The Construction of the Multilingual Internet: Unicode, Hebrew and Globalization

Dr. Nicholas A. John

nicholas.john@mail.huji.ac.il

Abstract

This paper examines the technologies that enable the representation of Hebrew on websites. Hebrew is written from right to left and in non-Latin characters, issues shared by a number of languages which seem to be converging on a shared solution—Unicode. Regarding the case of Hebrew, I show how competing solutions have given way to one dominant technology. I link processes in the Israeli context with broader questions about the 'multilingual internet,' asking whether the commonly accepted solution for representing non-Latin texts on computer screens is an instance of cultural imperialism and convergence around a western artifact. It is argued that while minority languages are given an online voice by Unicode, the context is still one of western power.

Keywords: multilingual internet, Unicode, globalization, Hebrew, infrastructure

The Construction of the Multilingual Internet: Unicode, Hebrew and Globalization

Introduction

In January 2010, The Internet Corporation for Assigned Names and Numbers (ICANN), the closest the internet has to a regulatory body, declared that it had approved four internationalized country code top-level domains (IDN ccTLDs). This meant that websites registered in Egypt, Saudi Arabia, the United Arab Emirates or the Russian Federation no longer had to end in .eg, .sa, .ae or .ru respectively, but rather could take on a suffix that was written in the country's native script. Given that Internationalized Domain Names (IDNs) had already been implemented in 2003, this declaration paved the way for certain URLs to be written entirely in non-Latin scripts, a move that has been heralded, not least by ICANN itself, as making the internet accessible to many more people around the world. In theory, people accessing sites registered in those four countries[1] can now surf the web while typing only in the script of their native language. This undoubtedly significant development can be seen as the culmination (for now, at least) of a long-standing effort to make the internet multilingual.

In this paper I examine two of the technological building blocks of this recent development, and I do so in the context of a specific, but generalizable, problem that faced a small group of web designers in the early 1990s: how to make Hebrew-language websites. The primary aim of this 'infrastructural inversion' (Bowker & Star, 1999) is to redress a certain technological blindness that is prevalent in the literature on the multilingual internet. The main reason for inverting an infrastructure is given by Star, who says, 'study a city and neglect its sewers […] and you miss essential aspects of distributional justice' (Star, 1999, p. 379).[2] Given the disagreements among theorists of globalization, and given the debates over 'language endangerment' in the context of the internet among sociolinguists, there is good *a priori* reason

to assume that carrying out an infrastructural inversion of the technologies behind the multilingual internet will indeed contribute to our understanding of 'aspects of distributional justice'. Moreover, if the results of the infrastructural inversion shed light on these debates—as I believe is the case here—then its deployment is doubly justified.

This suggests that the results of the infrastructural inversion—presented in the first part of this article—will contribute to our understanding of the phenomenon at hand, in this case, the multilingual internet, and specifically its association with questions of globalization and the discourse of language endangerment—presented in the second part of the paper. In other words, by focusing on the *technologies* of the multilingual internet—and especially Unicode—this article seeks to open up a new channel for debate in computer-mediated communication.

In theory, any technology or artifact whatsoever is amenable to the kind of deconstruction proposed by Science and Technology Studies (STS)—and the method of infrastructural inversion is certainly part of STS—and case studies in the field are extremely varied.[3] In such a context, it would seem acceptable to research any technology simply because it is out there. However, beyond 'being there' as a technological artifact, I believe that a study of how Hebrew is dealt with on the internet is of general interest for two main reasons. Firstly, settling these issues—or bringing about their closure (Misa, 1992)—has been an important part of the infrastructural institutionalization of the internet in Israel. That is, according to interviewees, the difficulties— now resolved—of publishing Hebrew-language content that could be read by people around the world (and not just in Israel) impeded the growth of the internet in Israel. Indeed, according to Doron Shikmoni, manager of the .il top level domain, the technologies discussed below constituted a 'tremendous scene changer for Hebrew content on the internet.'[4] In a similar vein, writing in 1993 about Japanese exclusionism on the internet, Jeffrey Shapard argued:

> Narrow vision, one-byte seven-bit ASCII biases, the assumptions about character coding that arise from them, inadequate international standards, and local solutions that disregard what international standards there are and that pay no heed to the ramifications for others—all these are serious related problems that inhibit, rather than enhance, increased connectivity and communication (Shapard, 1993, p. 256).

Moreover, given that developments in Israel were tightly tied in to global processes, as I shall explain below, and given that Hebrew is not the only language that throws up special complications in computing, one would expect to find that parallel controversies have also been part of the institutionalization of the internet in other countries. Indeed, the little that has been written about Unicode and directionality from a social sciences perspective suggests that this is the case (see, for instance, Jordan, 2002; Myers, 1995).

This leads us into the second and related reason for taking an interest in this subject—the so-called 'multilingual internet.' As the internet was emerging as a global phenomenon, commentators often noted that it appeared to be a primarily English-language domain (Crystal, 1997, 2001). Some critics even saw this as part of an ongoing colonialist attempt to dominate peripheral, non-English language speaking parts of the world—what Charles Ess called 'computer-mediated colonization' (Ess, 2006; Phillipson, 1992, p. 53). In the early 1990s, when the proportion of English-language websites was extremely high, concerns about 'language death' and the ascendancy of English might not have seemed too far off the mark (Brenzinger, 1992; Skutnabb-Kangas, 2000). Not only was the internet in English, but, as I explain below, the Latin alphabet seemed hard-wired into its very architecture.

From the outset, then, there were ideological concerns about the dominance of English in the internet that were being vocalized by high level global institutions, particularly UNESCO.[5] But while the effect of the internet on the world's linguistic diversity is an issue to which I shall return, feeding in as it does to key issues in theories of globalization, I am particularly interested

in the *technology* behind the multilingual internet, and hope to show that this frequently neglected aspect actually has the potential to shed light on the relevant theoretical questions. Thus, while some researchers have asked how people use the internet in a language that is not their mother tongue (Danet & Herring, 2003; Kelly-Holmes, 2004; Koutsogiannis & Mitsikopoulou, 2003; Warschauer, el Said, & Zohry, 2002), this paper, through the method of infrastructural inversion, asks what makes it possible for them to use the internet to communicate in their mother tongue at all, especially when that language is written in a different script from and in a different direction to English.[6] More bluntly put, only three of the articles in Danet and Herring's Special Issue of the Journal of Computer-Mediated Communication on the multilingual internet (volume 9, issue 1, 2003: see Danet & Herring, 2003; Koutsogiannis & Mitsikopoulou, 2003; Palfreyman & al Khalil, 2003), and only one article in a similar Special Issue of the Journal of Sociolinguistics (volume 10, issue 4: see Androutsopoulos, 2006) mention Unicode, and even then they do so only in passing and without seriously interrogating the technological backdrop to the phenomena they analyze.

In what follows I shall outline the problems of working with Hebrew on the internet before presenting what has become the consensual way for dealing with them. Both the problems and the solution have two aspects, one more specific to Hebrew (directionality; though it applies equally to Arabic), and one more generally applicable to non-Latin scripts (encoding). After explaining the problems and their solution, I return to the problematic of the multilingual internet. Towards the end of the article, I propose that the way that different languages have come to be accommodated in the internet indicates a seemingly paradoxical process of localization through globalization, or heterogeneity through homogeneity and that this process must be understood in terms of the expansion of western socio-economic global influence.

Background: the problem of Hebrew and the internet

The problems of working with Hebrew in the internet are a subset of the problems of working with Hebrew in computing more generally on the one hand, and of the problems of working with non-Latin and right-to-left scripts in the internet on the other. The technological history of Hebrew in computers is long and involved, most of which lies beyond the scope of this study. However, the common themes running through it concern two simple facts about Hebrew: firstly, it is written in the Hebrew alphabet; and secondly, it is written from right to left.

Encoding

The first problem—that the Hebrew alphabet is a non-Latin alphabet—is known as the problem of *encoding*. I shall come to the historical source of this problem presently, but first we need to understand exactly what the problem is. When computers store letters, they encode them into numbers (which are rendered in binary form). If another computer wants to put those letters on its screen (as is the case whenever I read any document on a computer), it converts the numbers back into letters. It does this by consulting a map, which tells it, for example, that the code number 97 represents the letter 'a.' For many years, the dominant map was known as ASCII (American Standard Code for Information Interchange). This map was constructed in seven-bit code, meaning that it had room for 128 characters, which was plenty for anyone wishing to use a computer to write text in English.[7] However, this caused a problem for languages with extra letters, symbols or accents. How were Scandinavian languages, or even French and German, to cope?

Fortunately, it was not too difficult to turn the seven-bit ASCII code into an eight-bit code,[8] thus doubling the space available for new letters and characters. Different countries began exploiting the new empty spaces by filling them with their own specific characters. This created

a situation whereby each country had its own code sheet, where positions 0-127 reproduced the regular ASCII encoding, but where positions 128-255 were occupied by their own alphabet, accented letters, and so on.

This meant that a computer could no longer just consult one single map in order to work out which letters it should be showing on the screen. It must also make sure that it was looking at the correct code sheet, of which there were now a large number (up to 16). For instance, the code number 240 could mean the Greek letter π (*pi*), or the Hebrew letter נ (*nun*), depending on which code sheet you were looking at. This implies that if one wished to publish an internet site, one had to add something to the code of that site that told other computers which map to consult. Without directing the computer to the correct map, or, in other words, without telling it which encoding was used in writing the original document, an internet surfer might receive a page of gibberish.

This system—whereby different countries and alphabets had their own specific code sheet—bore other potential problems: for instance, a document encoded in Hebrew would not be able to include Greek letters—you could not refer to both maps at the same time.[9] In addition, if you did not have the necessary code sheet installed on your computer, then you simply would not be able to read documents encoded according to that sheet.[10] While not a problem that faces Hebrew, it should be mentioned at this juncture that there are of course languages that have far more than 127 characters (the 'vacant,' second half of the eight-bit code sheet). For such languages (especially Asian languages, such as Chinese, Japanese and Korean), which might have thousands of different characters, the ad hoc eight-bit method adopted by European countries for encoding their special characters is obviously inadequate.

The first problem, then, is one of encoding. The tiny ASCII map quickly turned out to be inadequate with regard to other languages' letters and characters, resulting in the ad hoc creation of a large number of language-specific code sheets, a solution which in itself created new problems.

Directionality

The second problem is that of *directionality*, and is much more specific to Hebrew (and Arabic, of course). Computers were mainly developed in the United States, and just as this explains why ASCII was considered a good encoding system, it also explains why computer code is written from left to right. However, we can see how this creates problems for Hebrew if we consider internet pages. When we surf the net, our browser reads a source page and does what that page tells it to do (for instance, put this picture here, make this word link to another site, and so on). This source page, which includes the text that we see on our screen, is written from left to right. This creates special problems for languages that are written from right to left. Specifically, it poses the question of how to write them in the source page.

Less significant variations notwithstanding, there have been two main ways of working with the directionality of Hebrew on the internet. The first method is known as *visual Hebrew*, and involves writing the Hebrew in the source page back to front so that the browser will display it the right way round. This method was dominant for the first few years of the internet, and received governmental backing in Israel in June 1997, which was not formally rescinded until 2003.[11] The second method is known as *logical Hebrew*, and allows Hebrew to be written from right to left in the source page. It is this method that has become the standard.

Experts have pointed out the technical disadvantages of visual Hebrew as compared to logical Hebrew. They note that the kind of line breaks that word processors automatically put in

text that we type do not work with visual Hebrew. Instead, one has to insert 'hard' line breaks. If a surfer reading such a site then changes the size of either the font or the browser's window, this can have the undesirable consequence of jumbling up the order of the lines and rendering the text quite illegible. More significantly, however, they note that creating visual web pages is more costly than creating logical ones, as some kind of flipping mechanism is required to turn the text back to front before it can be put in the source page. Similarly, copying text from a visual Hebrew web page into another program, such as a word processor, requires the same flipping process, otherwise the copied text appears back to front (see, for instance, Dagan, n.d.).

Given that logical Hebrew predates the popularity of the internet, these serious problems with visual Hebrew raise the question of how it remained dominant for so long. If, as is common knowledge among programmers and web designers in Israel,[12] logical Hebrew is so much better than visual Hebrew, why did it take so long for it to establish itself as the standard? By way of a partial answer we can offer Wajcman's comments made against technological determinism: 'it is not necessarily technical efficiency but rather the contingencies of socio-technical circumstances and the play of institutional interests that favour one technology over another' (Wajcman, 2002, p. 352). It is to those socio-technical circumstances that we shall return.

The two-fold solution

Nowadays, there is little controversy over how to deal with Hebrew on the internet. The consensual solution has two institutionally interrelated but technologically distinct parts, each of which deals with a different aspect of the problems presented above. As I shall explain below, the solution involves using Unicode to solve the problem of encoding and logical Hebrew to solve the problem of directionality.

Directionality

The recognized solution for working with the problem of directionality in Hebrew websites is to build them using logical Hebrew. As mentioned, this is now the standard used in government websites, and, in addition, it was incorporated into HTML 4.01 as far back as 1999.[13]

The dominance of logical Hebrew nowadays has a lot to do with Microsoft. Microsoft wanted to enter the Middle Eastern market and quickly learnt that there were a number of ways of dealing with Hebrew (and Arabic) directionality issues. It understandably wanted one standard to be agreed upon that it could subsequently integrate into its operating system, but expressed no opinion as to which standard that should be. Indeed, this was Microsoft's policy in every country it entered. The standard that was decided upon in Israel was that of logical Hebrew, which was subsequently integrated within Windows. The fact that around 90% of the world's desktop computers have some version of Windows installed is part of the explanation for the dominance of logical Hebrew as the consensual standard. In other words, while Microsoft may have brought the issue to a head, it did not offer *a priori* support to one standard over another.[14]

Moreover, as powerful as Microsoft has become, its success has depended on the personal computer (PC) replacing mainframe computer systems. This suggests that the dominance of logical Hebrew can be partly explained in terms of actor-network theory (ANT), and particularly its recognition that non-human actors, termed actants, may play as important a role in constituting technology as human ones. As John Law puts it, a successful account of a technology is one that stresses 'the heterogeneity of the elements involved in technological problem solving' (Law, 1987). ANT also teaches us to view technologies as contingent ways of dividing up tasks between humans and objects (Latour, 1992). Accordingly, it seems appropriate to introduce into this analysis two objects—the dumb terminal of a mainframe computer system, and the personal computer—a move which requires a small amount of explication.

Simply put, representing visually stored text on a screen demands much less computer power than representing logically stored text. Indeed, dumb terminals—display monitors attached to mainframes with no processing capabilities of their own—were barely capable of doing so. Visual Hebrew was an apt solution for the world of mainframes. Personal computers with their own microprocessors, however, offered much more computing power and were more than capable of putting logical Hebrew on screen. Because using logical Hebrew saves time at the programming end—it cuts out the 'flipping' stage, because you do not need to turn the Hebrew back to front—programmers preferred it.[15] In terms taken from ANT, therefore, we can understand the shift from mainframes to PCs as also involving the redistribution of a certain task, namely, 'flipping' Hebrew text: the dumb terminal was not able to do so, and so a human had to do it; the PC, however, is able to take on that task by itself.

This, however, merely begs another question to do with the timing of the dominance of logical Hebrew in the internet: if 'everyone knew' that logical Hebrew was better, and if Microsoft started integrating it into its operating system from the early 1990s, how do we explain the Israeli government's decision in 1997 that all government sites must be written in visual Hebrew? Why was logical Hebrew not adopted for use on the internet at this stage?

Again, the answer does not lie with the technological superiority of one solution over another, but rather with a particular aspect of computing at the time, namely, the political economy of web browsers. In 1997, the most popular internet browser by far was Netscape Navigator, a browser that did not support logical Hebrew. In other words, sites written in logical Hebrew would simply not be viewable in the browser that most people were using. Netscape Navigator did support visual Hebrew, however.[16] Therefore, if you wanted to write a website that would be accessible to the largest number of people—users of Netscape Navigator—you would

do so using visual Hebrew. Netscape Navigator only introduced support for logical Hebrew in version 6.1, which was released in August 2001. By then, however, Microsoft's own browser, Internet Explorer, had attained supremacy, and it, of course, had full support for logical Hebrew. Previously, then, there had been a trade-off between ease of website design and number of potential surfers to that site. With the emergence of Internet Explorer, there was no longer any need to make that trade-off.

In summary, for quite understandable technical reasons, programmers and web designers have long preferred to produce sites using logical Hebrew. However, the dominance of that standard has more to do with the shift from the dumb terminal to the PC, Microsoft's interest in the late 1980s and early 1990s in global expansion, and the success of that company's browser, at the expense of Netscape Navigator, than it does with its inherent technological qualities.

Encoding

The consensual solution to the problem of encoding has been provided by the Unicode Consortium, whose website declares: 'Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language.'[17] In other words, instead of each set of scripts and alphabets requiring their own code sheet, Unicode provides one huge code sheet for all of them. It offers a standardized way of encoding all documents in all languages. As Gillam writes in his guidebook on the subject, Unicode solves the problem of encoding

> by providing a unified representation for the characters in the various written languages. By providing a unique bit pattern for every single character, you eliminate the problem of having to keep track of which of many different characters this specific instance of a particular bit pattern is meant to represent (Gillam, 2003, p. 7).

Put even more simply, Unicode makes it impossible that the same number might refer to more than one character, as we saw with the example of code number 240 being able to represent

both the Greek letter *pi* and the Hebrew letter *nun*. Unicode also incorporates logical Hebrew within its standards, thereby providing a comprehensive solution not only for Hebrew, but also for Arabic (another script written from right to left), Greek, Russian, Swedish, Japanese, Swahili, and, so says the Unicode Consortium, every language on the planet.

Unicode has been widely implemented in websites. In January 2010 Google reported that 50% of web pages were encoded in Unicode,[18] and today over 70% of the web's 10,000 most visited sites use Unicode.[19] It is also used in the Java programming language (which, among others, is used to program web applets, which in turn include Flash movies and browser games). In addition, it is built in to XML,[20] which lies at the foundation of Microsoft Office and Apple's iWork software. Google is another technology giant that has adopted Unicode, where this has enabled it to roll out services in languages other than English. For example, having been released in April 2004 in English, Gmail was made available in Hebrew in May 2006;[21] in November 2008 Google Calendar was offered in Hebrew (and Arabic);[22] in August 2009 Google started offering search results in Hebrew and Arabic on mobile devices;[23] and in September 2009 Google Sites started operating in Hebrew and Arabic.[24] Google itself has said that the reason for its adoption of Unicode 'is to give everyone using Google the information they want, wherever they are, in whatever language they speak'.[25]

This is not to say that the user experience for users of non-Latin scripts is immediately equivalent to that of English speakers. Shortly after the release of the iPad in April 2010, for instance, Elizabeth Pyatt, wrote a blog entry about the state of play of Unicode in those new devices. Her overall impression was that there was still work to be done, but she acknowledged that improvements appeared to be on the way (Pyatt, 2010). For example, the ability to input Arabic was not supported on first generation iPads immediately upon their release, though one

could buy an Arabic keyboard from the App Store, and an official update from Apple in November 2010 included a native Arabic keyboard. One blogger's first impression was that, in general, the multilingual capabilities of the iPad lagged behind those of the iPhone (Gewecke, 2010). However, this was attributed to Apple's wanting to release the product quickly rather than lacking the required technological know-how for dealing with Arabic. Non-English speaking users of Android mobile devices have also experienced certain difficulties: while the Android operating system uses Unicode, not all fonts are preinstalled on all devices, meaning that not all languages are equally accessible on them.[26] As noted by John Paolillo, Unicode encoding 'causes text in a non-roman script to require two to three times more space than comparable text in a roman script' (Paolillo, 2005b, p. 47), which costs might be 'enough of a penalty to discourage use of Unicode in some contexts' (p.73). So readers and writers of non-Latin scripts would still appear to be at a disadvantage when it comes to using the newest communication devices. However, the adoption of Unicode by the world's largest technology companies means that this disadvantage is both smaller and much shorter-lived than that faced by the community of Hebrew-language webmasters in the early- to mid-1990s.

The appearance and increasing popularity of Unicode have been met with widespread (but not universal) approval across the computing industry. Indeed, articles in newspapers and trade magazines greeted the emergence of Unicode extremely warmly (for instance, Ellsworth, 1991; Johnston, 1991; Schofield, 1993). For example, one article in a trade magazine talks about how Unicode is 'bringing more of the world in' (Hoffman, 2000). The Unicode Consortium itself claims that '[t]he emergence of the Unicode Standard, and the availability of tools supporting it, are among the most significant recent global software technology trends' (Unicode Consortium, 2004). In a book on Unicode for programmers, Tony Graham deploys Victor Hugo to describe it

as 'an idea whose time has come' (T. Graham, 2000, p. 3), and says that for him, a computer programmer, it is a 'dream come true' (p. x).

However, there are three elements of the discourse surrounding Unicode which call for critical attention: the first is the determinist tendency to represent Unicode as the next stage in a natural process; the second is the existence of alternatives to the dominant technology; and the third is a discussion of the technology's purported 'impacts.'

The first element—the tendency to see Unicode as the outcome of an almost natural process of development—can be seen, for instance, in an article by a programmer and writer who terms Unicode 'the next evolution' for alphabets (Celko, 2003). In their books, both Graham and Gillam present Unicode as the natural and obvious solution to a problematic state of affairs. In keeping with technological determinist ways of thinking, they talk as if Unicode was out there waiting to be discovered.

One way of presenting an alternative to this discourse is to use the concept of 'relevant social groups,' and particularly Pinch and Bijker's assertion that 'a problem is only defined as such, when there is a social group for which it constitutes a "problem"' (Pinch & Bijker, 1987, p. 414). To this we could add that the relevant social group also needs to have the means of formulating and disseminating a solution. Unicode expert Tony Graham, for instance, offers an explanation of the origins of Unicode:

> The Unicode effort was born out of frustration by software manufacturers with the fragmented, complicated, and contradictory character encodings in use around the world. […] This meant that the 'other language' versions of software had to be significantly changed internally because of the different character handling requirements, which resulted in delays (T. Graham, 2000, p. x).

This is borne out by other histories of Unicode, which locate the origins of its invention in the difficulties of rendering a piece of English-language software into an Asian language. And

indeed, the Unicode project was initiated by programmers at Xerox and Apple.[27] The most obvious relevant social group, then, is clearly that of computer programmers working in multilingual environments.

Their superiors also had a clear interest in Unicode insofar as it can dramatically reduce the time it takes to turn software from English into another language. For instance, the American version of Microsoft Windows 3.0 was released in May 1990, but the Japanese version was shipped only 18 months later. Partly as a result of the technology being discussed here, the English and Japanese versions of Windows 2000 were released on the same date. Unicode is thus represented as aiding in computer companies' internationalization efforts.

However, another social group can be identified which is enjoying the spread of Unicode, but which could not have developed it itself. This group is made up of librarians, who have long been working with large numbers of scripts. The implementation of Unicode means that American students of Chinese literature, for example, can search for titles and authors using Chinese characters without having to guess at how they might have been transliterated into Latin text. It also means that libraries with multilingual holdings can maintain them all in a single database, which has obvious implications for more efficient information management and searching (see Nichols, Witten, Keegan, Bainbridge, & Dewsnip, 2005 for a description of Unicode-based software for libraries with multilingual holdings).

The concept of relevant social groups can also be deployed in order to understand the limits of Unicode, whose focus has largely been on scripts used in business. For the core membership of Unicode, the problem that it purports to solve is that of internationalization. However, alternative relevant social groups, such as UNESCO and the Script Encoding Initiative,[28] attribute different meaning to the project. For the latter, the importance of

successfully integrating a minority language into Unicode has nothing to do with business; rather, it 'will help to promote native-language education, universal literacy, cultural preservation, and remove the linguistic barriers to participation in the technological advancements of computing.'[29] Moreover, the reasons given by such groups for the absence of minority languages from Unicode are explicitly political and include references to the relative poverty of speakers of minority languages, the obvious barriers they face in attending standardization meetings and drawing up proposals, and the fact that they do not constitute a large consumer base. Indeed, as Gee notes, citing Anderson (2004), '[w]hile the business interests have been actively behind much of the character encoding. ... advocates for the lesser-known scripts have not had a similarly strong presence among the Unicode Consortium membership' (Gee, 2005, p. 249).

UNESCO devotes a large section of its website to the issue of multilingualism on the internet, and its framing of the issue is patently clear. A page entitled Multilingualism in Cyberspace opens with the following sentence: 'Today various forces threaten linguistic diversity, particularly on the information networks,'[30] thus locating their interest in encoding issues in the field of language preservation. However, it is also framed as pertaining to the digital divide:

> Increasingly, knowledge and information are key determinants of wealth creation, social transformation and human development. Language is the primary vector for communicating knowledge and traditions, thus the opportunity to use one's language on global information networks such as the Internet will determine the extent to which one can participate in the emerging knowledge society. Thousands of languages worldwide are absent from Internet content and there are no tools for creating or translating information into these excluded tongues. Huge sections of the world's population are thus prevented from enjoying the benefits of technological advances and obtaining information essential to their wellbeing and development.

Likewise, linguistics expert and UNESCO consultant John Paolillo writes that, '[f]or the Internet to allow equivalent use of all of the world's languages, Unicode needs to be more widely adopted' (Paolillo, 2005a, p. 73).

This is an example of 'interpretive flexibility' (Kline & Pinch, 1996; Pinch & Bijker, 1987). For one group Unicode is a way to simplify software internationalization and thus increase profit margins, while for another it is a means of preserving endangered languages and narrowing the digital divide. The former interpretation is currently dominant, though organizations such as the Script Encoding Initiative and UNESCO are trying to impose their interpretation as well.[31]

The second aspect that a student of technology must be sure to highlight is that of alternatives to the dominant technology; that is, we must avoid the tendency to see the 'victorious' technology as the only one in the field (Pinch & Bijker, 1987) and make room in our analyses for competing, though less successful technologies too. With the case of Hebrew, we saw how visual Hebrew constituted competition to logical Hebrew; with Unicode, the competition, such as it is, would seem to be coming from what is known as the TRON project, a Japanese-based multilingual computing environment.[32] Raising a theme to which I shall return below, a senior engineer from that project sees the ascendancy of Unicode as directly linked to its support from leading US computer manufacturers and software houses, who promoted Unicode for reasons of economic gain, not out of consideration for the end user. Indeed, the full members of the Unicode Consortium are Microsoft, Apple, Google, Yahoo!, HP, IBM, Oracle, Adobe, and Sun Microsystems.[33] Their economic gain, it is argued, lies in the development of a unified market, especially across East Asia, which would 'make it easier for U.S. firms to manufacture and market computers around the world.'[34] Programmer Steven J. Searle, leading

representative of the competing TRON project, makes the point, therefore, that Unicode did not become the dominant standard on account of its technological superiority alone (indeed, that in itself is questioned), but rather because of the alliance of US firms supporting it.

Discussion: The multilingual internet and globalization

Having 'inverted' the infrastructure of Unicode, which incorporates logical Hebrew through the Unicode Bidirectional Algorithm, I turn now to a discussion of its purported impacts—the third element of the discourse surrounding Unicode mentioned above—and in particular to arguments that, despite seemingly enabling the localization of the internet, Unicode is yet another instance of western cultural imperialism. This kind of argument is essentially a sub-set of claims made about English becoming a global language at the expense of local languages, now faced with an ever greater danger of extinction (Phillipson, 1992, 2003; Skutnabb-Kangas, 2000; Skutnabb-Kangas & Phillipson, 2001).[35] In turn, this debate is one aspect of broader questions about globalization, and in particular the relations between center and periphery.

While there are many ways of interrogating the literature on globalization—how new is globalization? Is it really globalization or just westernization?—a very popular framework explores the dynamic between the global and the local. In the first moment of this dynamic, globalization is understood as the extension of the (economic, political and cultural) structures of capitalism throughout the world, bringing new populations within their oppressive control (Bauman, 1998; Schiller, 1976; Tomlinson, 1991). According to this view, western, and especially American, cultural forms and economic modes of operation are seen to be relentlessly colonizing the rest of the world. Ritzer's neo-Weberian work on McDonald's is a case in point (Ritzer, 1993).

In its second moment, this perspective is challenged by cultural sociologists and anthropologists who counter views of the homogenization of the world by showing how locals playfully interpret new cultural forms, thereby maintaining their distinct identity. Early expressions of this position are particularly identified with researchers of hybridity such as Ulf Hannerz (1992) and Jan Nederveen Pieterse (1995). In the context of McDonald's, various studies have tried to show how the chain is locally adapted, or becomes a hybrid, in specific localities (for case studies from the Far East, see Watson, 1997).

The third moment in this dynamic includes two main responses to the global/local debate. The first can be seen as a kind of 'yes, but…' response. Here, homogenization theorists accept that locals do indeed interpret new cultural forms in original ways, but they argue that hybridization theory goes too far. Martell, for instance, notes that processes of hybridization nonetheless occur within the context of cultural imperialism (Martell, 2010). This is a similar position to that argued for by Ritzer in his work on the 'globalization of nothing' (where 'nothing' is defined as 'centrally conceived and controlled forms largely empty of distinctive content' (Ritzer, 2004, p. 13)). Indeed, Ritzer coins the term 'grobalization' precisely to counter what he views as the overstated claims of glocalization.

Another response to the debates over homogenization/hybridization has been to say that the terms of the debate are in themselves misguided, and specifically that one can no longer talk in terms of the global and the local. Instead, say Hardt and Negri, for instance, both globalization and localization should be understood as part of a single '*regime* of the production of identity and difference' (2000, p. 45 emphasis in the original). In fact, they argue, the global/local dichotomy collapses because it wrongly assumes the existence of a point outside the global. This is similar to Ritzer's point that '*it is increasingly difficult to find anything in the world untouched*

by globalization' (Ritzer, 2004, p. 169, emphasis in original), to which we might add: were we (the west) to find anything truly local, by finding it we would only be bringing it within the scope of the global.

These debates over globalization can be found in various forms in the discourses surrounding the multilingual internet. As already noted, the dominance of English in the early internet led to fears that it would severely harm minority languages and cultures around the world, a position strongly identified with Crystal (1997), Phillipson (1992), and Skutnabb-Kangas (2000), who, together with Phillipson, warned against the dangers of 'linguicide' (Skutnabb-Kangas & Phillipson, 2001). This view was also given expression in popular discourse. For instance, a New York Times article from 1995 commented that 'some countries, already unhappy with the encroachment of American culture—from blue jeans and Mickey Mouse to movies and TV programs—are worried that their cultures will be further eroded by an American dominance in cyberspace' (Pollack, 1995).

Unicode has long been seen as a solution to this problem. In this regard, it can be understood as the 'local' response to the global. This is certainly how the ICANN executives viewed their decision to enable internationalized ccTLDs—based on Unicode technology—as described at the beginning of this paper. For instance, at the ICANN board meeting held in Seoul in October 2009, one board member said,

> This truly is a momentous time, and although the Internet was developed in western culture and with a great deal of it being done in the U.S., I think there is an enormous amount of good feeling and intention that this be—that the Internet be a balanced, extended, and open access for all peoples of the world.[36]

While another commented: 'This helps us live up to our shared goals of: One world, one Internet, everyone connected. Now, in people's own script'.[37] This was also how the Chinese Domain Name Consortium argued in favor of internationalized domain names in a proposal

submitted to ICANN in 2005: 'it helps to preserve culture diversiveness [sic] and protect special interests of people in different regions' (Chinese Domain Name Consortium, 2005). So if the first iterations of the internet gave rise to fears of the Englishization of the world, the promise extended by Unicode to enable non-English speakers to participate fully in the internet can be seen as the 'local' moment in the global/local dynamic described above. This dynamic has been summed up tidily as follows:

> On the one hand, the emergence of the interregional networks and systems of interaction and exchange of globalization has encouraged the spread of English. […] On the other hand, the ease and relatively cheap cost of using information technology allows any language group to produce its own sites, journals and programmes (Wright, 2004, p. 5).

However, just as critical theorists have gone beyond the global/local distinction to theorize the world as a single system (such as Hardt and Negri's notion of Empire), in the debates over endangered languages certain critical sociolinguists have challenged the fundamental assumptions of those who seek ways of preserving minority languages. The arguments that one can extrapolate from these texts are that either Unicode is irrelevant, or, precisely by enabling minority language users to stay within their language, it is condemning them to continued economic disadvantage.[38]

The argument that Unicode might be irrelevant can be gleaned from the writings of sociolinguists who refute the Whorfian association between language, culture and identity. Sociolinguist Jan Blommaert exposes the assumptions behind the model of 'linguicide': 'wherever a "big" and "powerful" language such as English "appears" in a foreign territory, small indigenous languages will "die". There is, in this image of sociolinguistic space, place for just one language at a time' (Blommaert, 2010, p. 44). Similarly, Pennycook states that we 'need to think about language and globalization outside the nationalist frameworks that gave rise to 20th-century models of the world' and that we should question 'the very concepts of language,

policy, mother tongues, language rights and the like that have been the staples of language policy up to now' (Pennycook, 2010, pp. 199, 200). In other words, according to these writers, the language 'protectionists' are operating according to a faulty model of the relations between language, culture and society.

The second criticism of the view that Unicode can help bring more of the world in and preserve endangered languages sees as paternalistic the attempt to help, or maybe force, speakers of minority languages keep their languages alive (Ladefoged, 1992). Put simply:

> Why should someone from Friesland have to speak Frisian, or act Frisian? She might prefer to be a punk, or a Buddhist. […] The preservation of a language community very often means the continued oppression of women, children, young people, the dispossessed, deviants, and dissidents (de Swaan, 2004, p. 574).

De Swaan argues that 'parents (and children) prefer a language that they assume will maximize their opportunities on the labour market, rather than the minority language that linguists and educationalists so well-meaningly prescribe' (de Swaan, 2004, p. 572). And who, he would ask, are we to denounce that? This is a position lucidly articulated by University of Chicago linguist, Salikoko Mufwene, who describes language endangerment as a 'wicked problem' in that it 'sometimes boils down to a choice between saving speakers from their economic predicament and saving a language. […] Language shift must be treated as one of the speakers' survival strategies…' (Mufwene, 2002, p. 377). And Blommaert argues that, in certain circumstances, promoting a people's mother tongue can be 'seen as an instrument *preventing* a way out of real marginalization and amounting to keeping people in their marginalized places…' (Blommaert, 2010, p. 47, emphasis in original).

Now I believe we are fully able to appreciate the infrastructural inversion carried out in the first sections of this paper and unpack its contribution to debates about globalization in general and the (bemoaned) growing power of English at the expense of other languages. In this

article, through the case study of Hebrew, I have shown how the infrastructure of the multilingual internet has been created by western organizations. As McMahon says about globalization in general, 'while globalization is specifically about the rearrangement of global resources into a new system of production and consumption, the essential characteristics of this new system have been determined by at most a very few nations […] and mostly by one, the US' (McMahon, 2004, p. 72; see also Dor, 2004, on 'imposed multilingualism'). This is not to say that Unicode is a state-led effort. Indeed, Bartley notes that 'standard setting and regulation are increasingly being accomplished through private means' (Bartley, 2007, p. 297), and, as I have done above, urges us to explore the political economies of those standards.

So the infrastructure for today's global computer-mediated communication is both overwhelmingly western and privately driven. While positive benefits may accrue to minority languages from Unicode, as seen by UNESCO's endorsements and promotion of Unicode, the overall context is still one in which the technologies of the internet are developed by western organizations. Because the context for this entire discussion is that of a world dominated by the political, cultural and economic apparatuses of the west, benefits to minority languages do not necessarily translate into benefits to those who speak them. Unicode may give the linguistically oppressed peoples of the world an online voice, but, as our infrastructural inversion of Unicode and logical Hebrew has shown, that voice is mostly provided by American multinational corporations.

To add a final twist to this discussion, it is worth noting that Unicode is quite a singular case in terms of globalization theory. In terms taken from John Urry, it is a 'global network' (in that 'the same "service" or "product" is delivered in more or less the same way across the network') which, seemingly paradoxically, contributes to the global 'fluidity' and disorderliness

of the internet (Urry, 2005, p. 245). It is not a cultural good in the way that a Hollywood movie is, and nor is it a kind of identity. However, its paths of diffusion are similar to those of such global cultural forms, and it assists in the computer-mediated communication of local identities, especially between Diasporic groups (on the importance of the internet in maintaining ties between parts of diasporas in the cases of Iran, Trinidad and Wales respectively, see M. Graham & Khosravi, 2002; Miller & Slater, 2000; Parsons, 2000). Indeed, the success of Unicode has been dependent on precisely those mechanisms so often critiqued as transmitting cultural imperialism, while enabling the use of non-English languages based on a quintessentially global technology. Increased homogeneity and standardization at the level of software and encoding is enabling increased heterogeneity at the level of the end user and his community. This is a precarious heterogeneity, however: while Unicode is discursively supported by organizations such as UNESCO, it is an infrastructure controlled by American multinationals and which would probably not have come into being without them.

Conclusion

In this article, the findings of the infrastructural inversion of the multilingual internet have been used to contribute to debates about globalization in general and its implications for language preservation in particular. Using Hebrew as a generalizable case study, we have seen the important role played by leading global technology companies in developing and diffusing logical Hebrew and Unicode as solutions for a language that is written from right to left and with a non-Latin alphabet. This role played by western MNCs feeds into our reading of the place of Unicode in the debates over globalization and language endangerment. The technologies of contemporary computer-mediated communication can thus be seen as an important site for research not only in their own right but also, and importantly, because the power relations they

embody impinge on theoretical debates in the field under study—here, the multilingual internet. In this sense, this paper is a call for scholars not to forget the 'plumbing' that makes computer-mediated communication possible in the first place.

---

[1] At the time of writing, a number of other countries' IDN ccTLD applications to ICANN are pending, including Israel's (see http://www.isoc.org.il/domains/idn_eng/).

[2] In 2003, the Vice President of the Unicode Consortium said, 'Unicode is like plumbing' (see http://www.nytimes.com/2003/09/25/technology/for-the-world-s-abc-s-he-makes-1-s-and-0-s.html)

[3] Such case studies range from electric cars (Callon, 1987) to the color of microwave ovens (Cockburn & Ormrod, 1993).

[4] Personal correspondence, 9th June, 2011.

[5] In 2003, UNESCO even adopted a 'Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace', in which it explicitly states its belief in '[p]romoting linguistic diversity in cyberspace' (UNESCO, 2003). Again, this would not have been said unless there were fears of a lack of linguistic diversity in cyberspace.

[6] These are conditions that pertain to CJK (Chinese, Japanese and Korean) languages (Shapard, 1993; Zhao, 2005) and to Arabic (Allouche, 2003; Palfreyman & al Khalil, 2003), but also to Hebrew, the case around which this paper is organized.

[7] In slightly more technical terms: a seven-bit code is one that uses binary numbers of up to seven digits in length. Such a code will therefore have 2x2x2x2x2x2x2=128 different binary numbers.

[8] Again, the reason is technical, but simple: bits always come in groups of eight. Therefore, a seven-bit piece of code always starts with a 0. If you turn that 0 into a 1 (thereby making it an eight-bit code), you have another 128 numbers at your disposal. (See also Haible, 2001.)

[9] A solution to this problem was found that involved embedding different encodings within a single document, but it was far from satisfactory.

[10] In Russia, where there were several different ways for encoding Russian letters, this could be particularly problematic. Similar problems are reported with Hindi (Paolillo, 2005a, pp. 54-55).

[11] Though the process leading up to the decision was started in 2001. See http://www.itpolicy.gov.il/topics_websites/takam5.htm and http://takam.mof.gov.il/doc/horaot/takam5.nsf/0/369e2da72849cd7d422567ed002b8c66?OpenDocument (in Hebrew).

[12] I base this on interviews with actors involved in the development of technologies for dealing with Hebrew on the internet, and on extensive reading of websites on the issue.

[13] Though not at the expense of visual Hebrew, which is still supported by HTML 4.01, primarily to ensure backward compatibility with older sites.

[14] This was related to me in interview by Jonathan Rosenne, an Israeli programmer and expert on issues of Hebrew on the internet. This expertise is reflected in his participation in pertinent committees in Israel and overseas.

[15] This was explained to me in interview by senior IBM employee, Matitiahu Allouche.

[16] More precisely, Netscape Navigator would show sites written in visual Hebrew but without realizing it was doing so—this required the development of a special font that the end user had to install. By using this font, the browser would be made to think it was dealing with Latin text, while the surfer would see letters in the Hebrew font.

[17] http://www.unicode.org/standard/WhatIsUnicode.html

[18] See http://googleblog.blogspot.com/2010/01/unicode-nearing-50-of-web.html.

[19] See http://trends.builtwith.com/encoding. The data on that site are updated weekly. The figure of at least 70% refers to June 2011.

[20] Extensible Markup Language: 'A way of organizing text information by labeling it with specified variables in a fixed format', from A Glossary of Online News Terms, http://www.ojr.org/ojr/wiki/glossary.

[21] See http://www.jpost.com/Business/BusinessNews/Article.aspx?id=23421.

[22] See http://www.webpronews.com/google-sites-goes-hebrew-and-arabic-2009-09.

[23] See http://www.intomobile.com/2009/08/27/google-search-results-for-feature-phones-now-available-in-arabic-and-hebrew/.

---

[24] See http://www.webpronews.com/google-sites-goes-hebrew-and-arabic-2009-09.

[25] See http://googleblog.blogspot.com/2008/07/hitting-40-languages.html.

[26] See http://code.google.com/p/android/issues/detail?id=5925 for a long discussion of issues to do with Unicode on Android devices, especially mobile telephones.

[27] See http://www.unicode.org/history/ for a more detailed account.

[28] See http://linguistics.berkeley.edu/sei.

[29] http://linguistics.berkeley.edu/sei/

[30] http://portal.unesco.org/ci/en/ev.php-URL_ID=16539&URL_DO=DO_TOPIC&URL_SECTION=201.html

[31] An attempt to include the Klingon script (from the Star Trek series) failed, showing how a certain interpretation of Unicode was rejected by that organization's institutions.

[32] See http://tronweb.super-nova.co.jp/characcodehist.html.

[33] See http://www.unicode.org/consortium/memblogo.html for a full list of the consortium's members.

[34] See http://tronweb.super-nova.co.jp/characcodehist.html.

[35] See UNESCO's site, http://portal.unesco.org/culture/en/ev.php-URL_ID=8270&URL_DO=DO_TOPIC&URL_SECTION=201.html. UNESCO claims that 50% of the world's 6,000 languages are endangered, and that a language 'dies' every two weeks.

[36] Dr. Stephen D. Crocker. Transcript downloaded from http://sel.icann.org/node/6751/.

[37] Rod Beckstrom, Transcript downloaded from http://sel.icann.org/node/6751/.

[38] Another criticism of the language endangerment school is that it ignores the fact that many indigenous languages are actually threatened by languages other than English, such as French or Portuguese, or by stronger languages within the country (this is a particulalry common phenomenon in Africa; see Mufwene, 2002). This criticism is noted here but not developed further.

Bibliography

Allouche, M. (2003). Re: Closing issue [arabicnum-03]   Retrieved 25/8/2005, from http://lists.w3.org/Archives/Public/public-iri/2003May/0016.html

Anderson, D. (2004). The Script Encoding Initiative. *SIGNA, 6*, 1-12.

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

Bartley, T. (2007). Institutional Emergence in an Era of Globalization: The Rise of Transnational Private Regulation of Labor and Environmental Conditions. *American Journal of Sociology, 113*(2), 297-351.

Bauman, Z. (1998). *Globalization: The Human Consequences*. Cambridge: Polity.

Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge, UK; New York: Cambridge University Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: classification and its consequences*. Cambridge, Mass.: MIT Press.

Brenzinger, M. (1992). *Language death: factual and theoretical explorations with special reference to East Africa*. Berlin; New York: Mouton de Gruyter.

Callon, M. (1987). Society in the Making: The Study of Technology as a Tool for Sociological Analysis. In W. E. Bijker, T. P. Hughes & T. J. Pinch (Eds.), *The Social construction of technological systems: new directions in the sociology and history of technology* (pp. 83-103). Cambridge, Mass.: MIT Press.

Celko, J. (2003). The Alphabet Achievement: Will Unicode be as easy as saying your ABCs?, *Intelligent Enterprise,* p. 48.

Chinese Domain Name Consortium. (2005). The Proposal for Internationalizing ccTLD Names.

Cockburn, C., & Ormrod, S. (1993). *Gender and technology in the making*. London; Thousand Oaks, Calif.: Sage.

Crystal, D. (1997). *English as a Global Language*. Cambridge: Cambridge University Press.

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

Dagan, N. (n.d.). Logical or visual? A practical approach to Hebrew content providing Retrieved 12/7/2005, from http://www.nirdagan.com/hebrew/compare

Danet, B., & Herring, S. C. (2003). Introduction: The Multilingual Internet. *Journal of Computer-Mediated Communication, 9*(1).

de Swaan, A. (2004). Endangered languages, sociolinguistics, and linguistic sentimentalism. *European Review, 12*(04), 567-580.

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.

Ellsworth, L. (1991). Three Cheers for Unicode, *InfoWorld,* p. 51.

Ess, C. (2006). From Computer-Mediated Colonization to Culturally-Aware ICT Usage and Design. In P. Zaphiris & S. Kurniawan (Eds.), *Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities* (pp. 178-197). Hershey, PA: Idea Publishing.

Gee, Q. (2005). Review of script displays of African languages by current software. *New Review of Hypermedia and Multimedia, 11*(2), 247 - 255.

Gewecke, T. (2010). iPad Language Capabilities Update. Retrieved from http://m10lmac.blogspot.com/2010/04/ipad-language-capabilities-update.html

Gillam, R. (2003). *Unicode demystified: a practical programmer's guide to the encoding standard*. Boston, MA: Addison-Wesley.

Graham, M., & Khosravi, S. (2002). Reordering Public and Private in Iranian Cyberspace: Identity, Politics, and Mobilization. *Identities: Global Studies in Culture and Power, 9*(2), 219-246.

Graham, T. (2000). *Unicode: A Primer*. Foster City, CA: M&T Books.

Haible, B. (2001). The Unicode HOWTO v1.0. Retrieved 5/7/2005, from ftp://ftp.ilog.fr/pub/Users/haible/utf8/Unicode-HOWTO-1.html#ss1.1

Hannerz, U. (1992). *Cultural Complexity*. New York: Columbia University Press.

Hardt, M., & Negri, A. (2000). *Empire*. Cambridge, Mass.: Harvard University Press.

Hoffman, P. (2000). Bringing in More of the World with Unicode. *Network World, 17,* 49.

Johnston, S. J. (1991, 25/2/1991). Unicode Promises An International Character Set, *InfoWorld,* p. 8.

Jordan, D. (2002). Languages left behind: Keeping Taiwanese off the World Wide Web. *Language Problems and Language Planning, 26*(2), 111-127.

Kelly-Holmes, H. (2004). An Analysis of the Language Repertoires of Students in Higher Education and their Language Choices on the Internet. *IJMS: International Journal on Multicultural Societies, 6*(1), 52-75.

Kline, R., & Pinch, T. (1996). Users as agents of technological change: The social construction of the automobile in the rural United States. *Technology and Culture, 37*(4), 763-795.

Koutsogiannis, D., & Mitsikopoulou, B. (2003). Greeklish and Greekness: Trends and Discourses of "Glocalness". *Journal of Computer-Mediated Communication, 9*(1).

Ladefoged, P. (1992). Another view of endangered languages. *Language, 68*(4), 809-811.

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: studies in sociotechnical change* (pp. 225-258). Cambridge, Mass.: MIT Press.

Law, J. (1987). Technology and Heterogeneous Engineering: The Case of Portuguese Expansion. In W. E. Bijker, T. P. Hughes & T. J. Pinch (Eds.), *The Social construction of technological systems : new directions in the sociology and history of technology* (pp. 111-134). Cambridge, Mass.: MIT Press.

Martell, L. (2010). *The sociology of globalization*. Cambridge ; Malden, MA: Polity.

McMahon, P. (2004). Money, Markets and Microelectronics: Building the Infrastructure for the Global Finance Sector. *Prometheus, 22*(1), 71-82.

Miller, D., & Slater, D. (2000). *The Internet: an ethnographic approach*. Oxford; New York: Berg.

Misa, T. J. (1992). Controversy and closure in technological change: Constructing "steel". In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: studies in sociotechnical change* (pp. 109-139). Cambridge, Mass.: MIT Press.

Mufwene, S. S. (2002). Colonization, globalization and the plight of 'weak' languages. *Journal of linguistics, 38*(02), 375-395.

Myers, S. (1995). Will Unicode Kill Japanese Kanji? Retrieved 1/1/2008, from http://www.japaninc.com/computingjapan/magazine/issues/1995/dec95/1295technica.html

Nederveen Pieterse, J. (1995). Globalization as Hybridization. In M. Featherstone, S. Lash & R. Robertson (Eds.), *Global Modernities* (pp. 45-68). London: Sage.

Nichols, D. M., Witten, I. H., Keegan, T. T., Bainbridge, D., & Dewsnip, M. (2005). Digital libraries and minority languages. *New Review of Hypermedia and Multimedia, 11*(2), 139-155.

Palfreyman, D., & al Khalil, M. (2003). "A Funky Language for Teenzz to Use:" Representing Gulf Arabic in Instant Messaging. *Journal of Computer Mediated Communication, 9*(1).

Paolillo, J. (2005a). Language Diversity on the Internet. In UNESCO Institute for Statistics (Ed.), *Measuring Linguistic Diversity on the Internet* (pp. 43-89). Paris: UNESCO.

Paolillo, J. (2005b). Language Diversity on the Internet: Examining Linguistic Bias. *Measuring Linguistic Diversity on the Internet*, 42-89.

Parsons, W. (2000). From Beulah Land to Cyber-Cymru. *Contemporary Wales, 13*, 1-26.

Pennycook, A. (2010). Rethinking origins and localization in global Englishes. In M. Saxena & T. Omoniyi (Eds.), *Contending with globalization in world Englishes* (pp. 196-210). Bristol, UK; Buffalo, NY: Multilingual Matters.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford [England] ; New York: Oxford University Press.

Phillipson, R. (2003). *English-Only Europe? Challenging Language Policy*. London: Routledge.

Pinch, T. J., & Bijker, W. E. (1987). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science, 14*(3), 399-441.

Pollack, A. (1995). A Cyberspace Front in a Multicultural War, *The New York Times*. Retrieved from http://www.nytimes.com/1995/08/07/business/a-cyberspace-front-in-a-multicultural-war.html

Pyatt, E. (2010, 2/7/2010). iPad & Unicode: B-. Retrieved from http://www.personal.psu.edu/ejp10/blogs/gotunicode/2010/07/ipad-unicode-b-.html

Ritzer, G. (1993). *The McDonaldization of society: an investigation into the changing character of contemporary social life*. Thousand Oaks: Pine Forge Press.

Ritzer, G. (2004). *The globalization of nothing*. Thousand Oaks, Calif.: Pine Forge Press.

Schiller, H. I. (1976). *Communication and cultural domination*. White Plains, N.Y.: International Arts and Sciences Press.

Schofield, J. (1993). Overcoming the Language Barrier, *The Guardian,* p. 19.

Shapard, J. (1993). Islands in the (Data)Stream: Language, Character Codes, and Electronic Isolation in Japan. In L. M. Harasim (Ed.), *Global Networks: Computers and International Communication* (pp. 255-270). Cambridge, Mass.: MIT Press.

Skutnabb-Kangas, T. (2000). *Linguistic genocide in education, or worldwide diversity and human rights?* Mahwah, N.J.: L. Erlbaum Associates.

Skutnabb-Kangas, T., & Phillipson, R. (2001). Linguicide. In R. Mesthrie (Ed.), *Concise Encyclopedia of Sociolinguistics*. Oxford: Elsevier Science.

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist, 43*(3), 377-391.

Tomlinson, J. (1991). *Cultural Imperialism*. Baltimore: Johns Hopkins University Press.

UNESCO. (2003). Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace. Paris: UNESCO.

Unicode Consortium. (2004). What is Unicode? Retrieved 7/12/2004, from http://www.unicode.org/standard/WhatIsUnicode.html

Urry, J. (2005). The Complexities of the Global. *Theory, Culture & Society, 22*(5), 235-254.

Wajcman, J. (2002). Addressing Technological Change: the Challenge to Social Theory. *Current Sociology, 50*(3), 347-363.

Warschauer, M., el Said, G. R., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer Mediated Communication, 7*(4), 0-0.

Watson, J. L. (Ed.). (1997). *Golden Arches East*. Stanford, Calif.: Stanford University Press.

Wright, S. (2004). Introduction. *International Journal on Multicultural Societies, 6*(1), 5-13.

Zhao, S. (2005). Chinese Character Modernisation in the Digital Era: A Historical Perspective. *Current Issues in Language Planning, 6*(3), 315-378.

Nicholas A. John (nicholas.john@mail.huji.ac.il) is a Brenda Danet Postdoctoral Fellow at the Department of Communication and Journalism, The Hebrew University of Jerusalem. He is currently researching for a book on the social logics of sharing in the new media and other social spheres.