such as the Council of Europe, the World Bank, the EU-Commission, the OSCE, and various foundations, have contributed to the gradual formation of a strong network of activists. These have been coordinating many of their activities across national borders. The fact that their primary sources of support – both funding, and political support – are international, rather than national, makes this kind of effort more rewarding than lobbying at the national level. On the whole, then, it is quite possible that the emerging model of codification in Romani may set a new kind of norm in language planning: trans-national networking of activities, as an umbrella for local solutions.

REFERENCES

Courthiade, M., 1989, *La langue Romani (Tsigane): Évolution, standardisation, unification, réforme*. In: Fodor, I./Hagège, C. (eds.), *Language reform. History and future. Vol. IV*, Hamburg, Buske: 79-109.

Elšík, V./Matras, Y., 2006, *Markedness and language change. The Romani sample*, Berlin, Mouton de Gruyter.

Friedman, V. A., 1999, *The Romani language in the Republic of Macedonia: status, usage, and sociolinguistic perspectives*. «Acta Linguistica Hungarica» 46: 317-339.

Halwachs, D. W., 2005, *Roma and Romani in Austria*. «Romani Studies», 5th series, 15: 145-173.

Hancock, I. F., 1993, *The emergence of a union dialect of North American Vlax Romani, and its implications for an international standard*. «International Journal of the Sociology of Language» 99: 91-104.

Hancock, I. F., 2002, *We are the Romani people*, Hatfield, University of Hertfordshire Press.

Hübschmannová, M., 1995, *Trial and error in written Romani on the pages of Romani periodicals*. In: Matras, Y. (ed.), *Romani in contact: the history and sociology of a language*, Amsterdam, Benjamins: 189-205.

Hübschmannová, M./Neustupný, J. V., 1996, *The Slovak-and-Czech dialect of Romani and its standardization*. «International Journal of the Sociology of Language» 120: 85-109.

Kenrick, D., 1996, *Romani literacy at the crossroads*. «International Journal of the Sociology of Language» 119: 109-123.

Matras, Y., 1999, *Writing Romani: The pragmatics of codification in a stateless language*. «Applied Linguistics» 20: 481-502.

Matras, Y., 2002, *Romani: A linguistic introduction*, Cambridge, Cambridge University Press.

Miklosich, F., 1872-1880, *Über die Mundarten und Wanderungen der Zigeuner Europas X-XII*, Wien, Karl Gerold's Sohn.

Pott, A., 1844-1845, *Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache*, Halle, Heynemann.

Sampson, J., 1926 [reprint. 1968], *The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood*, Oxford, Clarendon Press.

Turner, R. L., 1926, *The position of Romani in Indo-Aryan*. «Journal of the Gypsy Lore Society» third series, 5: 145-189.

YARON MATRAS

THE CHALLENGES OF LANGUAGE CODIFICATION
IN A TRANSNATIONAL CONTEXT