

Forthcoming in: Kuzmin, Evgeny, ed. *Linguistic and cultural diversity in cyberspace: Proceedings of the 3rd International Conference (Yakutsk, Russian Federation, June 28 – July 3, 2014)*. Moscow: Interregional Library Cooperation Centre.

**Multilingualism and minority language use in the digital sphere: The digital use of language
as a new domain of language use**

Anna Fenyvesi

Associate Professor, Institute of English and American Studies,

University of Szeged, Hungary

Introduction

In the past quarter of a century, things digital have changed numerous aspects of our lives – and language use is no exception. With the use of email, the world wide web, mobile technologies, and digitally mediated ways of communication, a new domain of language use has entered the lives of most of us – namely, what I call “digital language use” below. A lot of it is oral, mediated by mobile phones and voice-over-IP (like Skype, for instance), but a great part of it is written and involves both reading and writing (such as emailing, texting, instant messaging, blogging etc.). In fact, it is estimated that through the use of these ways of digital communication, we read and write more today than before their advent (Baron 2008: 183). This makes especially written forms of digitally mediated communication a highly important new aspect of language use that should be the focus of concern for sociolinguists, educators working in bilingual education, and, indeed, all professionals working with bilingual minority language communities, be they editors or writers of digitally present newspapers, computational linguists and computer scientists working on computational language tools, or social scientists studying the role of language in various aspects of community life.

“Domains of language use”

The concept of the domain of language use has been widely used in the study of bi- and multilingual communities ever since Fishman (1972: 441) adopted Schmidt-Rohr’s (1932) idea of “elements of dominance configurations” as a theoretical concept defined “in terms of institutional contexts and their congruent behavioral co-occurrences. They attempt to summate the major clusters of interaction that occur in clusters of multilingual settings and involving clusters of interlocutors”. The five domains originally differentiated by Fishman are family, friendship, religion, education, and employment. This set of domains has curiously remained unchallenged ever since and is widely used by sociolinguists to this day to describe patterns of language use by

bilinguals in their respective languages, indicative of language maintenance or language shift when compared intergenerationally. When linguists investigating the bilingualism of speakers discuss other domains in their work, they usually do so without explicitly stating that they have expanded their number (cf. some recent examples of domains, in Grosjean 2010: 29, “parents, children, siblings, distant relatives, work, sports, religion, school, shopping, friends, going out, hobbies, and so on”; in Bever 2011: “public domain”; and in Öpengin 2012: 160, “economy”).

Digital language use

But whatever the total number and range of domains to be differentiated, it seems inevitable that the “digital domain”, i.e. the use of language in digitally mediated communication, should be regarded and recognized as a separate domain of language use, for a number of reasons.

Digital language use has become a prominently important aspect of language use: it encompasses various forms of both formal and informal communication (rather than just the latter), it includes genuinely new functions of language use (cf. blogging), and, as scarce results into it already indicate, it can present patterns of language use which are markedly different from all other (traditional) aspects of language use by bilinguals. For instance, Huber (2013) has shown that while first generation Canadian Hungarians use Hungarian more in the traditional domains of family, friendship, and religion than do their second-generation children, the latter far outperform their parents in the use of Hungarian in the digital domain (in emailing and using it on the internet) – demonstrating that digital language use can indeed become an important factor of language maintenance for the young, “digital native” generation. Basharina (2013) has also shown that the digital domain of language use can present a space for users of the minority language Sakha in Yakutia, Russia, where new forms of old genres as well as new genres of story telling can contribute to the strengthening of the minority language user community and its cultural and linguistic adaptation, modernization, and vitality.

In a paper meticulously supported by ample empirical data and mathematical calculations, Kornai (2013) has argued that digital language death will be the fate of a great number of the languages existing today – primarily those that exist as minority languages only – unless their speakers (and the professionals supporting them) succeed in meeting some all important criteria like having a community of digitally literate users and a Wikipedia in the language.

Support for digital language use in minority languages

The possibility of use of minority languages in the digital domain is, of course, dependent on a number of factors which range from the technical (the existence and availability of hardware and software), through the educational (literacy in the traditional sense and in matters digital) and

personal (the presence of digitally competent language users interested in using the language in this domain) to those of prestige (whether language users regard their minority language as “worthy” of the effort of using it digitally). All of these aspects present arenas where minority language users can be supported in their language use by members of their own community such as language activists and by outsider professionals: computer scientists and computational linguists working on language tools for minority languages, educators, linguists etc. The work of the Norwegian Giellatekno company is a case in point: its computer scientists have been developing and making available a wide range of language learning tools, bilingual dictionaries, morphological and syntactic analyzers, and games for Saami and other endangered minority Finno-Ugric languages (cf. <http://giellatekno.uit.no/>).

An example: The FinUgRevita project

The “Computational tools for the revitalization of endangered Finno-Ugric minority languages, FinUgRevita” project was created in 2013 with the aim to provide computational language tools for endangered indigenous Finno-Ugric languages such as Udmurt and Mansi in Russia and to assist the speakers of these languages in using the indigenous languages in the digital domain (<http://www.ieas-szeged.hu/finugrevita/>).

The project involves two teams – one of the University of Helsinki, Finland (Principal Investigator: Roman Yangarber), the other of the University of Szeged, Hungary (Principal Investigator: Anna Fenyvesi) – comprising Finno-Ugrist linguists, computational linguists, and sociolinguists, and is funded for the period of September 1, 2013 to August 31, 2017, by the Academy of Finland (AKA) and the Hungarian National Research Fund (OTKA).

The two languages the project focuses on so far, Udmurt and Mansi, are both endangered, according to UNESCO’s classification of endangered languages (UNESCO 2010) although to a different extent. Udmurt is a “somewhat” endangered language, with almost 60% (or about 300,000) of the 552,000 ethnic Udmurts speaking the language (cf. the figures of the 2010 Russian census), spoken in the Udmurt Republic, or Udmurtia, west of the Ural Mountains. Even though it has official status in Udmurtia, it has limited power and rights in the public sphere and is used mostly in the family domain and among friends. In Udmurtia it is present in the media, education and culture as well as has an internet presence (e.g. it is one of the three Finno-Ugric minority languages that VKontakte, “the Russian Facebook” social networking site can be used in). Mansi is a severely endangered language with less than 1,000 speakers (among the 12 thousand strong ethnic Mansi population), spoken in the Khanti-Mansi Autonomous Okrug (informally known as Yugra) in western Siberia, east of the Urals. It has no official status whatsoever even in the Okrug, and although it has some minor presence in the media, education and culture of Yugra, it is used

primarily in the family and friendship domains. Perhaps surprisingly for such a small language, it does have an internet presence: the bi-weekly newspaper Luima Seripos is also published online.

Sociolinguistically, the speaker communities of both languages have been undergoing language shift, that is, the expanse of the majority language Russian at the expense of the minority language in the speakers' lives, ever since their ancestors came to be under Russian domination in the 16th and 17th centuries, also experiencing forceful assimilation and Russification in Soviet times (Bakró-Nagy, forthcoming). The discovery of oil and gas in the 1970s in the regions where Mansi and Udmurt are spoken also led to the in-migration of workforce from outside, making the regions multilingual and the Mansi a minority even in their own district). For instance, the number of people professing to be of Udmurt ethnicity decreased from 640,000 in 2002 to 552,000 in 2010, while the proportion of speakers fell from 67% to 59% during the same time. And while the number of those declaring Mansi ethnicity increased in the same period, from 11,500 in 2002 to 12,300 in 2010, the proportion of speakers fell from 23% to just 7.65%.

The main aims of the project are the development of open source, freely accessible computational language tools: electronic dictionaries, morphological and syntactic analyzers, language games, as well as learning tools. Computational linguistic work on these tools has started and is in progress.

In addition to the computational linguistic work, two online surveys have been undertaken as a part of the project. One survey, launched in June 2014, aims to survey the use of Giellatekno's computational language tools for Saami, with the goal of analyzing users' feedback regarding their use of and satisfaction with these tools, both for the sake of the developer company and their continuing improvement of the tools and for the FinUgRevita project being able to benefit from the experiences of the user community regarding tools similar to our future tools.

The other survey which the researchers involved in the project are preparing at the time of the writing of the present report, August 2014, and are planning to launch in the fall of the same year is a sociolinguistic survey aimed at mapping out the digital language use of Udmurt and Mansi speakers. Specifically, through the survey sociolinguistic and language use information will be collected from speakers of Udmurt and of Mansi about what language(s) they use in various forms of digitally mediated communication, i.e. using mobile phones, emailing, surfing, chatting, blogging, commenting, using social media, producing internet content etc. With detailed information about when speakers use the minority language (Udmurt/Mansi), the majority language (Russian), and/or other languages (English, or other minority languages spoken in Russia), it is hoped that the project's investigators will gain invaluable insight into users' habits of language use, needs of computation language tools in the minority languages, and, in general, a better

understanding of language use patterns of speakers of endangered indigenous languages in the digital domain.

Conclusion

The digital domain, as I have argued above, has become an all important domain of language use by bi- and multilinguals, especially from the perspective of minority languages. Their support is essential if they are to be “digital survivors” (in terms of Kornai 2013), although the most important prerequisite of such survival is, probably, the determination on the part of the speakers of the language themselves to save them from language shift and/or digital death – something that no outsider professional can achieve, however determined and skilled they may be.

References:

- Bakró-Nagy, Marianne. Forthcoming. Uralic languages. *Revue Belge de Philologie et d'Histoire*.
- Baron, Naomi S. 2008. *Always on: Language in an online and mobile world*. Oxford: Oxford University Press.
- Basharina, Olga. 2013. *The extent of the Sakha language on the internet: Digital genres and language revitalization*. Paper presented at the MinorEuRus conference, Helsinki, Dec. 15-17, 2013.
- Bever, Olga. 2011. Multilingualism and language policy in post-Soviet Ukraine: English, Ukrainian and Russian in linguistic landscapes.
<http://www.irex.org/sites/default/files/BeverSTG2011%20Final%20report.pdf>; Date of access: August 28, 2014.
- Fishman, Joshua A. 1972. Domains and the relationship between micro- and macrosociolinguistics. In: Gumperz, John J., and Dell Hymes, eds. *Directions in sociolinguistics: The ethnography of communication*. Oxford: Blackwell, 435–453.
- Grosjean, François. 2010. *Bilingual life and reality*. Cambridge, MA: Harvard University Press.
- Huber 2013. *Intergenerational transmission of Hungarian as a heritage language in Canada: The macrosociolinguistics of the Hungarian community in Hamilton, Ontario*. Szeged: University of Szeged, BA thesis.
- Kornai, András. 2013. Digital language death. *PLoS ONE*, 8, e77056.
doi:10.1371/journal.pone.0077056.
- Öpengin, Ergin. 2012. Sociolinguistic situation of Kurdish in Turkey: Sociopolitical factors and language use. *International Journal of the Sociology of Language*, 217: 151–180.

Russian census 2010. http://www.perepis-2010.ru/results_of_the_census; Date of access: April 22, 2014.

Schmidt-Rohr, George. 1932. *Die Sprache als Bildnerin der Völker*. Munich.

UNESCO. 2010. *Atlas of the world's languages in danger*.

<http://www.unesco.org/new/en/culture/themes/endangered-languages/atlas-of-languages-in-danger/>; Date of access: April 19, 2014.