

## **Social networks as centres of language codification: Romani on YouTube**

**Daniele Viktor Leggio and Yaron Matras**

*University of Manchester*

### **1. Minority languages on the Internet**

Following its worldwide adoption during the 1990s, the Internet opened up a new domain of language use. The first studies concerned with this domain expansion focused on English-based websites and various computer-mediated communication (CMC) modes. These studies showed that while text in static pages generally conformed to the norms of the standardized language, text occurring in multi-user, interactive CMC modes showed a high number of features so far deemed typical of oral, non-standard communication. These included the representation of dialectal features through creative spelling. Creativity also allowed users to cope with the limitations imposed by the textual nature of the Internet in conveying extra-linguistic cues. For example, written-out sounds (i.e.: *hahaha* for a laugh) or the usage of capitalization to represent shouting replaced supra-segmental features, while emoticons, combinations of punctuation symbols ---:) for a smile ;) for a wink, :( for sadness -- helped convey facial expressions and the writer's mood (Herring 2002).

The focus on English-based CMC was due to the early dominance of English on the Internet, which in turn derives from the medium's origin in the United States. Text transmission was originally based on the ASCII (American Standard Code for Information Interchange) character set which initially only supported the Roman characters used in English and thus discouraged usage of other languages, especially when their alphabets were significantly distinct from English. However, the use of languages other than English to on the Internet has since increased. This has prompted interest within sociolinguistics in what has been named the 'multilingual Internet'.

The papers collected in Wright (2004c) investigated the position of local languages in respect to English. Quantitative data were drawn from university and high school students in Tanzania, Indonesia, the United Arab Emirates, Oman, France, Italy, Poland, Macedonia, Japan and Ukraine. Their analysis showed a good deal of convergence toward English, an increasing presence of prestigious standardized languages and difficulties in the usage of smaller, non-standardized ones. These results, although limited to the specific populations studied, led the authors to conclude that the Internet was "not providing a space for diversity in language practice" (Wright 2004a: 12). At the same time, while noting the emergence of

innovative and spontaneous spelling practices for Arabic and Japanese, the contributions suggested that standardization would be needed for smaller, non-standardized languages in order for them to be effectively used on the Internet.

Various other studies (see for example those collected in Danet & Herring 2007) confirmed the role of English as the Internet lingua franca but showed that the medium offers spaces for vernacular language use. In particular it has been shown that low varieties of languages in diglossic situations, such as dialectal Greek (Tseliga 2007, Themistocleous 2010) and dialectal Arabic (Palfreyman & Al Khalil 2007), were favoured by users over the corresponding high varieties. Like their English-speaking counterparts, Greek and Arabic users found creative and playful ways to overcome the constraints imposed by the Internet on non-Latin scripts. In both cases users favoured innovative visual transcription patterns in which Roman letters and numbers are employed on the basis of their visual similarity with Greek or Arabic characters, rather than resorting to phonetic transcription norms. Such non-standard spellings remain in use despite the fact that more advanced text-transmission codes, such as UNICODE, now cater for non-Latin alphabets (Danet & Herring 2007b). In most cases, users attribute the choice in favour of non-standard spelling to the emerging, positive in-network values attached to these innovations (Lee 2007, Palfreyman & Al Khalil 2007, Tseliga 2007).

Even the second conclusion put forward in Wright (2004c), concerning the need to standardize smaller languages in order to facilitate their use on the Internet, has been challenged. Research on Jamaican Creole and Nigerian Pidgin (Deuber & Hinrichs 2007) as well as Mauritian Creole (Rajah-Carrim 2009) has shown how CMC is promoting the development of new forms of literacy for oral, non-standardized languages. In all these cases the Internet has become the arena for spontaneous, creative and decentralised processes of language codification. In such processes ‘users can converge on novel codes [...] without ever having to refer to a standard’ (Rajah-Carrim 2009: 504). As a result shared writing norms are spontaneously emerging, often leading to spellings increasingly differing from those proposed by language planning experts (Deuber & Hinrichs 2007).

This relaxation of norms concerning on-line spelling has an interesting counterpart in on-line language selection. While in traditional cases of domain expansion adherence to a single language was required and often linked to the representation of national identities (Kaplan & Baldauf 1997, Wright 2004b) this limitation is often disregarded on the Internet. Minority language speakers often switch into other languages when writing on the Internet. In particular, when English or another major language are dominant, switches to minority

languages tend to occur as displays of ethnic identity in fixed expressions (salutations and similar formulae), poetry and song lyrics. More creative usages of code-switch also occur, although less frequently, in jokes and attention-seekers such as insults and appeals (Paolillo 1996, Sperlich 2005, McClure 2001). Furthermore, Androutsopoulos (2006) notes how, on websites catering for immigrants in Germany, users switch between their home languages, German and even English. However, languages do not stand in a straightforward opposition, with home languages as symbols of ethnic authenticity and German and English signalling an uncritical adoption of German or of a global, homogenized culture respectively. Rather, switching between English, German and home languages indexes multiple cultural affiliations, each language used as an icon of individual identities, including but not limited to ethnicity.

Following Milroy & Milroy's (1992), Paolillo (2001) looked at the distribution of such vernacular language features on the #india Instant Relay Chat channel from a social network perspective. His finding showed that language variation on CMC shows features that are both similar to and different from face-to-face language variation. More precisely, he noted how, just like in face-to-face communication, 'standard' forms spread through weak ties in the users' networks he identified while 'vernacular' forms characterize the core of such networks where strong ties are prevalent. More precisely, he noted how code-switching into Hindi from English (the dominant language on the channel) and the usage of profanity appeared like vernacular language features: core users of the channels who shared a high number of strong ties resorted to them more frequently. On the other hand, innovative English spellings like *u* for you and *r* for are, appeared like standard language features: they were employed more frequently by peripheral members of the user networks who shared a high number of weak ties. One must in this connection consider the preponderance of weak ties in on-line networks. This in turn can be linked to certain properties of the medium: users' frequent mobility between different communicative channels, and the light-hearted tone of conversations favouring short-term, fleeting relationships. Considering, in addition, the absence of clearly identifiable power centres on the Internet, Paolillo suggested that with reference to the Internet domain, a 'standard' has to be defined in relation to frequency rather than prestige.

## **2. The case of Romani**

With upwards of 3-4 million speakers, Romani is one of the largest, if not indeed the largest minority language in the European Union. It is certainly the most dispersed language

in Europe, with communities of speakers scattered in virtually all European countries and beyond (in the Americas and Australia). Due to a history of social isolation and deprivation, poverty, low literacy rates, and very limited access to professions and positions of influence in public life, self-representation and consequently also public self-depiction of Romanies has been lacking until recently. Nonetheless, documentation of the Romani language by outsiders has a long and established tradition going back to isolated sources in the sixteenth century and intensified documentation and discussion of the language's structures and origin in the mid- to late eighteenth century. The discovery of the Indian origin of Romani is often wrongly attributed to Grellmann (1786). In fact it was Rüdiger who in a public lecture in 1777 first put forward the case for the linguistic affiliation of Romani to the modern Indo-Aryan languages. Acting on suggestions from colleagues and peers, Rüdiger carried out his own first-hand research on the language, working with a native speaker informant. In 1782 published the first systematic overview of Romani grammatical structures and a comparative discussion contrasting the language with "Hindustani" (Hindi/Urdu). Since then it has been accepted in mainstream scholarship that the Indian origin of the Romani language testifies to the origin of the Romani people as descendants of an immigrant population from India which arrived in Europe in medieval times.

Despite the absence of early attestation of the Romani language, comparative investigations of Romani dialects have in recent years enabled a reconstruction of the development paths from a common predecessor language, sometimes referred to as Early Romani, to the present-day varieties of the language (Matras 2002, 2010 Ch. 2, and Elšik & Matras 2006). Yet many popular sources continue to confuse Romani with a rather vague entity referred to as 'Gypsy languages'. There is no precise definition of the latter and it is usually understood as a wholesale reference to the diverse speech forms of peripatetic populations – 'Travellers' or 'commercial nomads' – who constitute tight-knit, endogamous communities that are typically engaged in mobile service economies. These populations (e.g. the Irish Travellers, the *gens du voyage* in France, the *Woonwagenbewoners* in the Netherlands, and the *Jenisch* and other numerous itinerant communities in Germany) do not have any particular language in common, and are indeed of diverse ethnic and geographical origins. However, they almost invariably maintain group-internal vocabularies, often referred to as 'secret languages' or 'special languages' (in German *Sondersprachen*), which are known in more popular terms as 'argot', 'jargon', 'Cant' and other labels. These vocabularies are useful in concealing key messages between group members from outsiders, but they also serve important identity-flagging functions.

While there is a historical interface between Romani and many such special vocabularies, in that Romani is often a source of lexical vocabulary for in-group jargons of this kind, Romani itself is a clearly defined, distinct linguistic entity. It is, for a start, a full-fledged language with its own grammatical inflections, many of them representing the most archaic stage of any modern Indo-Aryan language. Moreover, while the ancestors of the Roma are believed to have been peripatetics, affiliated with the *dom*-castes of India (the etymological cognate of the self-appellation *rom*), and while some Romani communities continue to engage in mobile service economies, most Romani populations of Europe have been settled for many centuries now and many show a rather diverse profile of economic activities. Romani is therefore much closer in its sociolinguistic typology (as the everyday language of the home) to the idioms of other settled, minority communities than to the in-group vocabularies used by itinerant communities for secret communication and identity flagging.

Structurally, Romani can be characterized as a combination of late Middle Indo-Aryan and early New Indo-Aryan morphology with Balkan-type syntactic structures (see Matras 2002). Romani populations migrated from the Balkans into all parts of Europe from the late fourteenth century onwards. They settled in the various regions during the fifteenth and early sixteenth centuries. It is believed that during this time their speech forms were rather homogeneous, reflecting an ancestral 'Early Romani' that formed in Anatolia and the Balkans during the Byzantine period. Geographical dispersion and the influences of separate social, cultural and linguistic contact settings contributed to the emergence of separate innovations sometime between the sixteenth and early eighteenth centuries. These particularities now define present-day dialectal differences in Romani. The processes involved are typically phonological changes, morphological levelling and internal grammaticalisation, as well as borrowings from the diverse contact languages especially in vocabulary but also in various areas of grammatical structure. Especially prone to contact influences are function words, word order patterns, phonetics and phonology. By and large, Romani dialects remain mutually intelligible. However, since all adult Roma are bilingual, some degree of code-switching is a permanent feature of family-internal conversation, and so code-switches into the respective majority languages may impede mutual intelligibility.

Generally speaking, Romani dialects constitute a language-geographical continuum, with structural differences accumulating with greater distance. A zone in central Europe, roughly following the old frontier zone that separated the Habsburg monarchy from the Ottoman Empire in the 16th-17th century, shows a dense bundle of isoglosses. The old political frontier hindered contacts and so also the diffusion of structural innovations during the

crucial period that followed settlement. As a result, this dialectal ‘Great Divide’ (see Matras 2005), which separates north-western from south-eastern dialects, limits mutual intelligibility between, for example, the Romani dialects of France, Germany, and Austria, on the one hand, and those of Poland, Hungary, and Slovenia, on the other.

Studies of domain usages of Romani and the degree of institutional support that the language receives tend to focus on individual countries or regions, and there are very few overview works that deal with Romani as a whole (but see Matras 2013a). Yet Romani offers a fascinating case study: it is a dispersed language that has been retained by its speakers for many centuries in what is and in all likelihood has always been a permanent situation of bilingualism, in the absence of a native territory, literacy, and any cultural institutions. Like the rise of literacy and codification in many languages, initiatives to promote the use of Romani in public and institutional contexts and in writing began in the form of activism by private individuals. The forging of networks of activists was boosted in the early 1990s following the political transition in central and eastern Europe, where the majority of Romani speakers live. But in the absence of any form of sovereignty and any large-scale resources, the process of language ‘institutionalisation’ in Romani remains dependant on outside intervention and support. In its specific social and geo-political context it also hinges on the prospects of trans-national cooperation and on the exploitation of opportunities offered by new communication technologies. The case of Romani in Europe today thus provides an opportunity to observe a rather unique interplay of a particular assembly of agents who are promoting change in status and usage domains, with a specific set of tools that are being employed to facilitate the process. So far the product of such efforts appears to be an organic network of initiatives that accept and in some ways actively promote linguistic pluralism as an alternative to a centralised approach to standardization and language planning. The status of Romani is thus being negotiated in a context that might be described as diasporic, trans-national, poly-centric or cosmopolitan, providing insights into what might possibly become a widespread alternative to a century-old concept of public and written language as a unified emblem of nationhood.

### **3. YouTube as a source of data**

YouTube, originally created in February 2005, has quickly become the most popular video sharing website available on the Internet. The website offers a series of networking opportunities for those users that create their accounts, or channels as they are called on YouTube. Besides the obvious possibility to upload one’s own videos, the networking

possibilities offered by YouTube include subscription to other users' channels, private messaging between registered users, access to the 'like' and 'dislike' buttons associated with each video and to the comment sections on both videos and channels. At the same time, the videos posted on YouTube are also accessible to the wider public who can watch them without registering. Because of these different access possibilities, the ways in which people use YouTube vary considerably.

Gallardo-Camacho & Jorge-Alonso (2010), for example, note how the majority of Internet users accessing YouTube are not taking advantage of the networking possibilities offered. Rather, they treat the website as a TV channel to be passively watched. Rotman & Preece (2010), on the other hand observe how users taking advantage of the networking possibilities offered by YouTube come to regard it as an on-line community. However, as noted by Shoham, Arora & Al-Busaidi (2013), Rotman & Preece's conclusions are influenced by the fact that they investigated video-blogs tagged with the terms 'YouTube,' 'community,' and 'on-line.'

In order to offer a more nuanced picture of users' approach to YouTube content, Shoham, Arora & Al-Busaidi investigated the interactions occurring around a single video. They noted how, out of around 84,000 viewers, only 225 liked or disliked the video, while 106 left a comment to the video. Of those commenting on the video, they noted only 24 engaged in dyadic or group conversations and posted more than one comment. The remaining commenters each posted a single message directed to a generalized other. Based on these observations, they suggest that the vast majority of YouTube users can be characterized as *passive*, simply watching videos. Of the remaining users, a tiny minority can be characterized as *interactive*, taking full advantage of the networking possibilities offered to engage in interactions with other users. The rest of YouTube users can be regarded as *active*, simply liking or disliking videos or broadcasting their comments to a generalized audience. Finally, echoing early findings by Thelwall, Sud & Vis (2012), Shoham, Arora & Al-Busaidi also notes that comments to YouTube videos are generally very short, often made up of just one sentence, and display many of the features encountered in other CMC environments.

Regarding language choice on YouTube, no reliable data is so far available. This is due, first and foremost, to the fact that only few studies looked at the content of either videos or comments on YouTube. Vergani & Zuev (2011) investigated the usage of YouTube by Uyghur activists and focus mostly on the content of the videos, in which the Uyghur language is used more frequently than any other language. They only marginally mention the problems encountered by Uyghur speakers in using their language, a Turkic language traditionally

written in Arabic script, in the comment section. Thelwall, Sud & Vis (2012) and Shoham, Arora & Al-Busaidi (2013) investigate the content of comments, but rely largely on English samples.

Whether users are active or interactive, their activities lead to the emergence of two different kinds of networks on YouTube: user networks and video networks. User networks are comprised of individuals' profiles that are connected through subscription to other users' profiles. They represent personal affinity and depict pre-existing relationships between individuals. Video networks consist of videos that are connected through a shared topic, identifiable through the tags, titles, descriptions and categories that users assign to the videos that they upload. They are thus connected through the comments that users post or through the fact that users reply to a video by linking it to a further video. Video networks therefore represent shared topicality and subjects of interaction.

The networking opportunities offered by YouTube and the different kind of networks they create provide for a unique opportunity to observe bottom-up processes of language codification. In traditional, static websites, the written language reflects the practices of the small group of people, if not of the single individual, who are responsible for the creation of a given site. By contrast, social networking sites are open to the contributions of thousands of users. Moreover, social networking sites provide a much more open and loosely structured frame of interaction compared to other multi-user CMC modes such as forums and newsgroups, which are often theme-based and moderated. For this reason, the popularity and openness of social networking sites can offer insights into the writing practices of a much wider and less regimented user community. Among all social networking sites YouTube is also the most easily accessible to researchers since all its user generated content is publicly available. In this paper we investigate the emergence of written forms of Romani among video networks on YouTube. We examine the choice of dialectal and spelling variants in users' interactions as a way of documenting the role of new communication technologies and the socio-cultural ties among actors involved in the emergence of a written language.

#### **4. Mapping user networks on YouTube**

In order to identify and investigate Romani networks on YouTube we used NodeXL. This open-source template for Microsoft Excel provides a series of automated tools that allow the mapping, analyzing and visualizing of social media networks. The template includes a tool to import both kinds of YouTube networks, user networks and video networks (Hansen, Shneiderman & Smith 2011).

NodeXL maps video networks by searching tags, titles, descriptions and user names that accompany videos for a particular keyword or expression. Based on preliminary observations we established that YouTube videos posted by Roma tended to fall into three main categories: weddings, religious (largely Pentecostal) videos, and live or studio performances by singers. In order to cover a range of countries of origin and different communities, we searched for videos containing the Romani word for ‘wedding’ in different dialectal pronunciations and spellings: *abav*, *abijav*, *abiav*, *abjav*, *bijav*, *biav*, *bjav*. To identify music videos we identified and searched for a series of names of singers: *Ševčet* (spelled 'sevcet' as commonly done by Roma from former-Yugoslavia`, see Leggio 2013) from Macedonia; *Francuzo* who originates from Poland, *Gypsy Kubanec* from the Czech Republic, and *Prinzo* from Germany. For Pentecostalism related videos, we searched for the keyword *khangeri* ‘church’ and for videos relating to *Ricardo Kwiek*, a Polish Romani pastor based in Germany. Finally, we had noticed a series of videos relating specifically to the Lovara Roma, a group whose origins are in Transylvania but whose networks have spread all across Europe and the Americas since the late nineteenth century, and so we also ran a search for the word *lovara*.

<b>Video Network</b>	<b>Keyword searched for</b>	<b>Videos</b>	<b>Edges<sup>1</sup></b>	<b>Total comments</b>
G1	<i>abiav</i>	10	15	68
G2	<i>lovara</i>	10	11	205
G3	<i>ricardo kwiek</i>	9	17	690
G4	<i>Sevcet</i>	9	19	769
G5	<i>Gypsy Kubanec</i>	8	22	78
G6	<i>Prinzo</i>	6	13	330
G7	<i>Abiav</i>	6	6	57
G8	<i>Abijav</i>	5	7	31
G9	<i>khangeri</i>	4	5	524
G10	<i>Bijav</i>	4	4	191
G11	<i>Bijav</i>	4	4	28
G12	<i>Francuzo</i>	3	2	37

**Table 1:** List of selected video networks.

Once the videos are identified, NodeXL maps the relationships between them. There are three available options: shared tag, shared commenter, and shared video reply. Our interest was in the relations among users who comment on videos in Romani, and we therefore chose to map shared commenters. We limited our searches to fifty videos per keyword, and then selected the networks of related videos for analysis. We selected networks that included between three and ten videos that had received at least twenty-five comments in total.

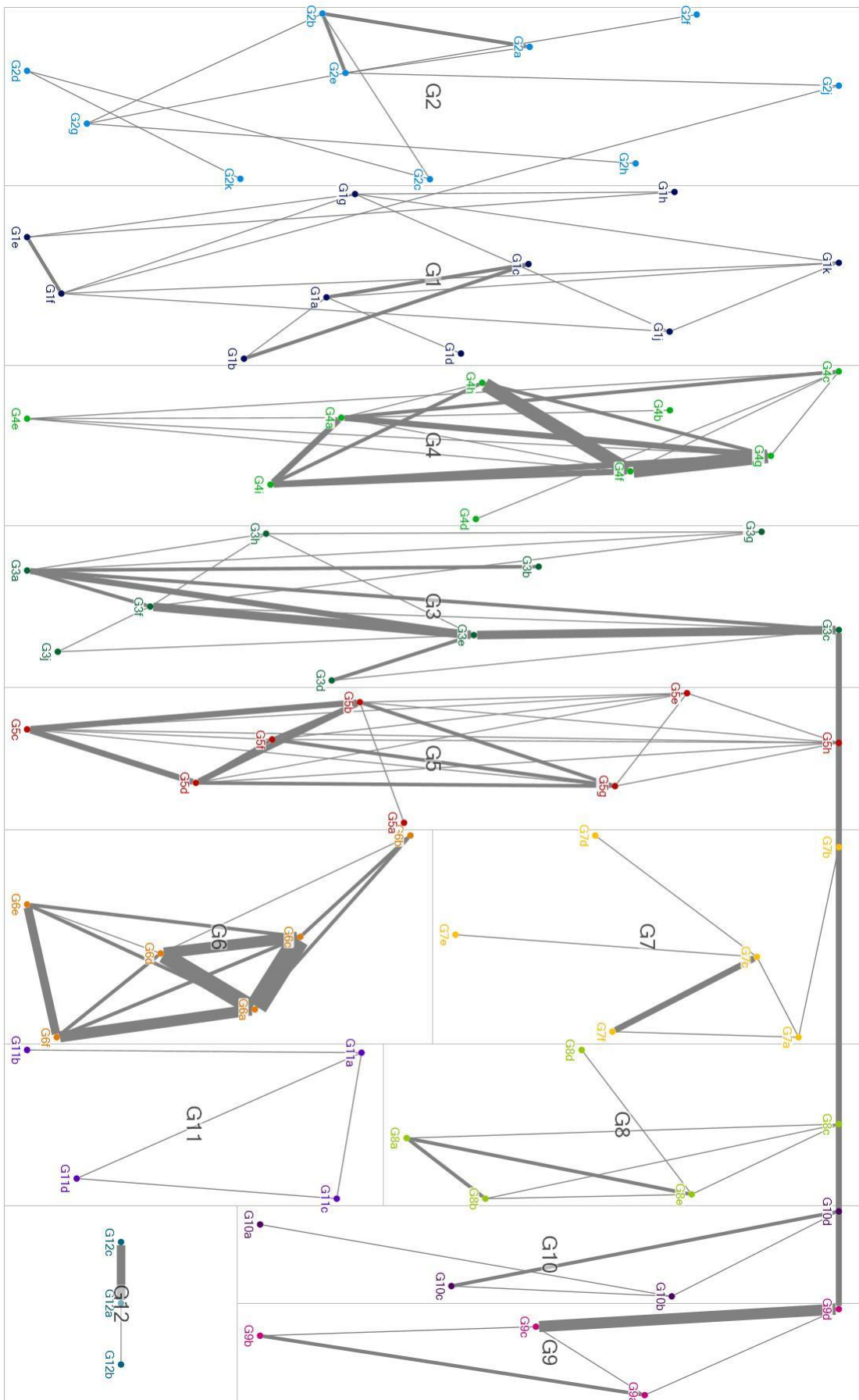
<sup>1</sup> The edges in each network represent users commenting on two videos, appearing as nodes (see Figure 1). The thinner edges represent a single user commenting on two videos, thicker edges represent a higher number of users.

These criteria were established in order to obtain a manageable corpus that could be analysed manually while at the same time containing enough data to allow for comparison between networks of different sizes and structures. The decision to analyse the corpus manually took into account the absence of a standardized spelling for Romani, which makes the use of corpus analysis software impractical.

The networks selected for analysis are listed in Table 1. Note that the searches for *abav*, *abjav*, *biav* and *bjav* did not return any network of videos matching our criteria for selection, while the searches for *bijav* and *abiav* each returned two networks of videos.

As each video network was found using different keywords, we were not expecting to see connections among them at this stage. However one of the *abiav* videos (G1f) also contained the word *lovara* in its title and one of the users commenting on it also commented on video G2j, thus creating an edge between the two networks. Similarly, one of the *Ricardo Kwiek* videos (G3c) also contained the word *khangeri*, and three of the users commenting on it also commented on video G9d, creating a thick edge between the two networks (see Figure 1).

It should be noted that NodeXL cannot check the language used by commenters. Considering the widespread bilingualism among Roma, the documented tendency to code-switch in CMC (see above) and the possibility that non-Roma were commenting on the videos, we expected that some of the edges represented users commenting exclusively in a language other than Romani. Similarly, we expected some edges to be missing as some users might be commenting in Romani on videos without common keywords, and were thus not identified by NodeXL searches. In order to obtain a more accurate picture of interaction in Romani, it was therefore necessary to delete edges representing commenters who did not use Romani, and to add edges representing users commenting on pairs of videos without common keywords. This was done as part of the analysis of comments. The comments for each network were copied from their YouTube pages and pasted into text documents. Users who did not employ Romani in any of their comments were identified and their contributions were excluded from the analysis, and any edge to which they contributed was removed from the network.



Created with NodeXL (<http://nodexl.codeplex.com>)

**Figure 1:** Selected video networks as identified by NodeXL. Videos represented as nodes. Shared commenters represented as edges: the thicker the edge the more commenters it represents.

## 5. Choice of variables

While checking comments for the use of Romani, we observed that some users only employed isolated Romani words or noun phrases while others posted entire sentences, which included inflected verbs. In order to capture possible variation among users and user groups in the extent to which Romani structures were employed, we divided the relevant contributions into three categories: comments containing isolated or individual Romani tokens (isolated words like *lacho* ‘good’ or noun phrases like *schukka gili* ‘nice song’), comments containing Romani sentences including inflected verbs, and comments without any Romani item. In order to obtain a picture of dialectal variation in the corpus, we compiled a number of key morphological and phonological variants which represent salient isoglosses in Romani (Matras 2005, 2002 ch. 9). In discussions of Romani dialect classification, some of these features are often considered to be ‘diagnostic’ of tentative dialect ‘branches’ (see discussion in Matras 2005, Elšik & Matras 2006, and for a critical discussion Matras 2013b).

While certain clusters of features are typical of particular regions, Romani dialects form, by and large, a geographical continuum and so a clear-cut mapping of branches or dialect groups based on a closed set of features is not possible. Nonetheless, the known geographical distribution of individual feature makes it possible to associate variants that appear in the online corpus with known varieties of Romani that are spoken in particular regions. The picture is complicated somewhat by the fact that some Romani communities constitute ‘diasporas’ in the modern sense, that is, they originate in a particular region and location, but have since dispersed into other regions and locations, where they continue to maintain a distinct identity profile that often includes in particular a distinct dialect. Some of these diasporas, such as the communities of the Lovara, Kelderasha, Churara and other related groups (usually referred to as ‘Vlax’ in association with historical Wallachia), dispersed more than a century ago from Banat and adjoining regions in western Romania. Others, such as the Romani community of Mitrovica in Kosovo, represent more recent migrations, in this case on connection with the Kosovo conflict from the mid-1990s onwards.

The fact that we can associate particular structural features with known dialectal varieties of Romani allows us to recognise, tentatively at least, the origins of users in certain Romani population groups and sometimes regions, based on the dialectal variants that these users employ in their communication. A high degree of consistency of variants within a distinct network of users or videos can point to the coherence of the dialect employed by these users. This can suggest membership in a coherent Romani community, whether regional or diasporic, or alternatively it might suggest a motivation to accommodate to a particular

dialectal variety. Conversely, noticeable variability of features within the communications of a single user network can serve as an indication of the diverse dialect backgrounds of the users as well as the absence of any motivation to accommodate to a single norm. The mapping of dialect features onto user networks therefore serves as our key indicator for the considerations that users entertain in a process of language codification that is spontaneous, i.e. not regulated by institutions or institutionalised norms.

It is important to remember that Romani YouTube users share the documented tendency to broadcast short messages (see above). This, combined with the general light tone of the messages, results in a style that is closely reminiscent of oral communication (a characteristic common to various CMC modes). For this reason, tokens for some of the features investigated occurred rarely in the sample. In particular, perfective verbs or verbs inflected for plural persons (both present and perfective) were extremely rare. Similarly, very little comparative or superlative adjectives occurred in the sample. This resulted in extremely low token counts for certain salient features like the usage of Romanian-derived comparative/superlative marker *maj* and the inherited *-eder* or other forms, the levelling of forms in the past-tense person inflection paradigm or the usage of relative pronouns and complementizers. Features with an extremely low token count, such as the above, were therefore excluded from consideration.

Attention was given to the following set of salient structural features, for which sufficient documentation was usually found within the comments to enable us to identify the user's chosen regional or group-specific variety of Romani (see Table 2): The abstract nominal marker takes the consonantal form *-ipen* (*sastipen* 'health') north of the Great Divide and the reduced form *-ipe* (*sastipe* 'health') to the south. An alternative variant (a Greek-derived marker) *-imos* (*sastimos* 'health') is widespread predominantly in the region of Romania, Moldova and Ukraine, as well as in the out-migrant so-called Vlach dialects originating from that region. The consonantal prothesis in *j-* (*javel* 's/he arrives', *joj* 'she') predominates north of the Great Divide and is absent to the south (*avel* 's/he arrives', *oj/voj* 'she'). Vowel prothesis in *a-* (*abjav* 'wedding', *ašun-* 'to hear') is generally found in the dialects of Romania, Moldova and Ukraine, as well as out-migrant Vlach dialects.

The past-tense 2SG person marker is *-al* (*tu kerdal* 'you did') in the Sinti-Manouche dialects of western Europe (Germany, France, Belgium, the Netherlands, Austria, northern Italy), in Finnish Romani, and in the so-called Central dialects of Slovakia, southern Poland, the Czech Republic, northern Hungary and parts of Romanian Transylvania, Slovenia and southeastern Austria.

The copula root *h-* is found for all persons (*hom* ‘I am’, *hal* ‘you are’, etc.) in the Sinti-Manouche dialects as well as in some isolated dialects spoken in Macedonia, Hungary, and Romanian Transylvania, while copula forms in *s-* (*som/sem* ‘I am’, *san/sal* ‘you are’, etc.) are found elsewhere. For the third person, forms in *h-* also appear in the central European and Finnish Romani dialects (often matching the distribution of the past-tense 2SG person marker *-al*) and occasionally, often as optional variants, in some of the out-migrant Vlax dialects of southeastern Europe. Copula stem intrusions in *-ij-* and *-in-* (*sijom/sinom* ‘I am’) are common in individual dialects of the southern Balkans and southern Italy, and for the third person also in some of the dialects of central Europe and Finland. Subjunctive forms of the copula verb derive from *ov-* (originally ‘to become’) in the dialects of southeastern Europe as well as in the so-called Romungro dialects of Hungary and Slovakia, while forms deriving from *av-* (originally ‘to arrive’) are common elsewhere, including the Vlax out-migrant dialects in southeastern Europe.

Finally, the distribution of individual demonstrative forms is complex due to the presence of a four-term system in most Romani dialects and to the high diversity of forms. Demonstratives in the corpus therefore help identify regional and sometimes local varieties of the language.

Feature	Variants	Distribution
Abstract nominal marker	<i>-ipen</i> <i>-ipe</i> <i>-imos</i>	north of Great Divide south of Great Divide Romania, Moldova and Ukraine
Consonant Prothesis	<i>j-</i> None	north of Great Divide south of Great Divide
Vowel Prothesis	<i>a-</i> None	Romania, Moldova and Ukraine, out-migrant Vlax other
2SG.PERF marker	<i>-al</i> <i>-an</i>	western and central Europe, Finland other
Copula root	<i>h-</i>  <i>s-</i>	Sinti-Manouche, some dialects of Transylvania, Macedonia, (all persons); central Europe, Finland, Balkans (third person) other
Intrusion in copula	<i>-ij-/in-</i> None	southern Europe elsewhere
Subjunctive copula	<i>ov-</i> <i>av-</i>	Balkans, Romungro elsewhere; Vlax
Demonstrative pronouns	<i>kada, kava, kado,</i> <i>akava, ada,</i> <i>kadava, etc.</i>	Dialect specific

**Table 2:** Dialectal features investigated

Examples of demonstratives systems are *akava/okova/adava/odova* as well as *adavka/odovka* in the southern Balkans, *kava/kova/kadava/kodova* in the so-called southern Vlax or out-migrant Vlax dialects of southeastern Europe, *kado/kodo/kako/kuko* in the dialects of the Banat and Transylvania region and out-migrant Vlax dialects in central and northern Europe, *dava/adava/dova/odova* in the dialects of Poland, Russia, and the Baltics regions, *ada/oda/kada/koda* in the Central dialects of southern Poland, Slovakia, northern Hungary, Slovenia, and northwestern Romania (Transylvania), and *kava/kova* in the Sinti-Manouche varieties.

Nonetheless, for each video network we were able to identify a number of features which token count generally allows for establishing which variety is dominant. The features identified and the diffusion spaces they characterize are summarised in Table 2.

Concerning the spelling used, as nearly all comments were written in the Latin alphabet, we focused our attention on those sounds for which Latin-based alphabets show variation and for which we can therefore expect to see variation in the choice of spontaneous spelling used by commenters for Romani: These include the postalveolar affricates /č/ and /dž/, the postalveolar fricatives /š/ and /ž/, the velar fricative /x/, the semi-vowel /j/, the labial fricative /v/, the centralized vowels /i/ and /ə/ and palatalized consonants /lʲ/, /rʲ/, /kʲ/, /gʲ/. We also looked at those sounds that are distinctive of Romani, and which have often been the subject of intense discussions among language activists: the aspirates /ph/, /th/, /čh/ and /kh/ and the sound /ř/, which is realised in different dialects as a prolonged trill, as a uvular trill, or in some as a retroflex. Attention to the Graphemic representation of these phonological features allows us to establish the degree to which YouTube commenters tend to follow the norms of their respective contact languages or apply creative solutions.

For each user, we manually recorded the number of occurrences of individual structural features and of spelling solutions. The majority of users only posted one or two short comments, in which dialectal features and ‘diagnostic’ spelling solutions (representing ‘problematic’ sounds) were usually represented by an average of three tokens each. It was thus difficult to assess whether individual users were always consistent in their choices, as the impression remained limited to a small sample for each variable. However, the assembly of relevant tokens for each variable across the entire group of users within a given network provided an impression of the degree of coherence within each network and allowed us to carry out comparisons between the networks.

## 5. Network structure and characterization

The network structure illustrated above (Figure 1) was modified through two procedures: First, by deleting edges that represented users who did not employ Romani in their comments. And second, by manually adding edges that had not been identified by NodeXL, and which represented users' comments in Romani that did not contain the keywords employed in the automatic search. The result was a modified network structure, which is shown in Figure 2.

Network G1 (identified by searching for *abiav*), for which a connection had already been established with G2 (see above), was now found to share commenters also in G3, G4 and G10: video G1a is connected to video G4g; G1f, already linked to G2j, is also connected to G10b; G1h also shares a commenter with Gb10; finally G1j is connected to G3j. It is worth noting that all videos in G1 contain the English word 'Gypsy' in their titles or descriptions. This shows that the users who uploaded the videos are not targeting just Romani speakers as their audience. Various edges that represented a high number of shared commenters in Figure 1 were reduced to thinner edges in Figure 2, since they represented commenters who did not use Romani. However, the usage of English 'Gypsy' will have also attracted Romani speakers of different backgrounds. This will explain the relatively high number of connections created by comments that contained some Romani, among different video networks. For this reason, we expected to find a high degree of variation both in Romani features and in Romani spelling in the comments to G1 videos.

G2 (identified by searching for the keyword *lovara*) remained mainly isolated but for G2j, the most peripheral video in the network, which beside the connection with G1f also showed connections with G3f, G6c, G7b and G12c. Given this network's group-specific theme and the fact that the Lovara constitute a tight social network of diasporic communities that are dispersed across numerous urban communities throughout Europe, we expected to find a dominance of characteristic structural features of the Lovara dialect (one of the so-called Northern Vlax dialects, originally from the Banat-Transylvania regions, and closely related to the dialects of the Kelderasha and Churara groups). Given the number of connections with other video networks and the considerable geographical dispersion of the Lovara (and therefore acquisition of literacy in a variety of different contact languages), we also expected considerable variation in orthographic representation.



G4 (identified by searching for the keyword *sevčet*) merged completely with G10 (based on the search for *bijav*). Various edges were lost or thinned down as various users only employed German in their comments to these videos. The frequent usage of German, coupled with the fact that Ševčet is a Macedonian singer living in Germany, suggests that users within this network might be part of the widespread Balkan Romani diaspora described by Silverman (2012). This is partially confirmed by the fact that *bijav* is a spelling commonly employed in Macedonia, representing the pronunciation of the word in Southern Balkan dialects of Romani. On this basis we expected both the dialectal features and the spelling conventions used in these videos to reflect, broadly at least, the Macedonian model of Romani standardization (see Friedman 1995, 1996, 1997, 2005), along with influences from German spelling. Due to the connections with G1 (G4g is connected to G1a, G10b is connected to G1f and G1h), we also expected that some features of other spelling systems as well as dialect features from outside the Southern Balkan region might occur in the comments to these videos.

G7 (identified by searching for *abiav*) and G12 (compiled by searching for *francuzo*) also merged completely. Francuzo is a Polish Romani singer, and some users wrote their comments exclusively in Polish. This suggests that G7 videos attract a Polish Romani audience. We therefore expected the comments to show dialectal features that are characteristic of the Romani dialects of the so-called ‘North-eastern’ group comprising the Romani varieties of central and northern Poland, the Baltics and northern Russia, or of the ‘Central’ dialects of southern Poland, or else of Vlax Romani communities based in Poland.

A more complex merger between networks identified through different searches occurred between G3, G5 and G9. G3 (compiled by searching for the keyword *ricardo kwiek*) and G9 (found by searching for *khangeri*) both include videos that revolved around Romani Pentecostalism. Some edges were lost or thinned down when excluding non-Romani content, while on the other hand a new connection between G3j and G9d was also found, reinforcing the existing link between these two networks (see above).

In G5 (identified by searching for *gypsy kubanec*), we found that only a single user inserted Romani words into messages in Czech while all other commenters wrote exclusively in Czech. This resulted in videos G5a, G5d and G5g being excluded from the sample as they contained no Romani in their comments. Although the remaining videos in G5 did not relate to Pentecostalism, those users who commented on them in Romani also employed some Romani in their comments to videos G9c, G9d and G3j. The addition of these edges led to the merger between G5 and G9, in which G3j was also grouped. This confirmed the impression

that the connection between G3 and G9 is a strong one. Given this strong connection, clearly due to the shared topic between G3 and the now expanded G9, we decided to treat these two networks as a single network. Considering the potential of the religious topic to attract Roma from various backgrounds, we expected to find the highest variation in dialectal features and spelling in these networks of videos.

The remaining three networks, G6 (found searching for *prinzo*), G8 (found searching for *abijav*) and G11 (found searching for *bijav*) remained mostly isolated except for the connection between G6c and G2j. A few edges in G6 and G11 were also thinned down, but this did not alter the networks' structures. Given the isolation of G6 and the fact that it includes videos by a Sinti singer, we expected to find the dialectal features of the Sinti-Manouche varieties and predominantly German spelling influences (reflecting the presence of the majority of the speakers of this dialect in Germany). The videos in G8 and G11 all depicted weddings, which took place in various European cities. Many featured artists who enjoy popularity among the Balkan Roma diaspora, such as *Ćita*, *Mladi Talenti* and *Džansever* (see Silverman 2012 Ch. 2). For this reason we expected both networks to show similar spellings based on a combination of norms including Serbian/Croatian, Italian, French and German (as documented in Leggio 2013). As for dialectal features, the usage of *abijav* in G8 videos suggested that users may be of Vlax background, while *bijav* in G11 pointed to a Southern Balkan background. The adjusted video networks are summarised in Table 3.

Video Network	Keyword searched for	Videos	Edges	Comments containing Romani
G1	<i>abiav</i>	10	15	40
G2	<i>lovara</i>	9	9	93
G3+G5+G9	<i>ricardo kwiek+gypsy</i> <i>kubanec+khangeri</i>	18	44	250
G4+G10	<i>sevcet+bijav</i>	12	24	251
G6	<i>prinzo</i>	6	13	183
G7+G12	<i>abiav+francuzo</i>	10	11	62
G8	<i>abijav</i>	5	7	30
G11	<i>bijav</i>	4	4	15

**Table 3:** List of adjusted video networks

It is interesting to note how all the video networks with the exception of G6 lost a considerable degree of density once non-Romani comments were filtered out. This shows how the networks attract an audience that include both speaker of Romani and others, or at least users who write in Romani as well as those who prefer to write in another language. It is noteworthy that the number of comments containing some Romani (Table 3) is generally half

the number of the total comments initially found in each video network (compare with Table 1) This confirms that YouTube, like other CMC environments, offers a space for multilingual, vernacular language practices. In the case of our sample, the other languages employed alongside Romani were Serbian/Croatian, Macedonian, Albanian, Polish, Czech, Slovak, Russian, Hungarian, Romanian, German, French, Italian, English and Spanish.

The other noticeable characteristic of the networks is that most of them seem to attract users originating from the same region; this could be established by assessing the distribution of dialect features (see discussion below). However the fact that there were also connections between videos posted by users who were based in, or originated from, apparently, different regions, indicates that the use of Romani in writing allows users to communicate across national boundaries and the barriers of different languages in which members of the Romani minority community acquires literacy. This is evident in the case of the Pentecostalism videos (G3 and G9) and the complex merger that involved G5, the Czech network, and it suggests that some networks have the potential to attract a pan-Romani audience.

## 6. The content and format of Romani comments

Romani-language comments on YouTube videos do not differ largely from those posted in other languages as far as content and format are concerned. They are overall short, in some cases down to just a single word, and they usually address an unidentified audience, as can be seen in Example 1:

Network	Comment	Translation
G1	<b>milordo007 3 years ago</b> - te aven bahtale	May you all be lucky
G2	<b>Majuschi strauss 2 years ago</b> - sar buschon e gila ????	How are the songs called????
G3+G5+G9	<b>Marcel Tokoli 5 months ago</b> - Bach talimo muro pral	Good luck my brother
G4+G10	<b>JunnyTheGame2010 2 years ago</b> - haha ov sasto amala..poz	haha may you be well friend..bye
G6	<b>Jessy Negroni 1 year ago</b> - schukkaaaa <3	Niceeee <3
G7-G12	<b>MILOCIAWO 3 years ago</b> - dawa nani heladytka roma	These are not Xaladitka Roma

**Example 1:** Selection of short, broadcast comments, collected 29/07/2013

At times, however, people engage in conversations resorting to the reply function provided by YouTube.

Comments	Translation
<p><b>SIEJLALKA 5 years ago</b> - nais ⇒)</p> <p><b>LoljaBernalJohnson 5 years ago</b> - Te aves sasto aj baxtalo, nashti tradesa/bishavesa mange phrala kacha shukar djili?, del tu o Del bax!, anda savo them san tume?, e voja aj e bukuriya te avel pasha tumende. O Lolja. Argentina. -</p> <p>• <b>in reply to SIEJLALKA</b></p> <p><b>SIEJLALKA 5 years ago</b> - TE AVES BAXTALO!!! ⇒) me busiuvav Estera, ame sam LOVARA anda London, te kames bishavul tuke kadi djili feri de ma cho msn/email ⇒) GOD BLESS YOU !!! ⇒)</p> <p>• <b>in reply to LoljaBernalJohnson</b></p> <p><b>LoljaBernalJohnson 5 years ago</b> - No, Nais tuke, kothe bishadem le tuke, vi o msn vi o e-mailo, vi murri dey si loveritsa, murro dad sas Kalderash. God bless you too. Ando London beshen tume?, numa tumare dada sas Polskake, nay?</p> <p>• <b>in reply to SIEJLALKA</b></p>	<p>thanks ⇒)</p> <p>May you be well and lucky, can't you send me this nice song brother?, may God give you luck!, from which country are you?, may all your desires and good fortunes be fulfilled. Lolja. Argentina. -</p> <p>MAY YOU BE LUCKY!!! ⇒) My name is Estera, we are LOVARA from London, if you want I'll send you this song if you give me your msn/email ⇒) GOD BLESS YOU !!! ⇒)</p> <p>Well, Thank you, I sent it to you there, both by msn and by e-mail, my wife is also a Lovara, my father was Kalderash. God bless you too. Do you live in London? but your ancestors were Polish, weren't they?</p>

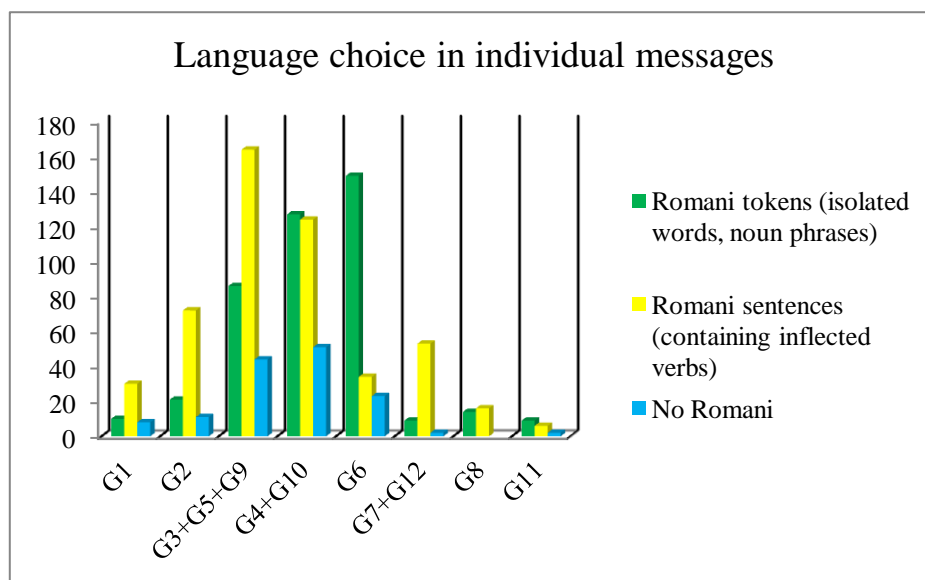
**Example 2:** Conversation using the reply function, from G2 (*lovara*), collected 29/07/2013

Users also engage in conversations without resorting to the reply function. This, however, results in the comments being displayed in the reverse order, with the earlier ones further down on the page than the latest ones, as can be seen in Example 3.

Comments	Translation
<b>TheLilou1501 1 year ago</b> - me sem me ta kava si mo abav alors	This is me and so this is my wedding
<b>riza muslija 1 year ago</b> - ko san tu ?? theLilou1501	who are you ?? theLilou1501
<b>TheLilou1501 1 year ago</b> - but lace sasa kov abav ⇒)	this was a very nice wedding ⇒)

**Example 3:** Conversation without using the reply function, from G8 (*abijav*), collected 29/07/2013

Finally, once users with exclusively non-Romani comments were excluded it became clear that users who do write in Romani do so consistently, as it can be seen in Figure 3, and so we can say that Romani is their preferred language. The majority of messages in almost all the networks contain entire Romani sentences, including fully inflected verbs, even when Romani is used alongside other languages. Furthermore, even in those messages in which a different language dominates in an individual comment, users also tend to insert Romani tokens – single words or noun phrases. The frequency of Romani sentences generally outranks the insertion of single Romani tokens. An exception to this behaviour is found in G6, where Romani tokens occur nearly five times as frequently as Romani sentences.



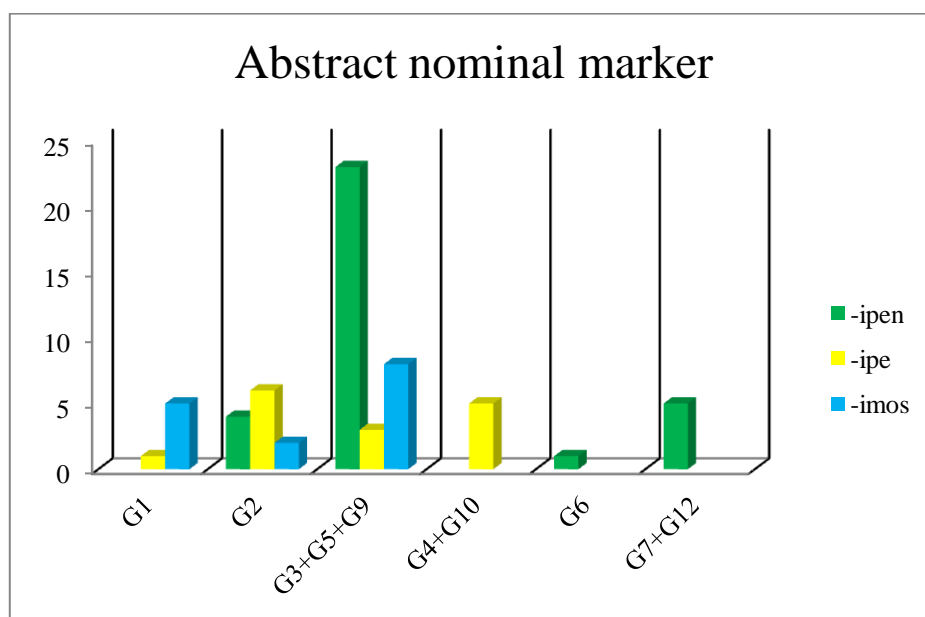
**Figure 3:** Romani usage. Numbers indicate the total count

In G8 and G11 the number of comments is very low and there is not much difference between the frequency of messages that contain full sentences in Romani and of those that merely contain Romani tokens. This is probably due to the fact that these two networks contain more recent videos, all uploaded between the end of 2012 and mid-2013, and so they have not yet had time to attract a larger audience. As a consequence, the token count for dialectal features and spelling variants was extremely low in these two networks, which hindered a proper comparison with the other networks. We therefore decided to exclude them from further analysis.

## 7. Choice of dialect variants

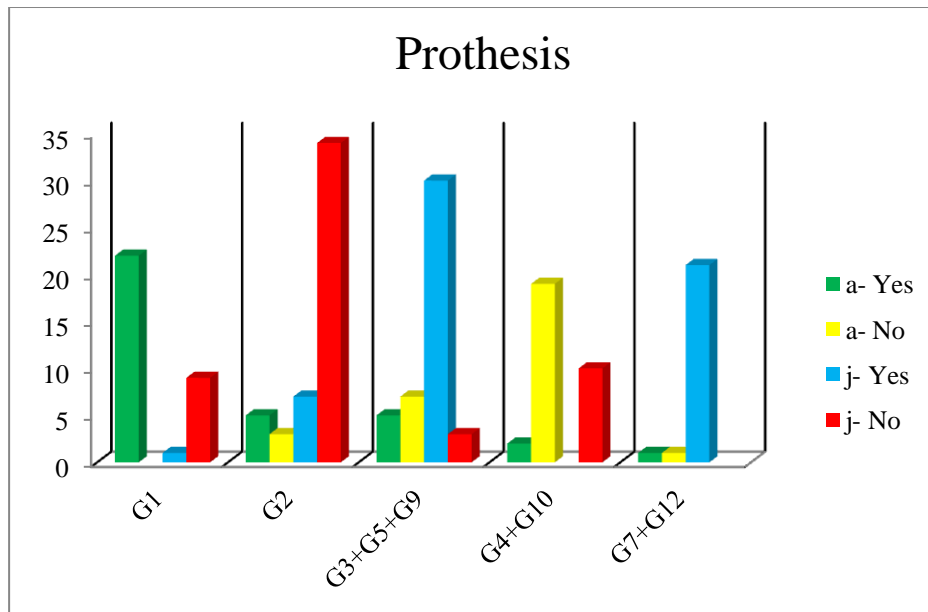
For the abstract nominal marker, we see that some networks show consistency in the choice of variants, while others are mixed. Figure 4 shows how G4+G10 are characterized by the exclusive usage of the Southern innovation *-ipe*, while G7+G12 show exclusive usage of the Northern conservative form *-ipen* and G1 shows predominantly the Vlax innovation (Romanian-Moldovan-Ukrainian Romani) form *-imo(s)*. The other networks G2 and G3+G5+G9 show all the variants. However, in G2 Southern and Vlax innovations together account for twice as many tokens as the conservative Northern form. The reverse pattern is found in G3+G5+G9, where Northern forms account for twice as many tokens as the Southern and Vlax forms combined. The higher token frequency of abstract nominal markers in this last network, which revolves around Pentecostalism, correlates with the frequent

occurrence of greetings such as *te del o Del sastipen* ‘may God give you health’. The position of G6 is difficult to assess, since only one single occurrence of a Northern form was found.



**Figure 4:** Abstract nominal marker. Numbers indicate the total count

Figure 5 shows results for the consonantal prothesis in *j-* and the vowel prothesis in *a-*. Like the distribution of the abstract nominaliser, the occurrence of prothesis phenomena shows clear patterns and the network profiles were similar. Thus, users in G4+G10 show forms that are typical of Southern Balkan dialects: the absence, by and large, of both *j*-prothesis and *a*-prothesis. In G7+G12 all relevant tokens showed *j*-prothesis, indicating that the users originated from Northern or Central Europe. In G1, the high frequency of *a*-prothesis and absence of *j*-prothesis point to a Vlax background. In the case of G2 and G3+G5+G9 we again found all the different variants. However, we also see, once again, that in G2 variants that are consistent with Southern Romani (i.e. the absence of *j*-prothesis) dominate, while in G3+G5+G9 variants that are consistent with Northern dialects (i.e. the presence of *j*-prothesis) are in the majority. G6 did not include any of the relevant target words.



**Figure 5:** Prothesis phenomena

For the past-tense person markers of the 2SG (Figure 6), we found *-al* only in G3+G5+G9, albeit in a small number of cases compared to the counterpart variant *-an*, and in G6, where it is used exclusively. For G6, the results may suggest an origin of most users in the Sinti-Manouche community, though the extremely low token count does not allow a clear conclusion on the basis of this isolated feature (but see discussion below). For G3+G5+G9, considering the tendency seen above to pattern with Northern dialects, the low frequency of *-al* suggests that ‘North-eastern’ speakers (originating from Romani communities in central and northern Poland, the Baltics or Russia) may constitute a majority. For G7+G12, the absence of *-al* despite the presence of various features that are shared by the Central and Northern dialects similarly suggests an origin in the North-eastern cluster, as it rules out both the Sinti-Manouche (and Finnish) or ‘Northwestern’ cluster and the central European dialects. For the other networks (G1, G2, G4+G10), the complete absence of *-al* tokens is in line with the tendencies observed above to pattern with Southern usages.

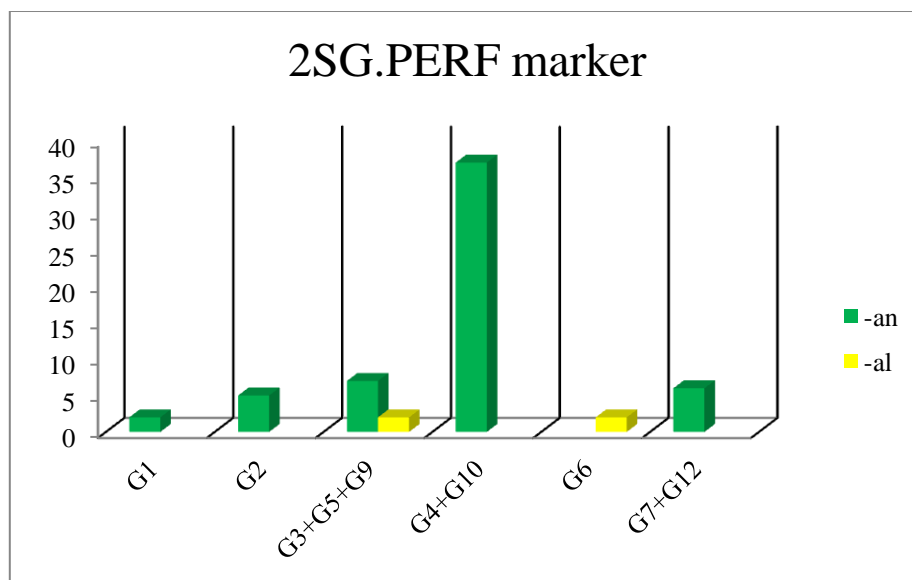
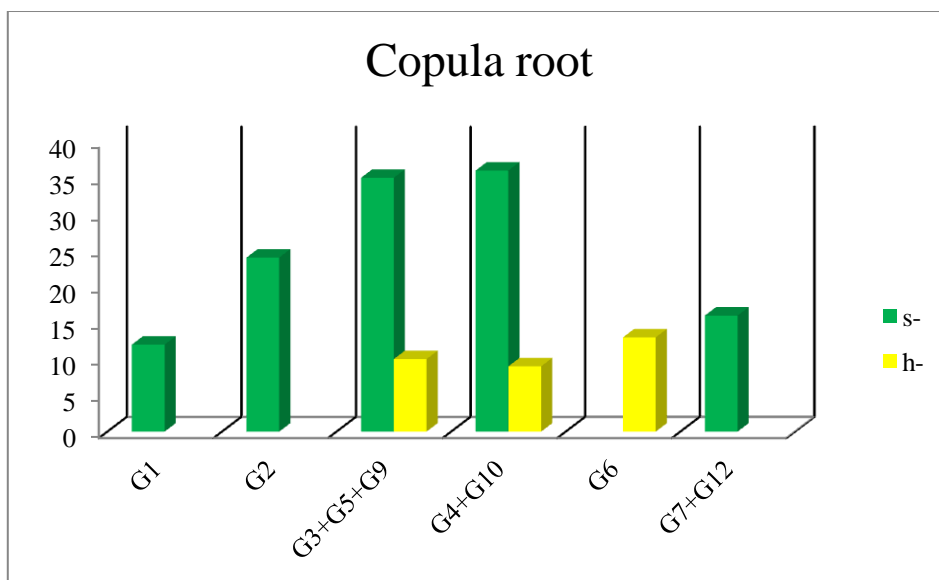


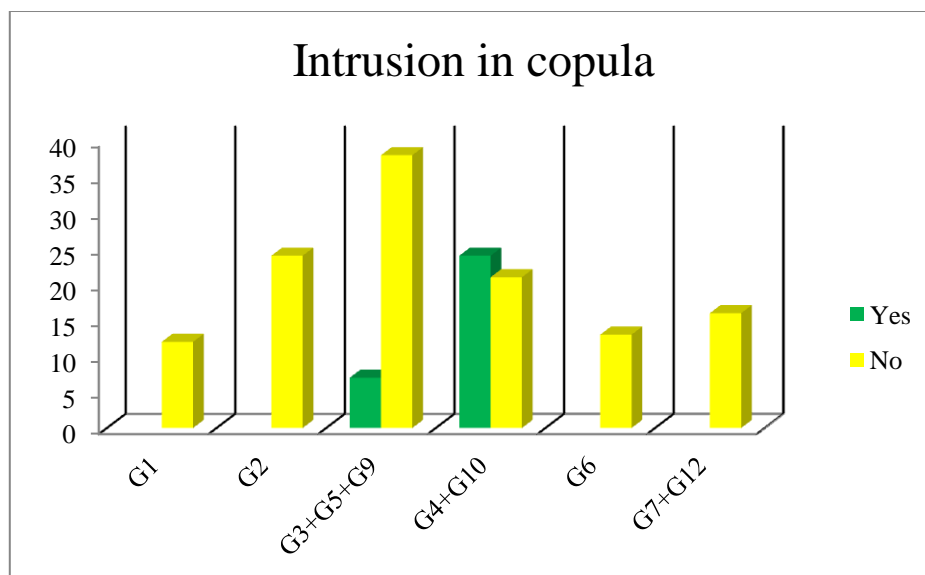
Figure 6: 2SG.PERF marker

In our sample *h-* copulas were encountered in G3+G5+G9, G4+G10 and G6 (Figure 7). In both G3+G5+G9 and G4+G10, *h-* copulas constituted a minority of tokens and appeared alongside counterpart forms in *s-*. The variation confirms the impression gained earlier that users in G4+G10 originate from the Southern Balkans, and more specifically, that they are speakers of the Arli varieties (spoken mainly in Macedonia and Bulgaria), in which both copula stems can be found. In the case of G3+G5+G9, the occurrence of *h-* copulas coupled with the presence of typically Northern forms seen before indicates a presence of speakers of the Sinti-Manouche (Northwestern) or Central dialects among the users. In G6, *h-* copula forms were used exclusively. This is the first case in which users commenting on G6 videos show a clear selection pattern that is backed up by a significantly high token frequency. Taking into account the distribution of other forms, the exclusive usage of *h-* copulas suggests the users in this network are Sinti-Manouche speakers. The complete absence of *h-* copulas in G1, G2 and G7+G12 is consistent with the usages so far identified for these networks, namely Vlax (Romanian-Transylvanian-Moldovan-Ukrainian cluster) for G1 and G2 and Northeastern (Polish-Baltic-Russian) for G7+G12.

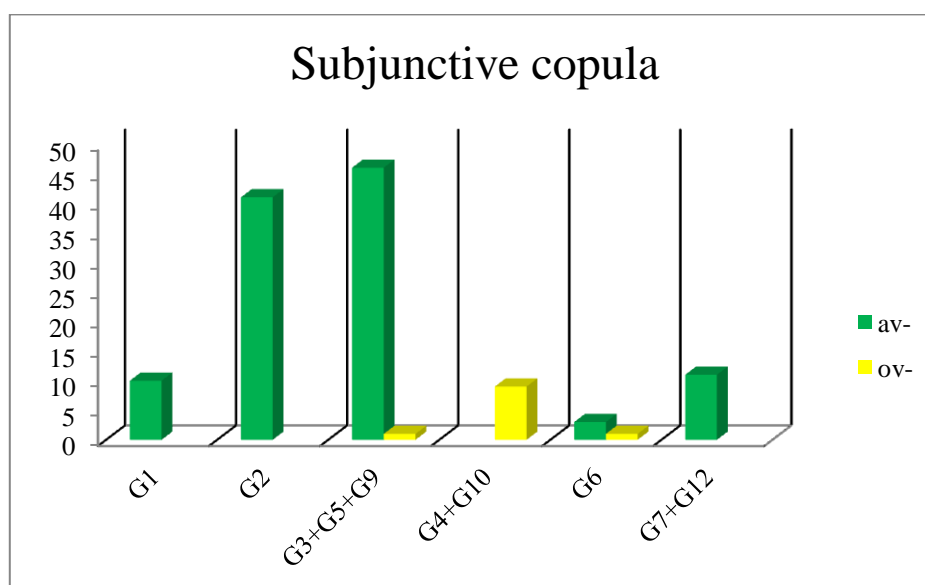


**Figure 7:** *s/h* alternation in the copula paradigm

Two more distinguishing features still relating to the copula are the intrusion of a stem extension *-ij-/in-* and the selection of *ov-* ‘to become’ or *av-* ‘to come’ as subjunctive copulas. The two features tend to overlap: Central and Southern Balkan dialects show insertion of the stem extension and selection of *ov-*, while in the remaining spaces no extension is encountered and the subjunctive copula is *av-* (Matras 2002: 230). For both features (Figures 8 and 9) the Central and Southern Balkan features co-occur only in G3+G5+G9 and G4+G10. By contrast, in G3+G5+G9 the occurrence of these features, but also the fact that their overall token count is low, together suggest the presence of a minority of users in this network of a central European (or Central dialect) background. In G4+G10, copula forms with intrusion occur only slightly more often than those without intrusion. This suggests that users of different backgrounds are commenting on these videos. Since so far users in this network have been seen to use Southeastern features, we can assume that, alongside the Southern Balkan speakers identified so far, a minority of Vlax speakers also form part of this network. The exclusive usage in this network of *ov-* as the subjunctive copula confirms that Southern Balkan speakers constitute the majority of users.



**Figure 8:** Intrusion of stem extension in the copula



**Figure 9:** Subjunctive copula

Once again G6 does not offer any conclusive indicators, as we find an overall low count of relevant tokens, and among them just one single occurrence of *ov-*.

For demonstratives (Figure 10), G1 shows only forms in *kad-* etc. In G2, the same forms are dominant, but we also found two tokens of *da-* forms. Since these reduced forms are typical of the North-eastern cluster of dialects (Poland, Baltics, Russia), and considering the occurrence of Northern tokens for the abstract nominalizer and in prothesis phenomena (see above), we can conclude that this network is primarily comprised of Vlach speakers but that it also attracts some speakers of North-eastern dialects. In G3+G5+G9 we found that Vlach

forms constituted the majority of the tokens. Central and North-eastern forms like *ada*, *da*, and *dava* were also frequent.

In G4+G10, the high occurrence of Southern Balkan demonstratives (*akava/adava*) confirms the dominant presence of southern Balkan speakers in this network. The occurrence of some *kava* tokens also confirms the presence of speakers from a different background. These forms, beside being characteristic of the Northwestern cluster, are also present in out-migrant Vlax varieties spoken in the southern Balkans, such as the dialect of the Gurbet (Boretzky 1986, Leggio 2011). Since no Northern feature was found in this network, the occurrence of *kava* demonstratives is a clear indicator of the presence of Gurbet speakers among G4+G10 users. G7+G12, with high occurrence of *dava* and *da* demonstratives, can be associated once again with North-eastern speakers, although two *kada* tokens indicates the participating of at least one Vlax speaker. As usual, G6 is the hardest network to evaluate due to the low token count. However, the fact that Northwestern demonstratives are slightly more numerous reinforces the impression that users in this network or of Sinti-Manouche background.

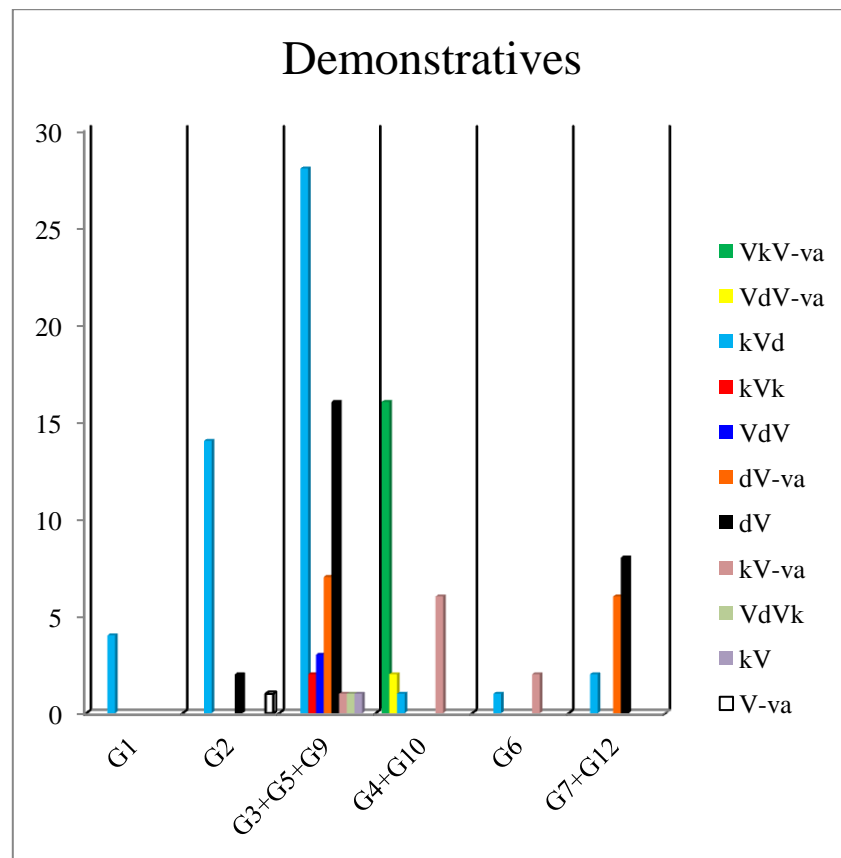


Figure 10: Demonstrative pronouns

The analysis of dialectal features confirmed most of our hypotheses with the exception of that for G1, one of the ‘wedding’ networks, which we expected to show high variation in dialect choice given the number of connections to other networks. Instead this network showed the consistent use of Vlax features.

G2, the Lovara network, showed predominantly Vlax forms, although some sporadic North-eastern features were found as well. In G7+G12, the exclusive presence of North-eastern features confirms our expectation, based on the content of the videos, (weddings and comments on the Polish Romani singer Francuzo), that users belong to the Polish Roma community. G4+G10, comprising wedding videos and comments on the Macedonian singer Ševčet, shows Arli (Southern Balkan) and Gurbet (Vlax) features, closely mirroring the Macedonian model of Romani codification in the selection of dialectal variants (Friedman 1995, 1996, 1997, 2005). G3+G5+G9, the Pentecostalism network, showed as expected the highest variation in dialectal features. The high occurrence of North-eastern features indicates that the majority of users share the same background, while a smaller number of features suggests a minority of users are of Central (East Slovak and Romungro) and Vlax backgrounds. It is noteworthy that although G3+G5+G9 is linked to G1 and G2, no Central features are found in G1 and G2 and only few Northeastern features can be found in G2. Since each edge represent users commenting on two videos, we can conclude that it is mostly the Vlax speakers commenting on G1 and G2 who also comment on the Pentecostalism network. On the other hand the Central and Northeastern users in G3+G5+G9 do not seem to contribute much to the comments in G1 and G2. This further highlights how the Pentecostalism network attracts Romani speakers from different backgrounds, obviously based on shared religious practices.

The background of users commenting in G6, comprising videos of the Sinti (German Romani) singer Prinzo, would have been difficult to determine if it was not for the exclusive usage of *h-* copulas, confirming our expectation to find Sinti users in this network. Since Sinti have historically been unwilling to write Romani (Matras 2004) it could be the case that we are witnessing a change of attitudes among Internet users. They, however, are still not as confident as other Romani speakers and therefore resort more to a tokenistic usage of the written language (see Figure 3), resulting in the very low token count for most dialectal features. This, however, matched our expectation that a predominantly tokenistic usage of Romani would make it more difficult to ascertain which Romani variety is used.

## 8. Choice of spelling

A point of interest in the analysis of the corpus is the representation of Romani sounds for which the various contact languages or national languages employ different graphemic representations. Diacritics such as {š}, {đ}, {ž} and {ć} are used in various alphabets of the relevant contact languages, and the symbols {č}, {dž}, {š} and {ž}) have a tradition of being used in the transliteration of Romani in academic and some activist publications. In the comments on YouTube videos, diacritics occur only rarely. Instead we find plain counterparts, that is, the relevant characters lacking diacritic symbols. This behaviour of Romani YouTube users mirrors the established CMC norm of majority language speakers to employ plain characters when diacritics are not available or difficult to transmit (Hentschel 1998 for Serbo/Croatian, see Anis 2007 for examples from French).

In the case of the unvoiced postalveolar affricate /č/ (Figure 11), we can observe how, with the exception of G2, G6 and G7+G12 the favourite choice is {c}. Their main model for imitation is, it appears, {č}, which is used in the majority of Romani printed materials following the usage of Slavic languages of central and southeastern Europe (Czech, Slovak, Slovene, Croatian). Another solution adopted by Romani speakers in CMC environments to represent /č/ is English-like {ch} (in some cases possibly modelled on Spanish; see below). This is found in all networks and is the favourite choice in G2, the Lovara network. In G6 (the Prinzo network) the German spelling {tch} is used almost exclusively. In G7+G12 (weddings and Francuzo videos), Polish grapheme combinations {ci} and {cz} have been adopted almost exclusively, suggesting that the tokens of {c} found in this network to represent the sound /č/ are most likely to be modelled on the Polish grapheme {ć}.

The occurrence of {ci} in G2 and G3+G5+G9 is in all likelihood also based on Polish norms. For G2, we know that there is a strong Lovara community based in or originating from Poland. For G3+G5+G9, most users show North-eastern dialect features, which are in line with the dialect of the so-called Polska Roma of central and northern Poland. On this basis, we can once again interpret the grapheme {c} used to represent the sound /č/ as modelled on Polish {ć}.

Network G4+G10 showed no North-eastern dialect features, and its content suggests that it attracts users who belong to diasporic Romani communities originating from former Yugoslavia. We can therefore assume that tokens of {ci} are modelled on Italian orthography, since various members of these communities have settled in Italy. Similar spelling selections are found in other CMC environments that are dominated by Romani

communities from former Yugoslavia (Leggio 2013). The choice of {c} to represent /č/ before front vowels in G4+G10 can thus also be attributed to Italian influence.

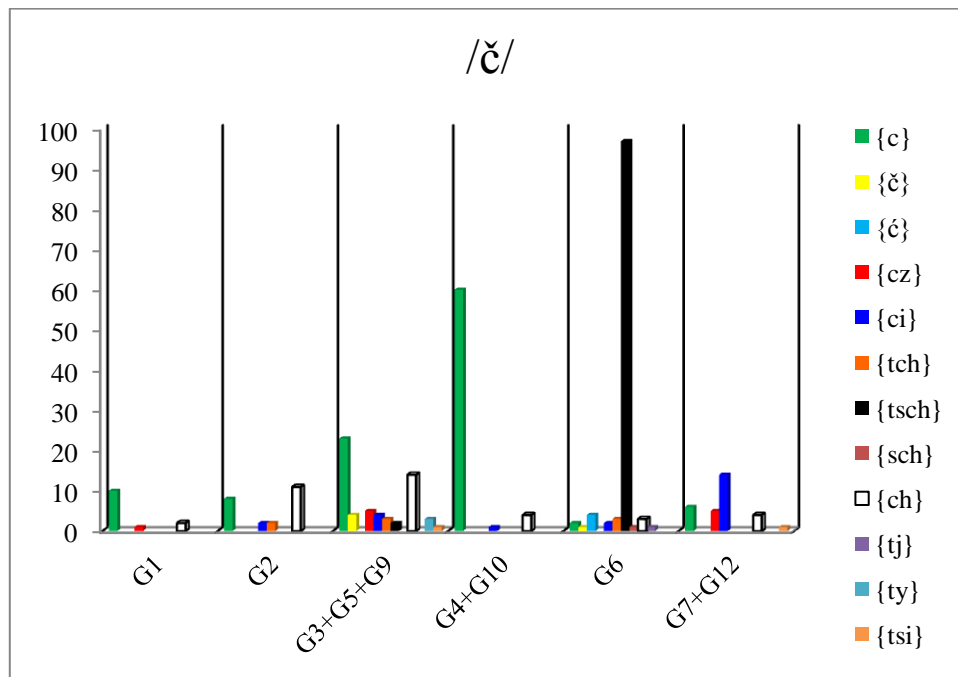


Figure 11: Unvoiced postalveolar affricate

The representation of the voiced postalveolar affricate /dž/ shows multiple sources within almost each and every network (Figure 12). The exception in this respect is G7+G12, which again shows Polish influence in the predominant use of {dzi}. This in turn suggests that {dz} is modelled on Polish {dź}. The same solutions are also prevalent in G3+G5+G9, but in this network we also find, in order of frequency, {d} modelled on Serbian/Croatian {đ}, Hungarian {gy} and Romanian {j} (note that in the Vlax dialects of Romania and neighbouring countries and region, the Romani phoneme /dž/ often undergoes de-affrication to /ž/, the sound that is represented in Romanian orthography by {j}). Considering this variety of sources, it is possible that some of the tokens for {dž} derive from Serbian/Croatian or Czech/Slovak {dž}. G2 shows {dj} and {d}, apparently based on the South Slavic mode, alongside conventions based on a Polish model. The same South Slavic spellings and one token of {đ} are also found in G4+G10. Since {dzi} is not used in this network, the tokens of {dz} appear to replicate South Slavic {dž} rather than Polish {dź}. The tokens of {g} and {gi} found in this network point again to influence from Italian, where {g} is used before front vowels and {gi} before central and back vowels to represent the sound /dž/. The tokens of {gj} represent either a compromise between Italian {gi} and South

Slavic {dj} or an attempt to represent palatalized pronunciations of /g/ before front vowels, a phonological development of the Romani dialects of former Yugoslavia that is shared with neighbouring Slavic languages (Matras 2002: 49). The use of {c} in positions where we would expect the phoneme /dž/ points to German influence. It is modelled, apparently, on the tendency in German for /dž/ to merge with /č/, and then on the choice of {c} as a substitute for or simplification of the diacritic modification of the South Slavic character {č}. One user in G6, whose dialect features are coherent with the Sinti-Manouche variety, opts for {dj} while otherwise using graphemes that are modelled on German. In G1, with just two tokens, we found {gi}, modelled on Italian, which suggests that, in this network, as in G4+G10, even some of the {c} for /č/ might copy Italian conventions.

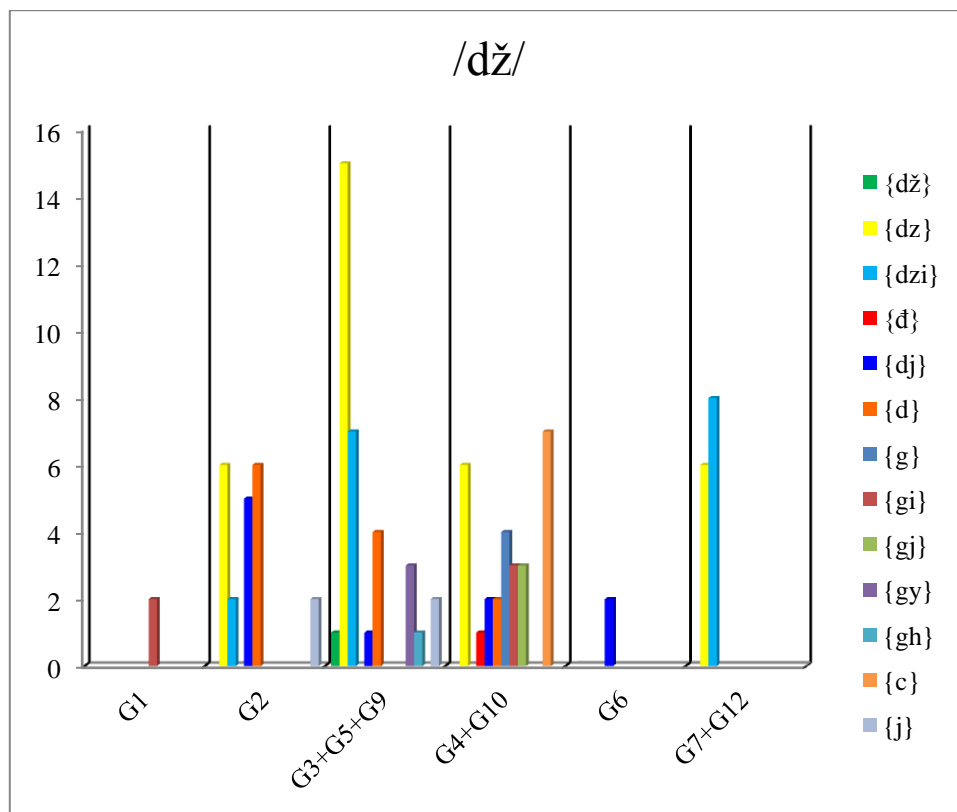


Figure 12: Voiced postalveolar affricate

The spellings used for the voiceless postalveolar sibilant /š/ show a similar distribution across networks to that observed for /č/: Most networks show predominantly {s} (Figure 13). Once again we find different solutions in G7+G12 and G6. In G7+G12, we find the Polish grapheme combinations {sz} and {si} and only a few tokens of {s}, presumably replicating Polish {ś}. G6 once again shows almost exclusive usage of a German-based grapheme

combination – {sch}. English-like {sh} is found as the second preferred choice in all networks except for G7+G12. In G3+G5+G9 we find diverse representations. The usage of {s} in this network is, however, more frequent, and it can be regarded as modelled on multiple sources. Some users in this network employed {s} for /š/, modelled on Hungarian, while at the same time using the Hungarian grapheme combination {sz} to represent the dental sibilant /s/. In the remaining networks, {s} is always the favourite choice. In G1 and G4+G10 it is in all likelihood modelled on the central and southeastern European Slavic Grapheme {š}. In G2, considering the Polish influences found in this network it might replicate either {š} or {ś}. The high frequency of /š/ in G4+G10 is due to the frequent use of the Balkan Romani exclamation *jasa* /jaša/ (from Turkish *yaşa!* ‘long live!’), common in group dances and music performances (see Leggio 2013 about the same behaviour in other CMC environments). It is interesting to note that in G2 we found a single user employing the Cyrillic alphabet, thus using {III}.

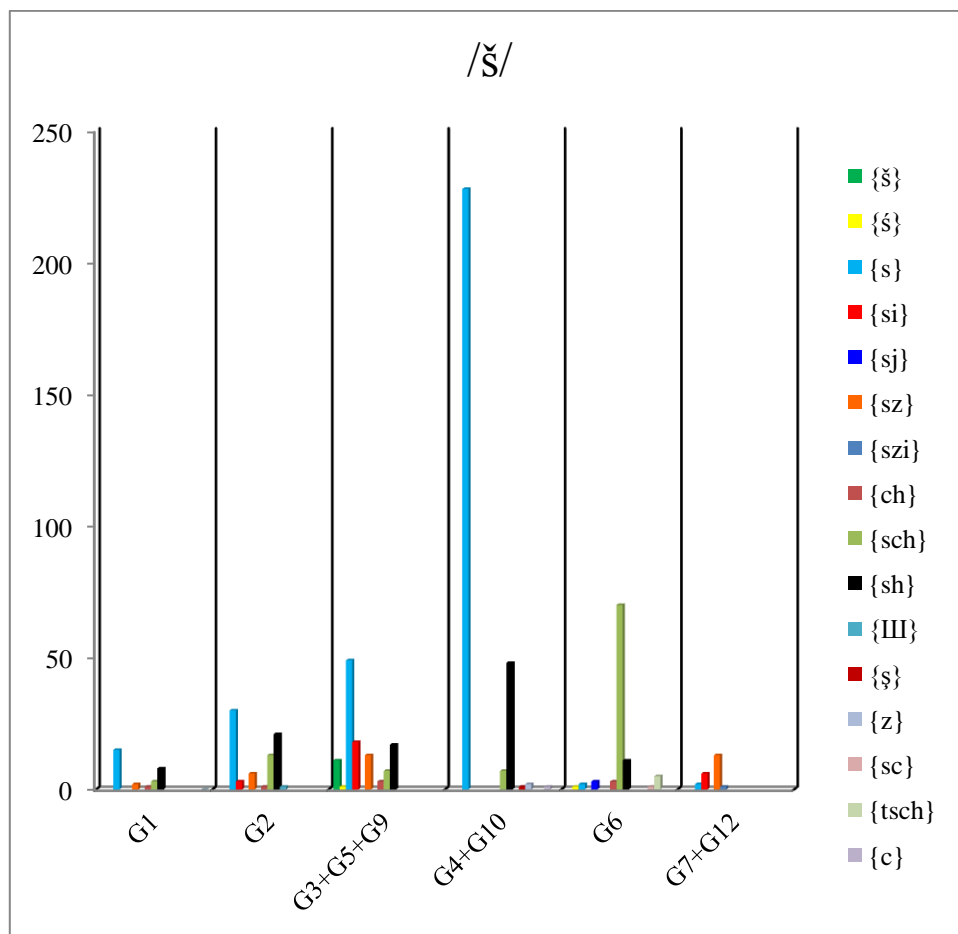


Figure 13: Unvoiced postalveolar fricative

The voiced postalveolar fricative /ž/ is a rare sound in Romani. In some dialects it emerges through de-affrication of /dž/. Elsewhere it is limited to loanwords. The most commonly employed spelling for this sound is {z} (Figure 14). In G7+G12, where we have found that spelling is modelled on Polish, {z} appears to replicate {ż} or {ź}. In G3+G5+G9 we find Romanian {j}, Hungarian {zs}, Polish {zi} and Serbian/Croatian or Czech/Slovak {ž}. The Grapheme {z} in this network appears to be modelled on more than one source, as a simplification of {ž}, {ż} or {ź}. In G2, where very few tokens were encountered, Hungarian {zs} is the more frequent, while the presence of Polish {zi} suggests that {z} in this network might be modelled on Polish {ż} or {ź}. A single token of {ż} in G6 appears in a message by a user (abcde814) who employs a series of ‘original’ graphemic solutions for Romani, some of them obviously modelled on academic conventions and/or various activists’ standardisation proposals, and which include {ř}, {ć}, {ś}, {č} and {ê}, all of which are rather exceptional in the corpus. In G1, G3+G5+G9 and G4+G10 we found some instances of graphemic solutions that are generally used for /š/ representing what we would expect to be the sound /ž/. This suggests that some users confuse the two sounds, possibly due to the rarity of /ž/ in Romani.

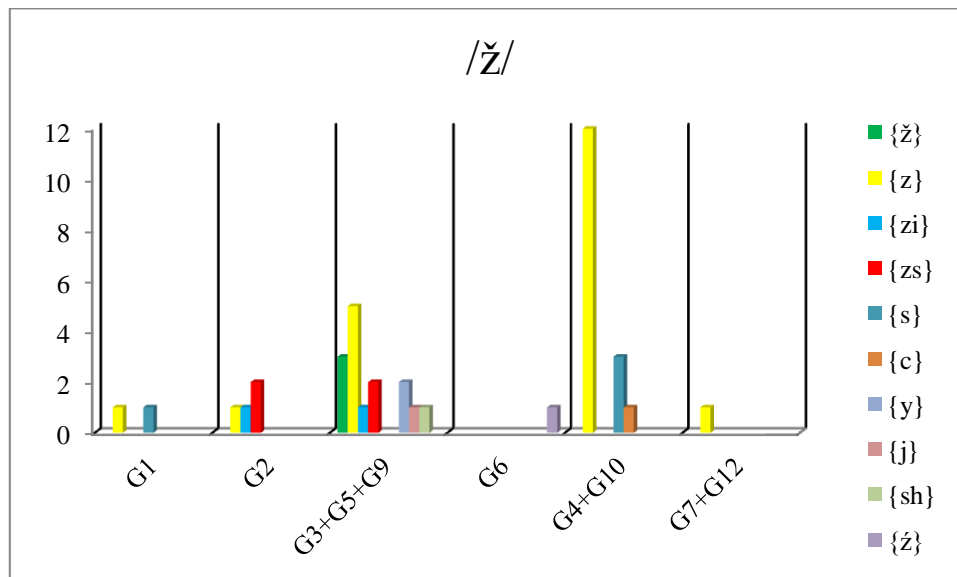


Figure 14: Voiced postalveolar fricative

Two main solutions are employed across the networks for the velar fricative /x/: {h} and {ch} (Figure 15). By contrast, the spelling employed in most academic and activist literature is {x}. Although this latter symbol is easily accessible on any keyboard, it is employed only occasionally in the corpus, suggesting that the vast majority of YouTube users has little

exposure to the published academic and activist literature in or on Romani, and is not influenced by it. It is noteworthy that in some Romani dialects of central and southeastern Europe, the phonemes /h/ and /x/ show a tendency toward merger, copying tendencies found in the contact languages. Both solutions used in the networks – {h} and {ch} – are available in Polish and Czech/Slovak. The fact that they show similar frequency in the networks G3+G5+G9 and G7+G12 shows once again how users in these two networks draw from these languages. In G6, the almost exclusive usage of {ch} points instead to a stronger German influence, while the exclusive usage of {h} in G4+G10 is based on Serbian/Croatian. The tokens of {j} and {g}, the latter always preceding front vowels, fund in G1, G2 and G3+G5+G9 show influences of yet another language, namely Spanish. Since many Roma, mostly Kalderash and Lovara are settled in Latin America and many of them belong to Pentecostal churches, it is not surprising to find them commenting on these networks. Furthermore, their presence implies that some of the tokens of {ch} for /č/ may be based not on English, but on Spanish. Finally, in G2 we find a single token of Cyrillic {X}

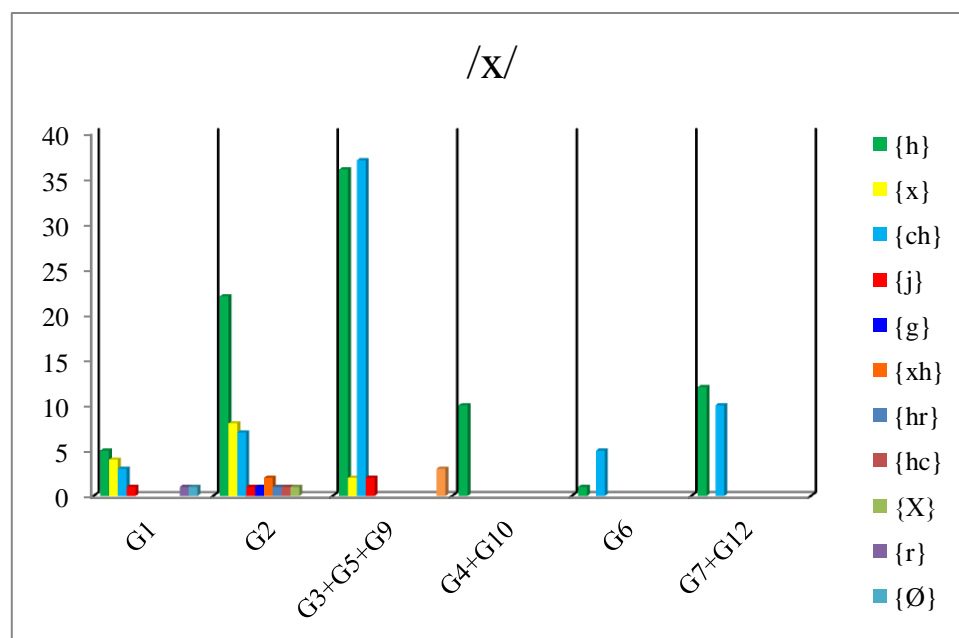


Figure 15: Velar fricative

For the semi-vowel /j/ in nearly all networks the preferred spelling is {j} (Figure 14). This is not surprising, considering that {j} is employed for /j/ in the majority of languages on which YouTube users are modelling their spelling. However, in Hungarian {j} alternates with {y} and in Polish it alternates with {i}. In Italian and Spanish {j} is not used for /j/ but

instead {i} is found, alternating with {y} in the case of Spanish. In the case of G1, the already observed Italian influence (see above regarding /dʒ/) and the instances of Spanish derived spellings for /x/, could explain why {i} is the preferred choice in this network. G7+G12 and G3+G5+G9 show a similar frequency of tokens for {j} and {i}, again confirming Polish influence. In G4+G10, the regular use of {ij} for /j/ in word-internal position once again shows how users in this network are modelling their spelling on Serbian/Croatian.

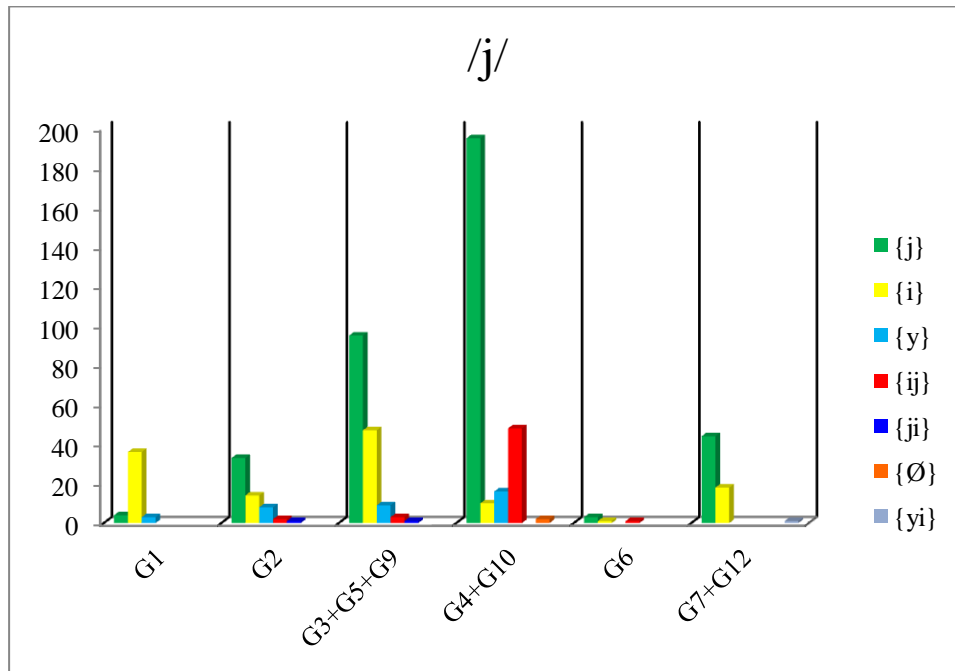


Figure 16: Palatal approximant

For the labio-dental fricative /v/ we found again that G6 and G7+G12 differed significantly from the other networks (Figure 17). In these two networks, {w} is used, showing again alignment with German (G6) and Polish (G7+GG12) orthographies. In the other networks, {v} is the preferred choice, following the usage of other Slavic languages as well as Hungarian and Italian. The tokens of {w} found in these networks can be attributed either to the presence of users of Polish background (G2 and G3+G5+G9) or to users currently living in Germany (G4+G10). The tokens of {b} in G1 and G3+G5+G9 derive from Spanish, in which {b} and {v} are used to represent the bilabial fricative /β/, the sound closest to /v/ in this language. Cyrillic {B} was found in comments by the consistent ‘Cyrillic’ user in G2.

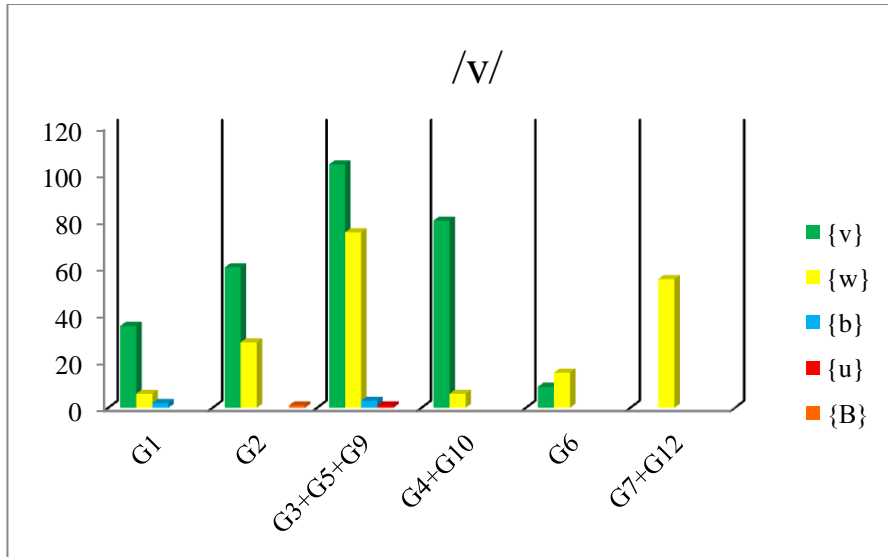


Figure 17: Labio-dental fricative

The centralized vowels /i/ and /ə/ have entered some Romani dialects through loanwords from the contact languages, and they often alternates with their non-centralized counterparts. Not all networks showed an indication of centralized vowels. As shown in Figure 18, the representation of /ə/ mostly follows the South Slavic practice of not representing it at all, or else the German norm that employs {e}. The tokens of {ê} in G6, were employed by user ‘abcde814’, and appear to reflect the influence of some academic transliteration conventions and/or activist standardization models. For {i}, on the other hand, the more common solution is {y} modelled on Polish (Figure 19).

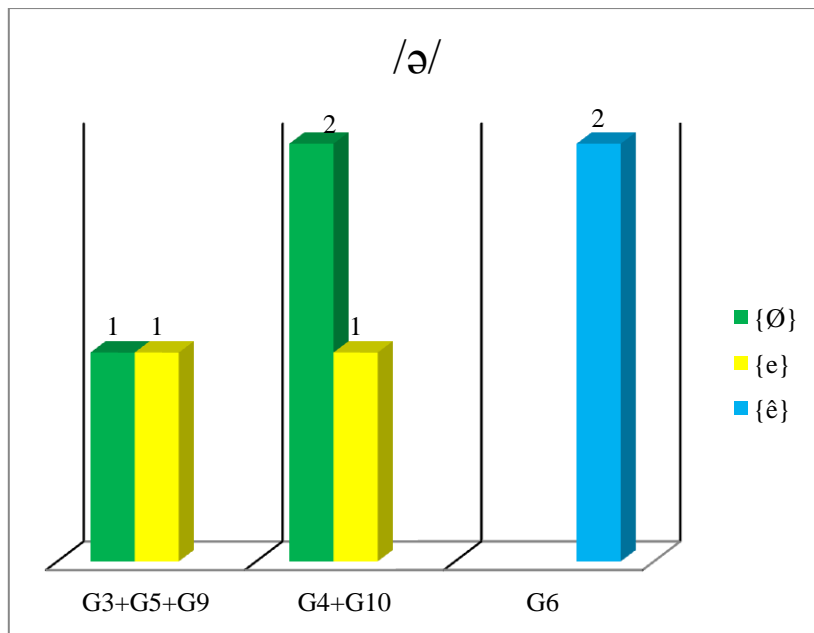
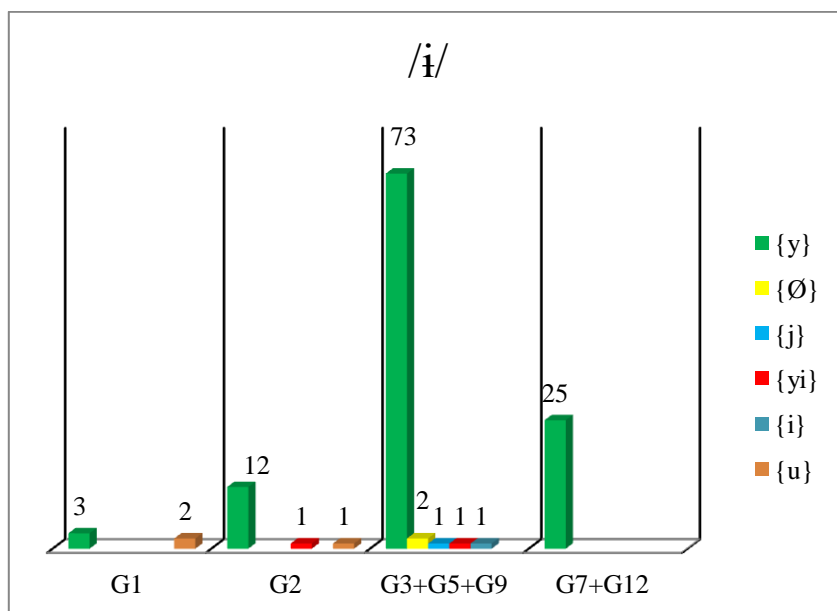
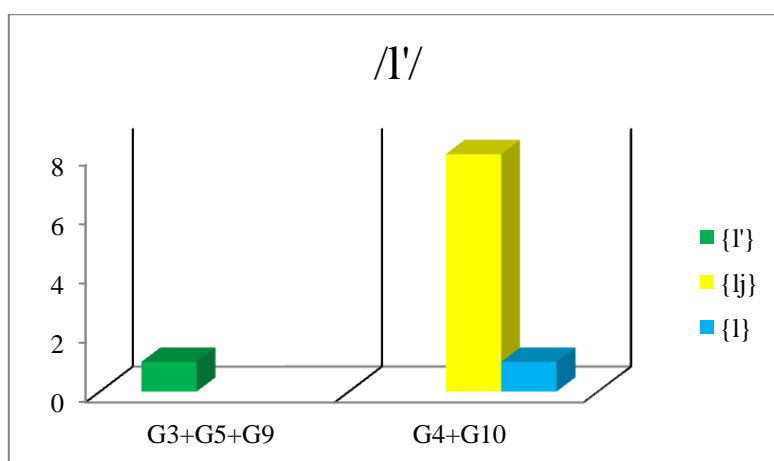


Figure 18: Mid-central vowel



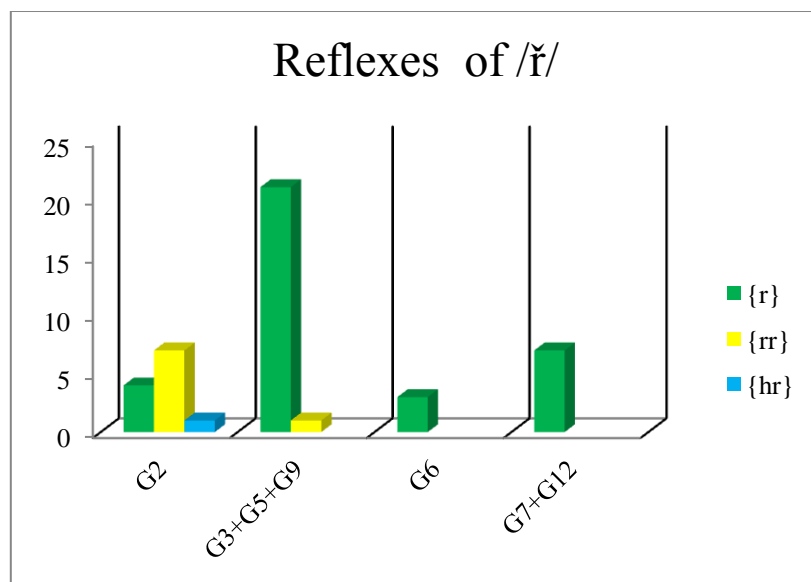
**Figure 19:** Close central vowel

Like centralized vowels, palatalized consonants /lʲ/, /rʲ/, /kʲ/ and /gʲ/ emerged in some Romani dialects through contact with other languages. We only identified single tokens of /rʲ/, /kʲ/ and /gʲ/, and never in more than one network. On the other hand, a relevant number of tokens for /lʲ/ were found in G4+G10. The appearance of /lʲ/ in G4+G10 is consistent with the dialectal profile of this network, since Arli and Gurbet speakers have acquired the sound through contact with Serbian/Croatian and Albanian. Even the spellings adopted mirror the norms of these two languages, with Serbian/Croatian {lj} being more frequent than Albanian {l}. The single token of {lʲ} found in G3+G5+G9, on the other hand, suggests the user employing it has had some degree of exposure to academic transcriptions of Romani.



**Figure 20:** Palatalized dental lateral approximant

While the sounds discussed so far are shared between Romani and the major European languages, the aspirated consonants /ph/, /th/, /čh/ and /kh/ and the distinctive /ř/, realised as [rr], [R] or [ɽ], are specific to Romani. Conscious of the symbolic value of these sounds as representing both the distinctiveness of Romani and its Indic legacy, activists have tended to emphasise their distinctive spelling. They tend to follow the academic transliteration conventions, where aspirated consonants are indicated by combining the stop grapheme with {h}, while /ř/ is usually rendered as {rr} or {ř}. As in the case of /x/, we found that academic and activist conventions are seldom followed in the YouTube corpus, either for aspirates or for the reflexes of /ř/ (Figures 21 to 25).



**Figure 21:** Continuation of early Romani /ř/

With some occasional exceptions, aspirated consonants are represented by respective simple consonant graphemes {p}, {t}, {k} and {č}. This trend among YouTube users confirms once again a lack of exposure to activist and academic literature, as was observed in the case of /x/. Furthermore, it shows that they are not using Romani as a symbol of unity among the Roma and of their shared Indian origin, as activists often do, but strictly as a communicative medium.

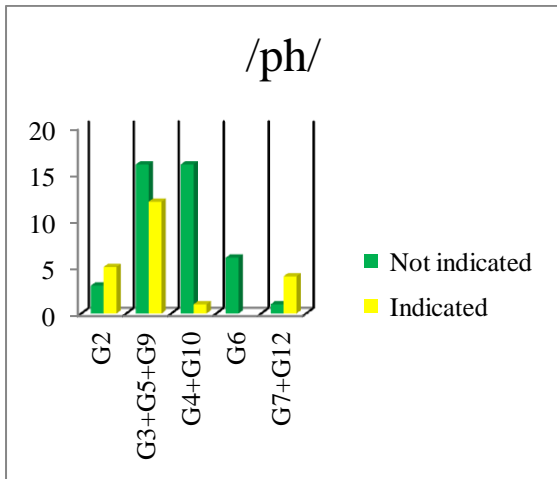


Figure 1: Aspirated unvoiced bilabial stop

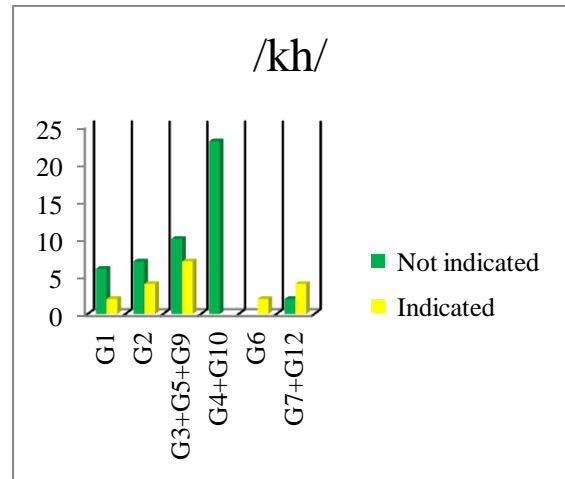


Figure 3: Aspirated unvoiced glottal stop

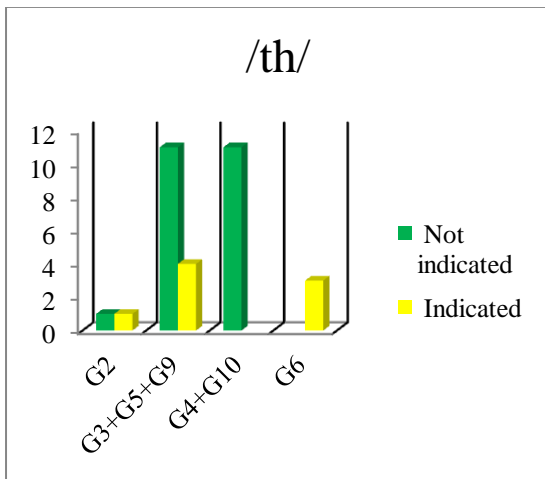


Figure 2: Aspirated unvoiced dental stop

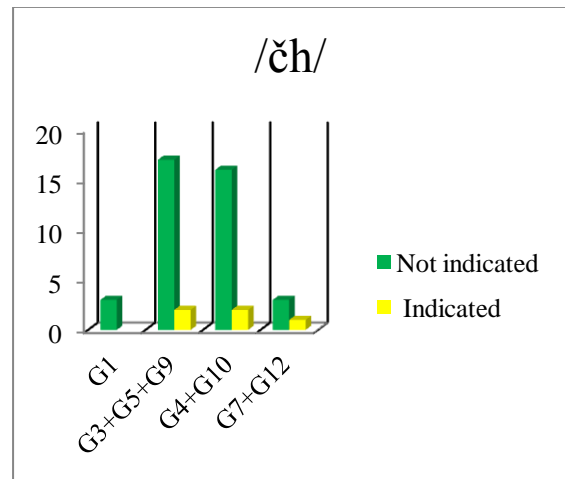


Figure 4: Aspirated unvoiced alveolar fricative

In G4+G10, G6 and G7+G12, the more regionally oriented networks comprising Romani speakers from former-Yugoslavia, Germany and Poland, respectively, users model their spelling on Serbian/Croatian, German and Polish orthographies. In G2 and G3+G5+G9, which include Romani speakers from different regions, users share the same influences as the ones in the previous three, but they also employ spellings derived from Hungarian, Czech/Slovak and Spanish.

G1, for which we expected a similar situation as in G2 and G3+G5+G9, instead showed a relatively high occurrence of Italian influenced spellings, suggesting many of its users are currently settled in Italy.

In spite of the variation in sources, the spelling emerging on YouTube is generally consistent across the networks. This is due to either similar solutions being employed in the model languages or to the common replacement of characters with diacritics by their plain counterparts, resulting for example in {č} and {ć} being replaced by {c}. In this respect it is interesting to notice that G6, the network that more consistently stray away from this consensus spelling, is also the most isolated one. Yet the preference for German spelling shown by the Sinti users of G6 does not stop users from different background from occasionally joining in and employing different spellings. However, while the Sinti users mostly employ Romani tokens, other users tend to write sentences in Romani, thus contributing more spelling tokens and slightly blurring the picture of spelling preferences in this network.

## **9. Conclusions**

The video networks discussed above generally represent Romani communities which, in historical perspective, share a tradition of direct, face to face oral communication in their locations of origin in particular regions in Europe. This can be clearly seen in the coherent use of dialectal features that are characteristic of distinct areas and of distinct Romani varieties or dialect clusters. However, with the exception of the Sinti Romani community of Germany and neighbouring regions, these communities are now dispersed among various geographical locations as a result of recent migrations from central and eastern European countries to the West. The coherence of user groups as communities of origin and hence as diasporas is evidenced by the coherence of the dialectal features employed in the networks, and to some extent in the shared interest in community-specific content, in particular family events and music, and in some cases explicit flagging of group identities like ‘Lovara’. The geographical dispersion of the commenters belonging to each network is evidenced by the mixture of linguistic repertoires, manifested by the choice of spelling and code-switches, as well as occasionally in the content of postings, which reveal the users’ locations. This character of the networks as diaspora user communities is thus represented by the combination of consistency of dialect features alongside pluralism in orthographic representation within each network.

In G1, one of the wedding-related networks, users employ Vlach dialectal features. The countries mentioned by some users as their location include Italy, France, Germany, Netherlands, Argentina, Chile and the United States, and these match some of the trajectories of the 19<sup>th</sup> century Vlach migration. It stemmed originally from the Banat-Transylvanian regions in western Romania, but reached western Europe and the Americas via eastern and southeastern Europe (Russia, Serbia, Greece); some of the Vlach communities in northern Europe and the Americas still refer to themselves as ‘Russians’, ‘Greeks’, or ‘Machwaya’ (the latter from Serbian Mačva). The national languages of these countries serve as spellings model for various users, yet the majority of users employ Serbian/Croatian- and Italian-based solutions. Since the most frequently used languages other than Romani are Italian and, occasionally, German and English, but no instances of South Slavic languages are found, we can conclude that a sizable portion of the users in G1 are Vlach speakers who settled in the Balkans in the 19<sup>th</sup> century and recently moved to western Europe, mostly into Italy.

In G2, the *lovara* network, users employ a variety of Romani that is closely related to that used in G1. American countries (Canada, Argentina, US, Brazil) were mentioned again, but while G1 users were identify their locations as based in western Europe, in G2 we found that users mostly referred to both eastern and western European countries (Russia, Ukraine, Poland, Sweden, Germany, France, UK). This difference is mirrored in the model languages for Romani spelling in G2. Central Slavic and Polish conventions are dominant, but instances of Romanian, Hungarian, German and Spanish can also be found. Moreover, the most frequently used languages beside Romani are Russian and Ukrainian, often employing the Cyrillic alphabet, and Polish.

The common origin of users in G1 and G2 in the population of so-called ‘Northern Vlach’ speakers is further confirmed by their shared use, although only occasional, of the negator *či*, a feature limited to Northern Vlach dialects only (Matras 2002: 114). Similarly, a typical Vlach sound change, /čh/ > /š/, is found to be represented in both networks by spellings like {savo} ‘boy’ or {sej} ‘girl’, rather than the more commonly found {cavo} and {cej}.

In G4+G10, combining wedding and music videos, the predominant usage of Arli (Southern Balkan) features, the occurrence of some Vlach features, and use of various Balkan languages (Macedonian, Serbian/Croatian) are consistent with the linguistic profile of Roma from former Yugoslavia. Furthermore, spellings like {dicel} ‘s/he hears’ or {dzili} ‘song’ show that some users palatalize velars before front vowels, a contact-induced sound change common in Džambazi and Gurbet, southern Vlach dialects of Macedonia, Montenegro and Kosovo. Similarly, instances of 3SG verbs in *-ol* ({cerol} ‘s/he does’) are a particular feature

of these same dialects (Matras 2002, Boretzky 2003, Leggio 2011). Like the users in G1 and G2, those in G4+G10 are also dispersed across various countries, although their settlement regions are limited to Western Europe. Germany, Italy, the Netherlands, Switzerland and Sweden are mentioned in the comments, but only German and marginally Italian are used alongside other languages. Similarly, the spelling employed in this network shows a smaller number of influences. The main model language is Serbian/Croatian, resulting in a spelling closely reminiscent of that employed in printed materials in Macedonia (Friedman 1995, 1996, 1997, 2005) but showing a number of influences from, in the first instance, Italian, as well as German.

External linguistic influences on the speech and writing of the Balkan Roma, who began emigrating from former Yugoslavia in the late 20<sup>th</sup> century, tend to be more consistent compare to those found in the various Vlax dialects, whose speakers had dispersed into various locations worldwide already in the second half of the 19<sup>th</sup> century. This partly accounts for the difference in the number of model languages between the ‘Vlax’ networks G1 and G2, on the one hand, and the ‘South Balkan’ G4+G10, on the other. This also accounts for the tendency toward uniformity of external influences on the dialect of the Polska Roma found in G7+G12, which similarly feature wedding and music videos. Beside the Northeastern features discussed above, the users in this network show consistent use of {pszal} ‘brother’ instead of the more commonly encountered {pral}, reflecting a pronunciation typical of the Polska Roma (Matras 1999a). Similarly, the frequent usage of {ɨ} where other users instead employ {ɪ} indicates the strongly velarized pronunciation of /ɪ/, resulting in substitution with the semi-vowel /w/, in the Polska Roma dialect (Matras 2002: 50). This development is borrowed from Polish, and the users’ choice to employ the Polish grapheme {ɨ} confirms their almost exclusive adherence to Polish writing norms. With the exception of some tokenistic usage of English words, the only other language beside Romani used in this network is Polish. Yet, in some of their comments some users mention living in countries other than Poland (Example 4).

Comment	Translation
grzesiek wlochkowski 1 year ago - sasto bahtalo te jawel da ciawo francuzo jow sy mire forostyr nottingham jaw sasto bahtalo de miro pszal hahahahahaha	may you be healthy and lucky this chap francuzo he is from my town nottingham may you be healthy my brother hahahahahaha

**Example 4:** Comment on migration from G7+G12, collected 29/07/2013

In G6, one of the music networks, use of Romani tokens rather than sentence prevails, yet we clearly recognized Sinti users. The prominent use of German and occasionally of French and Dutch alongside Romani confirms that users in this network are Sinti-Manouche speakers still residing in their historical locations. Since they did not migrate, they had no chance to come into contact with a multiplicity of writing norms, and this is clearly visible in their preferred spellings. A limited number of users modelled their spelling on Dutch (i.e.: {sjoeka gilie} ‘nice song’) and French (i.e.: {dourder} ‘further’) but the majority of them relied exclusively on German orthography. The few occurrences of non-German spellings in G6 always appear among individual users who also show dialectal features that are not typical to the Sinti-Manouche dialect cluster. Here we can see how, despite spelling differences, the choice of Romani allows people to communicate in writing even if they do not share a standardized contact language.

This opportunity offered by Romani is clearly taken up by users in G3+G5+G9, the Pentecostalism network. These users, by contrast to the ones in the previous networks, are not brought together by a shared regional origin, but rather by their common religious affiliation with Pentecostalism. Although Romungro, Slovak (Central) and Polska Romani (Northeastern) forms are prevalent, dialects from all over Europe and the Americas are employed in this network. The countries that users refer to as their current location once again range from the Americas to Eastern and Western Europe. Spelling conventions confirm this multiplicity of origins and current locations, as all influences identified in the other networks can be found in G3+G5+G9, although, following from the predominance of Central and North-Eastern Romani speakers, Czech/Slovak, Polish and Hungarian spellings occur more frequently.

Overall, we witness the emergence of a new domain of communication, made possible by the platform offered by YouTube and most likely also by other social networking sites. Romani speakers take to the technology in order to remain in contact or to establish new relationships with fellow Roma across distances. The expansion of Romani language communication into this new domain requires, of course, the use of a written language. This leads to unstructured literacy and to the emergence of new practices. The first and most obvious change is the use of Romani, so far the language of private communication within the family, in a public environment. We also see that Romani is being used to communicate with strangers (see Examples 2 and 3), while previously its use was largely limited to communication with acquaintances. More often than not, however, we witness Romani

speakers behaving just like YouTube users of other backgrounds, in that they broadcast their messages to an unidentified audience (see Examples 1 and 4).

It is important to emphasise that the availability of new communication technologies alone would have not made this language development possible. Crucial for the development is the need that Romani speakers feel to communicate across transnational networks, which arose as a result of migration and geographical dispersion. Moreover, users employ flexibility and creativity in employing and accepting all the resources available in their own and other users' linguistic repertoires. The lack of prescriptive attitudes regarding the appropriate way of writing Romani, clearly visible in the absence of comments criticising other users' choices, encourage a plurality of contributions. On this basis the emergence of Romani literacy can be described as an organic process rather than an engineered change implemented in the form of traditional or established language planning and policy procedures. The change in Romani is a bottom-up process that emerges spontaneously rather than at the initiative of an elite with a clear political agenda; it is unstructured, and none of the actors involved appear to be taking on any fixed institutional roles.

The development is pluri-centric, in that each YouTube network acts as an innovation centre for the selection of preferred dialect features and choice of orthography. Dialect features usually derive from the spoken variety of the users, while orthographic choices are modelled on the primary contact language or languages in which literacy has been acquired. In this respect, what we see happening on the Internet is not significantly different from the picture of printed media in Romani. Here too, the overwhelming tendency is for a pluricentric codification process, with centres consisting of individuals or small organizations that received recognition and some measure of support from national or European authorities (see Matras 1999b, Matras 2004, Friedman 2005, and Halwachs 2011, 2012). What is different on the Internet is the density of links between the various centres and the fact that all centres operate under the same constraint on the availability of particular characters. This results in considerably more reciprocal influences between the centres and in much more homogeneous, although still varying spelling conventions emanating from the different centres.

In this respect we can observe that the connection between territory and language, characteristic of traditional, established language standardization efforts, is still not entirely dismantled. However, the territory experienced by Romani speakers has never been a bounded space, but rather a diasporic space. In this diasporic space, each population cluster encompasses various spatial locations, which are in turn points of contact with different

languages and scripts. As all these languages and scripts are employed flexibly, pluralism in linguistic practice arises. Nonetheless, the choice of Romani has a clear identity construction function. This can be seen in particular in the case of the Pentecostal Roma, where the choice of Romani facilitates the development of non-territorial networks by including Romani speakers even if they lack a commonly shared contact language that can be used in written communication. Domain expansion via social networking sites is thus based on well-established off-line social groupings. Ideology – religious ideology, in our case – can, however, bring together people from different groupings based on a shared perception of their common identity. New media can therefore create language-based imagined communities, but these do not require either territorialisation or a uniform language.

## REFERENCES

- Androutsopoulos, J. (2006) Multilingualism, diaspora, and the Internet: codes and identities on German-based diaspora websites. *Journal of Sociolinguistics*, 10/4: 520-547.
- Anis, J. (2007) Neography: unconventional spelling in French SMS text messages. IN Danet, B. & Herring, S. (Eds.) *The multilingual Internet*. Oxford: Oxford University Press.
- Boretzky, N. (1986) Zur Sprache der Gurbet von Priština (Jugoslawien). *Giessener Hefte für Tsiganologie*, 3, 195-216.
- Boretzky, N. (2003) *Die Vlach-Dialekte des Romani. Strukturen - Sprachgeschichte - Verwandtschaftsverhältnisse - Dialektkarten*. Wiesbaden: Harrassowitz Verlag.
- Danet, B. & Herring, S. (Eds.) (2007) *The Multilingual Internet*. Oxford: Oxford University Press.
- Danet, B. & Herring, S. (2007b) Introduction. Welcome to the Multilingual Internet. IN Danet, B. & Herring, S. (Eds.) *The Multilingual Internet*. Oxford: Oxford University Press.
- Deuber, D. & Hinrichs, L. (2007) Dynamics of orthographic standardization in Jamaican Creole and Nigerian Pidgin. *World Englishes*, 26, 22-47.
- Elšik, V. & Matras, Y. (2006) *Markedness and language change*. Berlin: Mouton de Gruyter.
- Friedman, V. A. (1995) Romani standardization and status in the Republic of Macedonia. IN Matras, Y. (Ed.) *Romani in Contact. The History, Structure and Sociology of a Language*. Amsterdam: Benjamins.
- Friedman, V. A. (1996) Romani and the census in the Republic of Macedonia. *Journal of the Gypsy Lore Society*, 6/2, 89-101.
- Friedman, V. A. (1997) Linguistic form and content in the Romani-language press in the Republic of Macedonia. IN Matras, Y., Bakker, P. & Kyuchukov, H. (Eds.) *The Typology and Dialectology of Romani*. Amsterdam: Benjamins.
- Friedman, V. A. (2005) The Romani language in Macedonia in the Third Millennium: progress and problems. IN Schrammel, B., Halwachs, D. W. & Ambrosch, G. (Eds.) *General and Applied Romani Linguistics*. Munich: Lincom Europa.
- Gallardo-Camacho, J. & Jorge-Alonso, A. (2010) The low interaction of viewers of video clips on the internet The case study of YouTube Spain. *Revista Latina de Comunicación Social*, 65, 421-435. Retrieved 24/10/2013 from [http://www.revistalatinacs.org/10/art3/910\\_Malaga/32\\_GallardoEN.html](http://www.revistalatinacs.org/10/art3/910_Malaga/32_GallardoEN.html).

- Halwachs, D. (2011) Language planning and media: the case of Romani. *Current Issues in Language Planning*, 12 (3), 381-401.
- Halwachs, D. (2012) Functional expansion and language change: The case of Burgenland Romani. *Romani Studies*, 22 (1), 49-66.
- Hansen, D., Shneiderman, B. & Smith, M. A. (2011) *Analyzing social media networks with NodeXL: insights from a connected world*. Amsterdam: Morgan Kaufman.
- Hentschel, E. (1998) Communication on IRC. *Linguistik online*, 1(1). Retrieved May 11, 2011, from <http://www.linguistik-online.de/irc.htm>: last accessed on 26-06-2012.
- Herring, S. (2002) Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, 36, 109-168.
- Kaplan, R. B. & Baldauf, R. B. (1997) *Language planning from practice to theory*. Clavedon: Multilingual Matters.
- Lee, C. K. M. (2007) Linguistic features of email and ICQ instant messaging in Hong Kong. IN Danet, B. & Herring, S. (Eds.) *The Multilingual Internet*. Oxford: Oxford University Press.
- Leggio, D. V. (2011) The dialect of the Mitrovica Roma. *Romani Studies* 5, Vol. 21, No. 1, 57-113.
- Leggio, D. V. (2013) *Lace avilen ko radio*. Romani language and identity on the Internet. *School of Arts, Languages and Cultures*. Manchester: The University of Manchester.
- Matras, Y. (1999a) The speech of the Polska Roma: Some highlighted features and their implications for Romani dialectology. *Journal of the Gypsy Lore Society, Fifth Series*, 9 (1), 1-28.
- Matras, Y. (1999b) Writing Romani: the pragmatics of codification in a stateless language. *Applied Linguistics*, 20/4, 481-502.
- Matras, Y. (2002) *Romani. A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Y. (2004) The future of Romani: toward a policy of linguistic pluralism. *Roma Rights Quarterly*, 1:31-44  
[http://romani.humanities.manchester.ac.uk/downloads/2/Matras\\_Pluralism.pdf](http://romani.humanities.manchester.ac.uk/downloads/2/Matras_Pluralism.pdf).
- Matras, Y. (2005) The classification of Romani dialects: a geographic-historical perspective. IN Schrammel, B., Halwachs, D. W. & Ambrosch, G. (Eds.) *General and Applied Romani Linguistics*. Munich: Lincom Europa.
- Matras, Y. (2010) *Romani in Britain: The afterlife of a language*. Edinburgh: Edinburgh University Press.
- Matras, Y. (2013a) Language and the rise of a trans-national Romani identity. *RomIdent Working Paper 24*.
- Matras, Y. (2013b) Mapping the Romani dialects of Romania. *Romani Studies* . , Vol. 23, No. 2, 199-243.
- McClure, E. (2001) Oral and written Assyrian-English codeswitching. IN Jacobson, R. (Ed.) *Codeswitching Worldwide II*. Berlin: Mouton de Gruyter.
- Milroy, L. & Milroy, J. (1992) Social network and social class: toward an integrated sociolinguistic model. *Language in Society*, 21, 1-26.
- Palfreyman, D. & Al Khalil, M. (2007) "A funky language for teenzz to use": Representing Gulf Arabic in instant messaging. IN Danet, B. & Herring, S. (Eds.) *The Multilingual Internet*. Oxford: Oxford University Press.
- Paolillo, J. C. (1996) Language choice on soc.culture.punjab. *Electronic Journal of Communication*, 6:3 <http://ella.slis.indiana.edu/~paolillo/research/paolillo.publish.txt>.
- Paolillo, J. C. (2001) Language variation on Internet Relay Chat: a social network approach. *Journal of Sociolinguistics*, 5/2: 180-213.

- Rajah-Carrim, A. (2009) Use and standardisation of Mauritian Creole in electronically mediated communication. *Journal of Computer Mediated Communication*, 14: 484-508.
- Rotman, D. & Preece, J. (2010) The 'WeTube' in YouTube—creating an online community through video sharing. *International Journal of Web Based Communities*, 6, 317-333.
- Shoham, M. D., Arora, A. B. & Al-Busaidi, A. (2013) Writing on the Wall: An Online "Community" of YouTube Patrons as Communication Network or Cyber-Graffiti? *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE.
- Silverman, C. (2012) *Romani routes. Cultural politics and Balkan music in diaspora*. Oxford: Oxford University Press.
- Sperlich, W. B. (2005) Will cyberforums save endangered languages? A Niuean case study. *International Journal of the Sociology of Languages*, 172: 51-77.
- Thelwall, M., Sud, P. & Vis, F. (2012) Commenting on YouTube videos: From Guatemalan rock to el big bang. *Journal of the American Society for Information Science and Technology*, 63, 616-629.
- Themistocleous, C. (2010) Writing in a non-standard Greek variety: Romanized Cypriot Greek in online chat. *Writing Systems Research*, 2, 155-168.
- Tseliga, T. (2007) "It's all Greeklish to me!". Linguistic and sociocultural perspectives on Roman-alphabetized Greek in asynchronous computer-mediated communication. IN Danet, B. & Herring, S. (Eds.) *The Multilingual Internet*. Oxford: Oxford University Press.
- Vergani, M. & Zuev, D. (2011) Analysis of YouTube Videos Used by Activists in the Uyghur Nationalist Movement: combining quantitative and qualitative methods. *Journal of Contemporary China*, 20, 205-229.
- Wright, S. (2004a) Introduction. *IJMS: International Journal on Multicultural Societies*, 6:1, 5-13. UNESCO [www.unesco.org/shs/ijms/vol6/issue/art1](http://www.unesco.org/shs/ijms/vol6/issue/art1).
- Wright, S. (2004b) *Language policy and language planning. From nationalism to globalisation*. Houndmills: Palgrave Macmillan.
- Wright, S. (Ed.) (2004c) *Multilingualism on the Internet*. *IJMS: International Journal on Multicultural Societies* 6:1: UNESCO Social and Human Sciences. [www.unesco.org/shs/ijms/vol6/issue1](http://www.unesco.org/shs/ijms/vol6/issue1).

## APPENDIX 1 - NETWORK SUMMARIES

Network	G1
Keywords	<i>abiav</i>
Video content	weddings
Accessed on	29/07/2013
Earliest video posted on	31/07/2008
Latest video posted on	16/06/2011
Earliest comment posted on	31/07/2008
Latest comment posted on	29/05/2013
Total comments	68
Comments with Romani	40
Predominant Romani dialect/variety	Vlax
Likely region of origin	Balkans (Serbia, Macedonia); earlier Vlax migration from eastern Europe
Other languages in the messages	Italian, German, English,
References to users' locations	Italy, Argentina, France, Netherlands, Chile, Germany
Preferred spellings for Romani: Model language(s)	Serbian/Croatian, Italian, Czech/Slovak, Spanish, Hungarian, French, Polish

Network	G2
Keywords	<i>lovara</i>
Video content	weddings, music, car dealing
Accessed on	29/07/2013
Earliest video posted on	18/08/2007
Latest video posted on	01/04/2010
Earliest comment posted on	18/08/2007
Latest comment posted on	29/06/3013
Total comments	205
Comments with Romani	93
Predominant Romani dialect/variety	Vlax
Likely region of origin	Transilvania; early migration
Other languages in the messages	Russian, Ukrainian, Polish, Spanish, English, German
References to users' locations	Sweden, Russia, Canada, Germany, Argentina, UK, France, US, Brasil, Ukraine
Preferred spellings for Romani: Model language(s)	Czech/Slovak, Polish, Serbian/Croatian, Romanian, Hungarian, German, Spanish

Network	G3+G5+G9
Keywords	<i>ricardo kwiek+gypsy kubanec+khangeri</i>
Video content	Pentecostalism, music
Accessed on	20/07/2013
Earliest video posted on	27/12/2007
Latest video posted on	18/02/2013
Earliest comment posted on	27/12/2007
Latest comment posted on	20/07/2013
Total comments	1292
Comments with Romani	250
Predominant Romani dialect/variety	Slovak Romani, Romungro, Polska Romani, Vlax
Likely region of origin	Hungary, Czech Republic, Slovakia, Poland, Romania, Germany
Other languages in the messages	Czech, Slovak, Hungarian, Polish, English, German, Russian, Romanian, Spanish, French, Brazilian Portuguese
References to users' locations	Spain, Romania, Czech Republic, Slovakia, Chile, Brazil, Germany, Latvia, UK, US
Preferred spellings for Romani: Model language(s)	Czech/Slovak, Polish, Serbian/Croatian, Romanian, Hungarian, German, Spanish

Network	G4+G10
Keywords	<i>bijav+sevcet</i>
Video content	weddings, entertainment music
Accessed on	20/07/2013+22/07/2013
Earliest video posted on	03/04/2007
Latest video posted on	26/12/2012
Earliest comment posted on	03/04/2007
Latest comment posted on	14/07/2013
Total comments	960
Comments with Romani	251
Predominant Romani dialect/variety	Arli, Gurbet
Likely region of origin	Macedonia
Other languages in the messages	Macedonian, Serbian/Croatian, Italian, German, English
References to users' locations	Kosovo, Macedonia, Germany, Italy, Netherlands, Switzerland, Sweden
Preferred spellings for Romani: Model language(s)	Serbian/Croatian, Italian, German

Network	G6
Keywords	<i>prinzo</i>
Video content	entertainment music
Accessed on	22/07/2013
Earliest video posted on	23/07/2008
Latest video posted on	12/03/2010
Earliest comment posted on	23/07/2008
Latest comment posted on	22/06/2013
Total comments	330
Comments with Romani	183
Predominant Romani dialect/variety	Sinti
Likely region of origin	Germany, Netherlands, Italy, France
Other languages in the messages	German, French, Dutch, Romanian, English
References to users' locations	Germany, Italy
Preferred spellings for Romani: Model language(s)	German, French, Dutch, Italian

Network	G7+G12
Keywords	<i>abiav+francuzo</i>
Video content	weddings, entertainment music
Accessed on	29/07/2012+22/07/2012
Earliest video posted on	27/08/2009
Latest video posted on	07/09/2011
Earliest comment posted on	27/08/2009
Latest comment posted on	29/06/2013
Total comments	94
Comments with Romani	62
Predominant Romani dialect/variety	Polska Romani
Likely region of origin	Poland
Other languages in the messages	Polish, English
References to users' locations	UK
Preferred spellings for Romani: Model language(s)	Polish

Network	G8
Keywords	<i>abijav</i>
Video content	weddings
Accessed on	29/07/2013
Earliest video posted on	26/06/2012
Latest video posted on	22/11/2012
Earliest comment posted on	26/06/2012
Latest comment posted on	29/02/2013
Total comments	31
Comments with Romani	30
Predominant Romani dialect/variety	Vlax (southern)
Likely region of origin	Balkans (Serbia, Macedonia)
Other languages in the messages	Serbian/Croatian
References to users' locations	Italy
Preferred spellings for Romani: Model language(s)	Serbian/Croatian

Network	G11
Keywords	<i>bijav</i>
Video content	weddings
Accessed on	22/07/2013
Earliest video posted on	26/12/2012
Latest video posted on	28/05/2013
Earliest comment posted on	26/12/2012
Latest comment posted on	28/04/2013
Total comments	37
Comments with Romani	15
Predominant Romani dialect/variety	Southern Balkan
Likely region of origin	Balkans (Kosovo, Macedonia)
Other languages in the messages	Macedonian, German
References to users' locations	France, Macedonia, Bulgaria, Germany
Preferred spellings for Romani: Model language(s)	Serbian/Croatian

## APPENDIX 2 - VIDEO URLs

Video ID	URL
G10a	<a href="http://www.youtube.com/watch?v=47_BPShKIdg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=47_BPShKIdg&amp;feature=youtube_gdata_player</a>
G10b	<a href="http://www.youtube.com/watch?v=LbvX-zCtrvA&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=LbvX-zCtrvA&amp;feature=youtube_gdata_player</a>
G10c	<a href="http://www.youtube.com/watch?v=EmmWvqOrFe0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=EmmWvqOrFe0&amp;feature=youtube_gdata_player</a>
G10d	<a href="http://www.youtube.com/watch?v=1k_O3JcBi2s&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=1k_O3JcBi2s&amp;feature=youtube_gdata_player</a>
G11a	<a href="http://www.youtube.com/watch?v=xMatYtGzmQg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=xMatYtGzmQg&amp;feature=youtube_gdata_player</a>
G11b	<a href="http://www.youtube.com/watch?v=jpiHi_ELzAw&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=jpiHi_ELzAw&amp;feature=youtube_gdata_player</a>
G11c	<a href="http://www.youtube.com/watch?v=lHFxW7uG0Ds&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=lHFxW7uG0Ds&amp;feature=youtube_gdata_player</a>
G11d	<a href="http://www.youtube.com/watch?v=kSn1Q4k__8&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kSn1Q4k__8&amp;feature=youtube_gdata_player</a>
G12a	<a href="http://www.youtube.com/watch?v=b0EIB_YgHqA&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=b0EIB_YgHqA&amp;feature=youtube_gdata_player</a>
G12b	<a href="http://www.youtube.com/watch?v=RtJCOadXg18&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=RtJCOadXg18&amp;feature=youtube_gdata_player</a>
G12c	<a href="http://www.youtube.com/watch?v=FBOXfZ4ePrk&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=FBOXfZ4ePrk&amp;feature=youtube_gdata_player</a>
G1a	<a href="http://www.youtube.com/watch?v=jPIVZZwpjew&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=jPIVZZwpjew&amp;feature=youtube_gdata_player</a>
G1b	<a href="http://www.youtube.com/watch?v=qjKA8vegNsE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=qjKA8vegNsE&amp;feature=youtube_gdata_player</a>
G1c	<a href="http://www.youtube.com/watch?v=ilegPMtkMXg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=ilegPMtkMXg&amp;feature=youtube_gdata_player</a>
G1d	<a href="http://www.youtube.com/watch?v=KXO7SkeTHmc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=KXO7SkeTHmc&amp;feature=youtube_gdata_player</a>
G1e	<a href="http://www.youtube.com/watch?v=aoiTmJReQiE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=aoiTmJReQiE&amp;feature=youtube_gdata_player</a>
G1f	<a href="http://www.youtube.com/watch?v=UgSQBGRRsAc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=UgSQBGRRsAc&amp;feature=youtube_gdata_player</a>
G1g	<a href="http://www.youtube.com/watch?v=EabKcUylo18&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=EabKcUylo18&amp;feature=youtube_gdata_player</a>
G1h	<a href="http://www.youtube.com/watch?v=AcFmKHYLR8w&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=AcFmKHYLR8w&amp;feature=youtube_gdata_player</a>
G1j	<a href="http://www.youtube.com/watch?v=TFsEOuOD-ug&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=TFsEOuOD-ug&amp;feature=youtube_gdata_player</a>
G1k	<a href="http://www.youtube.com/watch?v=pkLSrF820tQ&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=pkLSrF820tQ&amp;feature=youtube_gdata_player</a>
G2a	<a href="http://www.youtube.com/watch?v=PEuwPCZbaPg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=PEuwPCZbaPg&amp;feature=youtube_gdata_player</a>
G2b	<a href="http://www.youtube.com/watch?v=tmfDZvH155I&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=tmfDZvH155I&amp;feature=youtube_gdata_player</a>
G2c	<a href="http://www.youtube.com/watch?v=NKcXm1787ss&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=NKcXm1787ss&amp;feature=youtube_gdata_player</a>
G2d	<a href="http://www.youtube.com/watch?v=sqjKXQp3xc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=sqjKXQp3xc&amp;feature=youtube_gdata_player</a>
G2e	<a href="http://www.youtube.com/watch?v=zEopcNb7bBY&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=zEopcNb7bBY&amp;feature=youtube_gdata_player</a>
G2f	<a href="http://www.youtube.com/watch?v=UtsU2JMI13Q&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=UtsU2JMI13Q&amp;feature=youtube_gdata_player</a>
G2g	<a href="http://www.youtube.com/watch?v=H4g1x_Pjwr0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=H4g1x_Pjwr0&amp;feature=youtube_gdata_player</a>
G2h	<a href="http://www.youtube.com/watch?v=BZ0meNM-8SU&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=BZ0meNM-8SU&amp;feature=youtube_gdata_player</a>
G2k	<a href="http://www.youtube.com/watch?v=zZKr60-vlN0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=zZKr60-vlN0&amp;feature=youtube_gdata_player</a>

G2j	<a href="http://www.youtube.com/watch?v=Gs_8yOtBL4Q&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=Gs_8yOtBL4Q&amp;feature=youtube_gdata_player</a>
G3a	<a href="http://www.youtube.com/watch?v=Q2s_R_O0GRo&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=Q2s_R_O0GRo&amp;feature=youtube_gdata_player</a>
G3b	<a href="http://www.youtube.com/watch?v=KqyxPpV1mqY&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=KqyxPpV1mqY&amp;feature=youtube_gdata_player</a>
G3c	<a href="http://www.youtube.com/watch?v=_6F_uJZeLfM&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=_6F_uJZeLfM&amp;feature=youtube_gdata_player</a>
G3d	<a href="http://www.youtube.com/watch?v=jIAaPKiByiA&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=jIAaPKiByiA&amp;feature=youtube_gdata_player</a>
G3e	<a href="http://www.youtube.com/watch?v=XxuWMhQzjnU&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=XxuWMhQzjnU&amp;feature=youtube_gdata_player</a>
G3f	<a href="http://www.youtube.com/watch?v=T3R2dPfi15Q&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=T3R2dPfi15Q&amp;feature=youtube_gdata_player</a>
G3g	<a href="http://www.youtube.com/watch?v=mIKmI1Eqh5c&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=mIKmI1Eqh5c&amp;feature=youtube_gdata_player</a>
G3h	<a href="http://www.youtube.com/watch?v=ZstLbwHgvTE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=ZstLbwHgvTE&amp;feature=youtube_gdata_player</a>
G3j	<a href="http://www.youtube.com/watch?v=kcse-WWR2zA&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kcse-WWR2zA&amp;feature=youtube_gdata_player</a>
G4a	<a href="http://www.youtube.com/watch?v=Ak-qq-hyIq8&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=Ak-qq-hyIq8&amp;feature=youtube_gdata_player</a>
G4b	<a href="http://www.youtube.com/watch?v=6I0xqXYMtSQ&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=6I0xqXYMtSQ&amp;feature=youtube_gdata_player</a>
G4c	<a href="http://www.youtube.com/watch?v=qvjbcbk8nQ&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=qvjbcbk8nQ&amp;feature=youtube_gdata_player</a>
G4d	<a href="http://www.youtube.com/watch?v=y4U-LEE-p7s&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=y4U-LEE-p7s&amp;feature=youtube_gdata_player</a>
G4e	<a href="http://www.youtube.com/watch?v=kcYh94aEMLc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kcYh94aEMLc&amp;feature=youtube_gdata_player</a>
G4f	<a href="http://www.youtube.com/watch?v=E6gSsK_TcYs&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=E6gSsK_TcYs&amp;feature=youtube_gdata_player</a>
G4g	<a href="http://www.youtube.com/watch?v=8nDufEfilTc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=8nDufEfilTc&amp;feature=youtube_gdata_player</a>
G4h	<a href="http://www.youtube.com/watch?v=LmDXfMetWFE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=LmDXfMetWFE&amp;feature=youtube_gdata_player</a>
G4i	<a href="http://www.youtube.com/watch?v=HqmNFWtewfc&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=HqmNFWtewfc&amp;feature=youtube_gdata_player</a>
G5a	<a href="http://www.youtube.com/watch?v=7UTjyXfr7nE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=7UTjyXfr7nE&amp;feature=youtube_gdata_player</a>
G5b	<a href="http://www.youtube.com/watch?v=rFFqAdv96zQ&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=rFFqAdv96zQ&amp;feature=youtube_gdata_player</a>
G5c	<a href="http://www.youtube.com/watch?v=6ufGEGFYn5s&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=6ufGEGFYn5s&amp;feature=youtube_gdata_player</a>
G5d	<a href="http://www.youtube.com/watch?v=zRSSHscWFyw&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=zRSSHscWFyw&amp;feature=youtube_gdata_player</a>
G5e	<a href="http://www.youtube.com/watch?v=GyJFwqA0OUI&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=GyJFwqA0OUI&amp;feature=youtube_gdata_player</a>
G5f	<a href="http://www.youtube.com/watch?v=WBqQ0R114bg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=WBqQ0R114bg&amp;feature=youtube_gdata_player</a>
G5g	<a href="http://www.youtube.com/watch?v=kottUZtd860&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kottUZtd860&amp;feature=youtube_gdata_player</a>
G5h	<a href="http://www.youtube.com/watch?v=q1ZqRmjpaX0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=q1ZqRmjpaX0&amp;feature=youtube_gdata_player</a>
G6a	<a href="http://www.youtube.com/watch?v=cyRJCq-Qqzs&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=cyRJCq-Qqzs&amp;feature=youtube_gdata_player</a>
G6b	<a href="http://www.youtube.com/watch?v=kxJD55hHzG0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kxJD55hHzG0&amp;feature=youtube_gdata_player</a>
G6c	<a href="http://www.youtube.com/watch?v=cwUyVhkJyHM&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=cwUyVhkJyHM&amp;feature=youtube_gdata_player</a>
G6d	<a href="http://www.youtube.com/watch?v=3hSTyvaYxfM&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=3hSTyvaYxfM&amp;feature=youtube_gdata_player</a>
G6e	<a href="http://www.youtube.com/watch?v=IIUAFNGEXA0&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=IIUAFNGEXA0&amp;feature=youtube_gdata_player</a>
G6f	<a href="http://www.youtube.com/watch?v=kThwsSGQu7E&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=kThwsSGQu7E&amp;feature=youtube_gdata_player</a>
G7a	<a href="http://www.youtube.com/watch?v=PJKVsBXnJxs&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=PJKVsBXnJxs&amp;feature=youtube_gdata_player</a>
G7b	<a href="http://www.youtube.com/watch?v=VTfQMnzUZRQ&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=VTfQMnzUZRQ&amp;feature=youtube_gdata_player</a>
G7c	<a href="http://www.youtube.com/watch?v=Qs-X5HJWQQg&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=Qs-X5HJWQQg&amp;feature=youtube_gdata_player</a>
G7d	<a href="http://www.youtube.com/watch?v=JpH0eKjyXW4&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=JpH0eKjyXW4&amp;feature=youtube_gdata_player</a>
G7e	<a href="http://www.youtube.com/watch?v=OuCO4rOLNB8&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=OuCO4rOLNB8&amp;feature=youtube_gdata_player</a>
G7f	<a href="http://www.youtube.com/watch?v=KjcD0TICsxU&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=KjcD0TICsxU&amp;feature=youtube_gdata_player</a>
G8a	<a href="http://www.youtube.com/watch?v=U_p4XErjbro&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=U_p4XErjbro&amp;feature=youtube_gdata_player</a>
G8b	<a href="http://www.youtube.com/watch?v=AZwuVMAd4Fs&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=AZwuVMAd4Fs&amp;feature=youtube_gdata_player</a>
G8c	<a href="http://www.youtube.com/watch?v=QDbswW7yDPA&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=QDbswW7yDPA&amp;feature=youtube_gdata_player</a>
G8d	<a href="http://www.youtube.com/watch?v=EBXjs6cjLCE&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=EBXjs6cjLCE&amp;feature=youtube_gdata_player</a>
G8e	<a href="http://www.youtube.com/watch?v=mgOGiqaMw5w&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=mgOGiqaMw5w&amp;feature=youtube_gdata_player</a>
G9a	<a href="http://www.youtube.com/watch?v=wLINBkkPqXw&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=wLINBkkPqXw&amp;feature=youtube_gdata_player</a>
G9b	<a href="http://www.youtube.com/watch?v=BdG7PPPkfrw&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=BdG7PPPkfrw&amp;feature=youtube_gdata_player</a>
G9c	<a href="http://www.youtube.com/watch?v=ghVV1dHD59Y&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=ghVV1dHD59Y&amp;feature=youtube_gdata_player</a>
G9d	<a href="http://www.youtube.com/watch?v=6s1EPiVZR7o&amp;feature=youtube_gdata_player">http://www.youtube.com/watch?v=6s1EPiVZR7o&amp;feature=youtube_gdata_player</a>