(In press, 2016). In C. Lee, *Multilingualism Online*. London: Routledge.

Epilogue to *Multilingualism Online*

Susan C. Herring
Indiana University, Bloomington USA

An Auto-Glosso-Technobiography

This book began with Carmen Lee's auto-technobiography. Similarly, I will start this epilogue by briefly relating my auto-glosso-technobiography, or the story of my interests in language and technology. Greek *glosso*- 'language' is infixed to limit the scope of the account, since some of my interests in technology-mediated communication are orthogonal to multilingualism per se. Nonetheless, multilingualism has been a persistent thread running through my research program for the past 15 years.

I decided to study linguistics many years ago because of an interest in foreign languages, and by the time I graduated with my doctorate, I had studied a dozen or so languages. When I began researching computer-mediated communication (CMC), however, it was in my native language, English, in which I felt most confident about my linguistic intuitions. CMC was already taking place in other languages at that time, for example, in Usenet newsgroups by speakers of other languages living in the western diaspora, but it was not until internet access spread around the globe starting in the mid-1990s that I became more than casually interested in multilingual CMC. During those years, I edited collections that included contributions on cross-cultural and non-English-language CMC, and I started encouraging my foreign students to study the uses of their native languages online.

By the end of the decade, I was growing concerned about the dominance of English in CMC research and on the internet as a whole. I expressed that concern publicly in a keynote lecture at the Cultural Attitudes towards Technology and Communication (CATaC) conference (Herring, 2002). That lecture was the first of what would become several broad efforts to pull together existing work on, and lay out agendas for, online multilingualism research, the best known of which is the 2007 collection I co-edited with Brenda Danet, *The Multilingual Internet: Language, Culture, and Communication Online*.

Having laid out those agendas, I felt that I should do my part to contribute to fulfilling them. My own contributions to online multilingualism fall into three streams that emerged roughly in the following order, but with temporal overlap: broad overviews and agenda-setting; case studies of a single language; and cross-language comparisons. The first stream was alluded to above. The second and third have been facilitated by the native or native-like language expertise of various co-authors, since by the turn of the millennium, I had moved from a linguistics department to a department of information science, where, unlike in linguistics, multiple co-authorship is the norm. This had the wonderful unintended benefit of expanding the scope of my CMC research beyond English. My co-authors and I have studied Thai, Lithuanian, Italian, and Bengali CMC. We have compared CMC in Lithuanian and Croatian; in English and Arabic; in English and Polish; and in English and Mandarin Chinese, as well as conducting larger cross-language comparisons, one involving

more than 50 languages (discussed further below). Like multilingual CMC itself, this research has involved multiple CMC modes, including chat, blogs, text messages posted to interactive television programs, Twitter, Wikipedia, e-commerce sites, newssites, university websites, social network sites, multiplayer online games, and video.[1] Yet this body of work, albeit diverse, is only a drop in the bucket, so numerous are the languages used online now, the contexts of their use, and the possible approaches that linguists could take to study them.

Trends in Online Multilingualism Research

The Danet and Herring (2007) collection represented online multilingualism research broadly, with no specific focus; we sought to include the best work that was available at the time, drawing from a variety of linguistic and methodological domains.[2] It included, for example, chapters on writing systems, politeness, language choice, global linguistic diversity, and challenges posed by CMC for machine translation. The chapter by the author of the present monograph, Carmen Lee, compared the linguistic features of email and ICQ produced by Cantonese-English bilinguals in Hong Kong. That study, like most of the others in the 2007 collection, analyzed language use data from log files.

The present monograph makes a more focused and coherent contribution by privileging analyses that drill down below the surface of linguistic expression to incorporate context, in keeping with contemporary trends in micro-sociolinguistics (Androutsopoulos, 2006). By providing numerous rich, contextualized vignettes of multilingual online language use, this book complements the log-based studies that came before, and advances the field of study overall. Sadly, Brenda Danet did not live to see the flowering of research on online multilingualism as described in this book. I expect that she would have appreciated it greatly, all the more so in that her own approach to research inclined towards the qualitative and the ethnographic.

Lee makes a persuasive case for a context-rich, micro-sociolinguistic approach to online multilingualism; indeed, hers is a distinctive voice that has emerged in recent years in association with this approach. Her book also addresses some broader-scope issues such as online linguistic diversity and metadiscourses about internet multilingualism. At the same time, important macro-sociolinguistic questions remain, and these should not be lost sight of in the turn towards contextualized specificity in sociolinguistic CMC research.

The macro-level questions I have in mind concern, for example, universal vs. language-specific practices, language contact and spread, the status of minority and majority languages in the linguistic ecology of the internet, and multilingualism on a regional and global scale, as well as longitudinal trends in the foregoing. Comparative language research, sometimes on a large scale, is required to address these phenomena. If a study aims to go beyond simply counting languages, moreover, it may be resource intensive, requiring manual analysis of potentially massive amounts of data. Some working knowledge of multiple languages is also helpful. It follows that this line of research is often best undertaken by multilingual teams of researchers, rather than by single scholars.

Cross-Language Trends in Online Multilingualism

For some years, I have been drawn to comparative studies in a quest for "big picture" understandings of online multilingualism. In what follows I discuss three projects in which my collaborators and I analyzed four or more languages, and which illustrate some of the macro-sociolinguistic issues mentioned above. In addition, each study sheds light on a larger trend (or trends) and thus, I suggest, provides a basis for making predictions about future online language use. The first study concerns language networks across blogs; the second is a longitudinal study of language choice on university websites; and the third focuses on special language varieties associated with playful online subcultures.

Language Networks on a Blog-Hosting Site

LiveJournal.com is a blog-hosting platform and one of the first social network sites (boyd & Ellison, 2007). Created in 1999 by a 19-year-old American programmer, it was sold in 2005 to Six Apart in San Francisco. By the time of our study in 2006 (published as Herring, Paolillo, et al., 2007), the platform was international in scope and multilingual, with templates available in 32 languages. The platform's slogan was: "LiveJournal.com is a place where you can share your thoughts with the world." We wanted to know what languages were actually used on LiveJournal.com, and how robust non-English language networks were. A multilingual team of eight researchers coded a random sample of LiveJournals and identified the most frequent non-English languages (English-language blogs accounted for 84% of the sample): Russian (11%), followed by Portuguese, Finnish, and Japanese (less than 1% each). We then constructed a seed sample of monolingual LiveJournals for each of the four languages, and crawled two degrees of 'friends' links from the seed languages to approximate social networks. The main languages used in the more than 5,000 blogs in these networks were manually identified,[3] counted, and visualized in social network graphs.

We found that the better-represented a language was on the site, the more 'friends' bloggers in that language had on average, and the more likely it was that friends' blogs would use the same language. The Russian network was the densest, followed by the Portuguese and the Finnish networks. The Japanese network was sparse – few Japanese blogs linked to other Japanese blogs; most linked to English blogs. In contrast, Russian blogs linked overwhelmingly to other Russian blogs. Thus a monolingual speaker of Russian (and to a lesser extent, Portuguese) could enjoy much content on LiveJournal.com, whereas Finnish and Japanese speakers would not have much blog content to read or respond to unless they knew English. We concluded that LiveJournal.com was not as multilingual in practice as it presented itself as being.

Moreover, it turned out that we had inadvertently identified evidence of a larger trend. One year after our study, Six Apart sold LiveJournal.com to the Russian media company SUP, due to the popularity of the platform with Russian users. When I sampled LiveJournals randomly again in November 2010, 56% were in Russian and only 34% were in English. A similar pattern of U.S. social media platforms being co-opted by speakers of other languages can be observed for other platforms. The social network site Orkut, for example, was created in 2004 by Google in California, but after a few years, Portuguese speakers were more numerous than English speakers on the site, and in 2008, Google announced

that Orkut would be managed and operated by Google Brazil. As of April 2010, 48% of Orkut's users were from Brazil, followed by India with 39% and the U.S. with only 2%. Another example is Friendster, launched in the U.S. in 2002. Its initial success in the U.S. was cut short due to dissatisfaction with the site's management, but its popularity continued to grow in Asia, especially in the Philippines, and in December 2009, Friendster was acquired by MOL Global, one of Asia's largest internet companies. Another social network site, Hi5, was created in San Francisco in 2003 and was popular in the U.S. for a while, but now it is frequented mostly by users from Central America, South America, and Thailand. Only 14% of its current users are from the U.S., mostly Hispanics and African Americans.

Based on these trends, we might venture a prediction that the U.S. social network site Facebook, which has been adopted by speakers of many other languages around the world, will someday shift its operations offshore and be replaced in the U.S. by a new, popular platform. In the meantime, Facebook remains highly popular in the U.S., while its global spread[4] continues apparently unchecked. Indeed, a number of other social network sites have been shut down due to their inability to compete with Facebook, including Orkut (in 2014) and Friendster (in 2015).

The global linguistic ecology of social network sites is a fascinating topic. Sites launched in countries other than the U.S. do not exhibit the same tendency to be taken over by speakers of other languages – why is that? It is an interesting question, moreover, how particular U.S. sites come to be popular with speakers of particular languages. Early adopters and influence leaders may be two favoring factors. (In the case of LiveJournal, several early adopters of the site were Russians studying in the U.S. who carried word of it back to Moscow; subsequently, LiveJournal was adopted by a high-profile Russian journalist who wrote about it on his website.) Targeted marketing and economic resources, as in the case of Facebook, play a role as well. In addressing questions such as these, linguistic researchers might profitably draw from both macro-sociolinguistics and political economics (e.g., Dor, 2004).

Changing Language Choices on University Websites

Institutions with a global audience are increasingly making versions of their websites available in different languages. Universities, in particular, are interested in marketing their institutions to foreign students and promoting international contact. Callahan and Herring (2012) investigated how often universities in different countries provide foreign-language web pages, what languages they use, and whether there has been any change over time in the languages used. More broadly, we were interested in the question of whether the internet facilitates the use of some languages (such as English) at the expense of others, and if so, which languages are favored or disfavored?

To address these questions, we manually analyzed the webpages of more than 1,100 universities in 57 countries at three points in time over a five-year period. Each website was coded for country, primary language, and secondary language(s) used. My co-author speaks Polish, Russian, and German, in addition to English, and I have some familiarity with

a number of languages. That said, our broad sample included languages and writing systems that neither of us knew; in such cases, we made use of cues available on the website, such as flag icons next to the language options, and Google Translate.

Perhaps unsurprisingly, given that English is the international lingua franca of higher education, we found that 72% of the countries had English as the primary or secondary language on a majority of their university websites, and English was used as a primary or secondary language to some degree in all but one of the 57 countries. Yet the overall degree of multilingualism was also rather high: 52% of the websites had at least one additional page in another language, 16% presented information in three or more languages, and several sites also provided an option for machine translation through Google Translate. Analogously, our longitudinal analysis found that English is expanding, but the use of other languages is expanding more. Overall, the number of bilingual and multilingual sites increased between 2006 and 2011. Countries with multilingual sites included Spain (Spanish, English, Catalan/Galician), Iran (Farsi, English, Arabic), Israel (Hebrew, English, Russian), and Japan (Japanese, English, Chinese).[5]

Dor (2004) predicted that the globalization of the internet would lead to market-driven 'global diglossia,' with English as the language of international communication and commerce among speakers of different languages, and local languages used for, e.g., marketing products and services to local communities. But the situation for university websites might better be characterized as 'dynamic multilingualism' – multilingual, in that more than two languages are involved, and dynamic, in that multilingualism increases as different language versions are added over time.

These findings touch on a number of broader issues. What factors contribute to university websites in some countries becoming more multilingual than in other countries? The most linguistically diverse websites in our global sample were in Australia and New Zealand and in countries in the European Union, where social and political realities favor cross-linguistic outreach. Universities in the former countries attract a broad spectrum of students from South and Southeast Asia, and countries of the European Union are participating increasingly in cooperative international scholarship programs. What do these findings say about trends in the online global linguistic ecology? Since pressures towards internationalism are currently affecting most countries on earth, we can extrapolate that the internet will become increasingly diverse. Scholars' early fears of English spreading via the internet to the detriment of other languages (cf. Herring, 2002) thus appear less and less likely to be realized: English is spreading (mainly as a lingua franca), but other languages are not being marginalized; the trend is quite the opposite.

How much can we extrapolate from university websites to other online genres, though? It seems that the purpose and audience of the websites make a difference. For example, a high level of "globalization" (translation into other languages) has been observed for websites in the electronics, telecommunication, entertainment, news, and travel industries, but less for websites in the banking, food and drink, insurance, and finance industries (Singh & Boughton, 2005). The latter provide services that are arguably oriented more toward local consumers within a single linguistic area than the former are, consistent with

Dor's (2004) global diglossia model.

Finally, I noted above that several websites in the Callahan and Herring (2012) study incorporated automated translation tools. It is not a stretch to predict that before long, visitors to any website[6] will be able to view its contents in an array of languages with the click of an icon. This technology is already available as a free Microsoft widget that can be installed on any website to translate it into more than 30 languages, including Haitian Creole, Hmong Daw, and Yucatec Mayan.[7] Widespread adoption of automated translation has the potential to alter many of linguists' basic assumptions about language choice in online communication. Choices will be tied less to the competence and identities of the users than in traditional offline communication, although perhaps the choices will be no less strategically motivated. This promises to be a richly rewarding area of study in future multilingualism research.

Cross-Cultural Comparison of Special Internet Language Varieties

The third study differs from the other two in that the units of analysis are much smaller (parts of words and sentences instead of webpages); nonetheless, it is a cross-linguistic study that I believe sheds light on macro-level issues. Several years ago I became interested in what I call special internet language varieties, or SILVs for short. A SILV, as defined by Herring, Kouper, et al. (2012),[8] is a highly playful, nonstandard variant of a standard language – more divergent from standard language norms and less comprehensible to non-initiates than typical "netspeak." Although SILVs can arise in different parts of the world and be based on languages that differ widely in their grammatical structure and writing systems, they had never before been studied from a cross-linguistic perspective.

In late 2011, I and four collaborators analyzed the linguistic characteristics and social contexts of four SILVs, each originating in a different culture and based on a different language: Leet Speak (U.S.), Padonki (Russia), Fakatsa (Israel), and Martian Language (Taiwan and China).[9] The research team included native speakers of Russian, Hebrew, and Mandarin Chinese, as well as a member with a background in computer science who was familiar with U.S. hacker culture. We asked: How similar are SILVs across cultures? If similarities are present, what accounts for them (contact across varieties, common generative linguistic principles, common social factors)?

Our data source was texts from the public internet contexts in which each SILV was used most often – bulletin boards, game chat logs, blogs, web forum posts and comments, etc., depending on the SILV. We categorized and compared the features of each language subsample at multiple linguistic levels: typographic, orthographic, phonological, lexical, morphological, syntactic, pragmatic, and rhetorical. With the exception of Padonki, the features most involved in SILV production were mostly typographic and orthographic. These types co-occurred so often that we coined a portmanteau term, *typthographic*, to refer to them.

Our comparison revealed a number of similarities across the SILVs but little evidence of cross-variety contact. Rather, the similarities appeared to derive from the exploitation of

common principles, such as substituting certain characters for other characters based on graphical or sound resemblance, combined with the metalinguistic awareness fostered by persistent textual CMC, which lends itself to language play. These principles might be considered candidates for typthographic universals; additional evidence for them can be found in several of the chapters in Danet and Herring (2007). Moreover, all four varieties arose in online sub-cultures where there was frequent in-group interaction, and they seemed to be motivated by a desire on the part of their users to make their writing distinctive (e.g., obscure, humorous, decorative). That is, SILV use signaled in-group identification.

Differences among the varieties appeared to be conditioned largely by the resources made available by the writing system of each base language (in the case of typthography) and the sociocultural context in which each SILV arose. Leet users were primarily young and male, and their subculture valued hacking and computer skills. Padonki users were educated males in their late 20s and 30s with antisocial, contrarian values. Fakatsa users were preteen girls, and the subculture was focused on cuteness, femininity, and perfection.[10] The use of Martian Language was also associated with female more than male users and expressed ironic cuteness, especially in regard to romantic love. These values were reflected in language use in the SILVs at all levels, for example in the affixation of nonsense syllables to Hebrew words in Fakatsa to represent a cute, childish style of speech, or the proliferation of creative variants of Russian profanities in Padonki.

In addition to suggesting potential universals of creative online typthography, the SILV study sheds light on how linguistic innovations arise and spread. Some features of the four SILVs have made their way into general internet language usage; for example, Leet n00b ('newbie') and w00t (an expression of excitement) are part of English netspeak. Some SILV expressions have also percolated up into offline use, e.g., on billboards, t-shirts, record albums, and book covers, where their usage is mostly tongue-in-cheek, to index hipness and internet savvy. By the time we observed such uses, however, the SILVs themselves had already become passé, much like youth slang when it is adopted by adults. As with slang and other language fads, SILVs are transitional phenomena; at the time we collected our data, most of the SILVs had already peaked in usage.

We might predict that new SILVs will arise to take the place of old ones in these languages, and that SILVs will be found in other languages, given that online subcultures arise in many linguistic contexts, and that CMC promotes metalinguistic reflection and language play. There are also other SILV-like varieties, such as offline special language varieties that have migrated online (e.g., German rapper language; Androutsopoulos, 2007) and graphics-based forms of subcultural communication. Indeed, SILVs today are more likely to incorporate multimodal elements than not. The phenomena of LOLspeak – e.g., 'I can has cheezburger,' superimposed over photos of cats – and Doge Speak – e.g., 'many happy,' 'very love' superimposed over images of a shiba inu dog[11] – are arguably graphical SILVs. I believe that an understanding of text-based SILVs can provide a useful comparative lens through which to identify and study these phenomena. More generally, how special language varieties online arise, spread, and die out; what, if anything, is universal versus

what is particular to each variety; and what role the internet plays in these processes, constitute questions that should be of interest to students of internet multilingualism.

The three studies described in the subsections above illustrate very different methodological approaches to cross-cultural comparison: social network analysis, content analysis, and close qualitative analysis of linguistic features combined with sociocultural interpretation. Through participating in these studies, I have come to understand multilingualism online in new and, I believe, ultimately complementary ways, despite how disparate the studies may appear. All are aspects of the "big picture" of online language use, arrived at by comparing across multiple cultures or subcultures in order to arrive at more general insights, which in turn help to predict future trends.

Closing Thoughts

With the publication of *Multilingualism Online*, the field of internet multilingualism research has attained a new level of maturity. In the future, we can expect to see increased specialization and development of sub-areas within this field. Areas that could give rise to future book-length works include CMC in lesser-studied and minority languages; ethnographic studies of language use and language choice; cross-cultural communication and nonnative language use; multilingualism in contemporary social media; and macro-level studies of degrees of, and trends in, internet multilingualism.

Meanwhile, new technologies raise new questions and can be expected to lead to new areas of inquiry. What will the effects of widespread online machine translation be, for example, on cross-cultural communication? When access to other languages is a click away, for what purposes will it be used, and to what effect? It seems to me that the future holds some rather exciting prospects.

Notes

1. A list of my publications, with links, can be found at: http://info.ils.indiana.edu/~herring/pubs.html.
2. An earlier version of the collection, containing a subset of the articles in Danet and Herring (2007), was published in 2003 as a special issue of the *Journal of Computer-Mediated Communication*.
3. Our research team included a native speaker of Russian and a speaker of Portuguese, and I have studied some Japanese. Finnish was identified, after some online research, through the use of characteristic words and letter sequences, especially those involving diacritics.
4. As of January 2016, it was the most popular social network site in 129 out of 137 countries worldwide (http://vincos.it/world-map-of-social-networks/, accessed March 21, 2016).
5. This is a simplified summary of a complex global picture. For further details, see Callahan and Herring (2012).

6. Any website that uses a standard language variety, that is. Nonstandard and creative language use may continue to pose challenges for machine translation (see Climent, et al. 2007).
7. https://www.microsoft.com/Web/solutions/mstranslator.aspx, accessed March 20, 2016.
8. This research is unpublished, but a video of a 2012 presentation is available that includes further information and examples of language use in SILVs.
9. The names of the varieties are endonyms, that is, those used in the subcultures themselves. The name 'Leet' (also written as 1337, 133+, etc.) is derived from the word 'elite.' 'Padonki' is a deliberate misspelling of the Russian word for 'scoundrels' or 'scumbags.' Padonki is also known as Padonkaffsky jargon. 'Fakatsa' is an acronym in Hebrew that translates roughly as 'a shallow, stupid, noisy girl.' 'Martian Language' is so named because it is thought to be complex and unreadable.
10. See Vaisman (2014) for a more in-depth analysis of Fakatsa.
11. For more on these varieties, see Gawne and Vaughan (2011) and McCullough (2014).

References

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics, 10*(4), 419-438.

Androutsopoulos, J. (2007). Style online: Doing hip-hop on the German-speaking Web. In P. Auer (Ed.), *Style and social identities: Alternative approaches to linguistic heterogeneity* (pp. 279–317). Berlin/New York: de Gruyter.

boyd, d., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230.

Callahan, E., & Herring, S. C. (2012). Language choice on university websites: Longitudinal trends. *International Journal of Communication, 6*, 322-355.

Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M., & Vallmanya, L. (2003). Bilingual newsgroups in Catalonia: A challenge for machine translation. *Journal of Computer-Mediated Communication, 9*(November). http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2003.tb00360.x/full

Danet, B., & Herring, S. C., Eds. (2007). *The multilingual Internet: Language, culture, and communication online.* New York: Oxford University Press.

Dor, D. (2004). From Englishization to imposed multilingualism: Globalization, the Internet, and the political economy of the linguistic code. *Public Culture, 16*(1), 97-118 .

Gawne, L., & Vaughan, J. (2011). I can haz language play: The construction of language and identity in LOLspeak. In *Proceedings of the 42nd Australian Linguistic Society Conference.* https://digitalcollections.anu.edu.au/bitstream/1885/9398/5/Gawne_ICanHaz2012.pdf

Herring, S. C. (2002). *The language of the Internet: English dominance or heteroglossia?* Keynote talk, Conference on Cultural Attitudes towards Technology and Communication (CATaC), Université de Montréal, July 13.

Herring, S. C., Kouper, I., Kutz, D., Vaisman, & Zhang, G. (2012). Linguistic creativity online: A cross-cultural study of special Internet language varieties. Pragmatics Festival, Indiana University, Bloomington. Video presentation available at http://vimeo.com/40743805

Herring, S. C., Paolillo, J. C., Ramos Vielba, I., Kouper, I., Wright, E., Stoerger, S., Scheidt, L. A., & Clark, B. (2007). Language networks on LiveJournal. *Proceedings of the Fortieth Hawaii International Conference on System Sciences.* Los Alamitos, CA: IEEE Press. http://ella.slis.indiana.edu/~herring/hicss07.pdf

McCulloch, G. (2014). A linguist explains the grammar of Doge. Wow. *The Toast,* February 6. http://the-toast.net/2014/02/06/linguist-explains-grammar-doge-wow/

Singh, N., & Boughton, P. D. (2005). Measuring web site globalization: A cross-sectional country and industry level analysis. *Journal of Website Promotion, 1*(3), 3–20.

Vaisman, C. (2014). Beautiful script, cute spelling and glamorous words: Doing girlhood through language playfulness on Israeli blogs. *Language & Communication*, *34,* 69–80.