2017

Early challenges to multilingualism on the Internet: the case of Han character-based scripts

Mark J. McLelland
University of Wollongong, markmc@uow.edu.au

Early challenges to multilingualism on the Internet: the case of Han character-based scripts

Abstract

In today's hyper-mediated world where computer software can deal seamlessly with a variety of the world's languages and scripts, it is difficult to recall the seemingly insurmountable computing problems raised by "Han" character-based scripts such as Chinese, Japanese (and to a lesser extent, Korean). In the early days of networked computing, some commentators even argued that the continued use of Han characters was a lost cause, and could only result in "intolerable inefficiencies" when used to communicate digital information. In this paper, I consider the orthographic factors that delayed the implementation of cross-platform protocols allowing for the input, display and transmission of character-based scripts across early computer networks (mid-1980s to mid-1990s). I note how Anglophone Internet histories have been largely oblivious to the inherent biases of Internet infrastructure that were built by programmers using ASCII (based on the limited range of characters provided by the Roman alphabet) who also assumed the QWERTY keyboard to be the obvious human-machine interface. Instead of stressing the deficiencies of character-based scripts, I invite the reader to consider how the Internet might look today had it not been founded upon assumptions based on Anglophone usage, and consider the potentialities of a non-phonetic character-based writing system.

Disciplines
Arts and Humanities | Law

Early challenges to multilingualism on the Internet: the case of Han character-based scripts

Mark McLelland

Abstract

In today's hyper-mediated world where computer software can deal seamlessly with a variety of the world's languages and scripts, it is difficult to recall the seemingly insurmountable computing problems raised by "Han" character-based scripts such as Chinese, Japanese (and to a lesser extent, Korean). In the early days of networked computing some commentators even argued that the continued use of Han characters was a lost cause, and could only result in "intolerable inefficiencies" when used to communicate digital information. In this paper I consider the orthographic factors that delayed the implementation of cross-platform protocols allowing for the input, display and transmission of character-based scripts across early computer networks (mid-80s to mid-90s). I note how Anglophone internet histories have been largely oblivious to the inherent biases of internet infrastructure that were built by programmers using ASCII (based on the limited range of characters provided by the Roman alphabet) who also assumed the QWERTY keyboard to be the obvious human-machine interface. Instead of stressing the deficiencies of character-based-scripts, I invite the reader to consider how the internet might look today had it not been founded upon assumptions based on Anglophone usage, and consider the potentialities of a non-phonetic character-based writing system.

Introduction

In today's hyper-mediated world where computer software can deal seamlessly with a variety of the world's languages and scripts across a range of applications, young people may find it difficult to imagine the seemingly insurmountable problems raised by the variety of scripts used to write the world's languages in the early days of

networked computing. As internet applications expanded in the 1980s, some commentators even argued that millennia-old native scripts such as Chinese (henceforth "Han") characters were a lost cause and would need to be replaced by Roman script if computerisation were to advance in the region (Unger, 1987, p. 8). I recall this situation very well myself though, due to my particular circumstances. From 1997 to 2000 I was a graduate student undertaking a PhD in Japanese Studies at the University of Hong Kong and so, despite being a native English speaker, I have first-hand experience of the frustrations of trying to get software to work across three different languages. As Nishigaki noted at the time "Any country that does not have English as the mother tongue will likely find the conditions of the Internet to be deeply unsatisfying" (1997, p. 6) and this was indeed my experience. Nothing about the internet was straight-forward at the time, including issues such as writing (and printing!) text that included both English and Japanese, sending (and receiving in an intelligible format) emails in Japanese or even displaying a Japanese website.

In this short overview I consider the linguistic and orthographic factors that delayed the implementation of cross-platform protocols allowing for the input, display and transmission of character-based scripts across early computer networks (mid-80s to mid-90s) in North-East Asia. It is useful for the project of internet histories to draw attention to how English speakers have been largely oblivious to the "invisible infrastructure of the Internet" (Pargman and Palme, 2009, p. 177) – since this helps us understand the bias that existed in early accounts of computer communication that saw languages and scripts other than English/Roman, particularly Han character-based scripts, as problems to be overcome. It is worth recalling that the foundation of today's networked computing was established by programmers using American Standard Code for Information Interchange (ASCII) that was based on the limited range of characters provided by the Roman alphabet. These programmers also adopted the QWERTY keyboard as the main human-machine interface, despite its inefficiency as an input system. Now that internet histories is a developing field with more attention being paid to regional histories (Goggin and McLelland, 2007; 2017) as well as pre-internet developments in computer mediated communication (CMC) (McLelland, 2017), it is useful to recall the hidden and persistent biases in early computer communication and to at least imagine how things might have turned out differently had English speakers not played such a key role in internet development.

ASCII Imperialism?

The dominance of English speaking countries in the development of computing in the middle of the last century led to a situation where the Roman alphabet and the English language were the default script and language used both for the construction of computer code and commands and for discussion concerning research and development (Breen and Tokita 2004, p. 1; UNESCO, 2005). The 7-bit code that was established by the US standards agency in the early 1970s, generally known as ASCII, was originally developed from telegraph code. It was a 7-bit code that allowed for 128 basic characters including lower and uppercase letters from a – z, numerals 0 – 9, and punctuation marks. With the extension to 8-bits, a further 128 characters were made available that included accented letters and additional punctuation marks (UNESCO, 2005, p. 71-73; Breen and Tokita, 2004, p. 1). Most countries using European languages were thus able to deploy the standard ASCII set while using the extra character capacity to configure diacritic, punctuation and other marks to suit the local writing system. Non-Roman alphabets such as Greek or Cyrillic were able to use the extended code space to configure their own alphabets. However, languages such as Chinese, Japanese and to a lesser extent, Korean, which use several thousand distinct Han ideographs that are not phonetic (in that the character itself contains only general indicators of how it might be pronounced) could not be written by such a restricted number of character options.

In 1983 ASCII became embedded in internet architecture via the Domain Name System (DNS) which maps the names of hosts or websites to their Internet Protocol (IP) addresses (Torsen 2005). Until 2003 it was not possible to link to addresses written with scripts that could not be accommodated by the limited ASCII code space. Another key feature that allowed networked communication via computers was the development in 1978 (and release onto ARPANET in 1983) of the TCP/IP set of protocols based on ASCII that allowed computers with different operating systems to communicate with each other via telephone lines (Carey and Elton, 2010, p. 217). The establishment of protocols enabled a range of text-based networks such as Fidonet, Compuserve and later America on Line to offer services outside of North America. However language options on these networks were limited by the software. Internet expansion across the Arab

world, for instance, was delayed by the technical difficulties of employing the traditional script, an issue not resolved until the release in 1997 of Microsoft's Arabic Office 97 (Daoudi, 2017, p. 229). Similarly, encoding issues across Chinese, Japanese and Korean limited the use of these scripts online until the roll-out of UNICODE in 1995.

Subsequent to the implementation of TCP/IP, the next development that increased the capacity for information to be sent and received via the internet was the World Wide Web (henceforth the Web), made available to the public in 1993. The Web had numerous advantages when compared with earlier applications, especially its capacity for handling formatted text, embedded graphics and sound and visual media. However Web programming and Hyper Text Markup language (HTML) carried over existing biases in terms of their reliance on Roman script. For instance, the lack of Arabic script HTML and browsers capable of displaying the script meant that "text had to be rendered as graphics, making it extremely slow to load" (Houissa, 2000, p. 59). This was also a problem for other languages using scripts other than Roman that initially needed to upload their Web pages as image files, resulting in slow download speeds and lack of searchability. Hence, as Pargman and Palme, in their analysis of "ASCII imperialism" conclude, "there does exist a bias among the organisations, institutions, and individuals" involved in setting the standards for computer communication "that works in favour of English-speaking Internet users and to the disadvantage of (in varying degrees) speakers of all other languages" (2009, p. 197). These concerns about the language bias written into internet infrastructures led authors of a 2005 UNESCO report to comment that "If digital literacy requires literacy in another language as a prerequisite, openness and universal access cannot be assured" (2005, p. 71).

Language problems on the early internet

Revisiting newspaper and magazine reports, academic papers and BBS discussions about the internet from the late 80s and early 90s reminds us just how novel and exciting the roll out of networked computer technologies seemed at the time. The assumptions made remind us that the implications and applications of new technologies are not always clearly understood, nor are the ways in which they will be adopted and

modified easily predicted. One key issue which has not so far been explored in detail in histories of the internet is the impact of language, specifically orthographic (that is, script) differences and the challenges posed to CMC by character-based scripts.

In Euro-American societies, the roll-out of public internet access, which was accelerated by the introduction of the Web and the first Web browsers in 1994, took place in a context where large segments of the population already had some prior experience of personal computing (Fouser 2001: 274). The fact that most high-school graduates already had a familiarity with the QWERTY keyboard, originally popularised by the Remington typewriter in 1873, meant that prior to the development of the mouse and the graphic interface, the use of the keyboard to interface with a computer screen was easily intelligible. Despite the fact that the QWERTY layout, originally devised to slow down typing and prevent the jamming of commonly occurring letters in English words, is not the most efficient for computer input, the widespread familiarity established by the typewriter has made the keyboard difficult to change (Castillo 2011, p. 613; Choi 2013, p. 37; Zhang 2016). Due to resistance from users, little adaptation of the original typewriter layout has taken place (Zhang 2016), other than the addition of function and arrow keys necessary for navigating a screen and for computer commands.

An existing familiarity with typewriters was not, however, the case for many non-European language users, particularly those in North-East Asia whose written languages were not alphabet based but depended to some extent upon the reproduction of complex Han characters – specifically Chinese, Japanese, and to a lesser extent, Korean (in Korea since the end of the Second World War emphasis has been place on the use of the native Hangul alphabet over Chinese characters). China and Japan did not see the widespread office automation characteristic of Western societies in the early postwar period. Korea, however, did develop numerous prototypes for a Hangul typewriter but standardisation was always a problem with "up to eleven competing keyboard designs" by the 1960s (Choi 2013, p. 42). Although typewriters for Chinese and Japanese scripts did exist they were cumbersome and could only be used by highly trained operatives, making it impossible to reproduce the typing pools that had developed across western businesses and government departments. Hence, the possibilities for easier text input, display and retrieval afforded by computers were of interest to East Asian governments and from the 1970s onward various schemes and protocols were explored by both

government and commercial agencies across the region, including engagement with agencies such as the Internet Engineering Taskforce, but not in any unified manner (Seo, 2013, p. 186; Contreras 2014).

One major problem was the restricted memory available in early personal computers making it difficult to develop code to deal with characters beyond the 26 letters of the Roman alphabet, Arabic numerals and a small number of key punctuation symbols associated with written English. The fact that the internet and CMC more generally were pioneered in the United States, a largely English-speaking jurisdiction, meant that there was little inclination to invest in technology to expand this character range. Hence, as Breen (2007, p. 1) points out, at the time of these early developments in computing, it made little sense to speak of "computing in English" – since the use of the English language was embedded in the very architecture of computer programs and English was the default language for international communication on the early internet (Breen and Tokita, 2004, p. 1; Parman and Palme, 2009, p. 184). However "computing in Japanese" (or Chinese or Korean) raised a whole set of issues particular to these languages and there have been a number of studies dedicated to the specific computing problems raised by use of East Asian scripts (*see for example*, Unger, 1987; Lunde, 1999; Gottlieb, 2000; W3C, 2012). As late as 1987 Unger was arguing that continued commitment to the use of Han characters as a major component of Japan's hybrid writing system would create "intolerable inefficiencies" (1987, p. 8) and that computerisation thus required script reform, including enhanced use of *romaji* or Romanized Japanese "in the majority of [data processing] applications" (1987, p.171).

Unger, and others arguing for the Romanisation of East-Asian languages, viewed character-based scripts as fundamentally "irrational" and ill-suited to modern communications systems. As Nishigaki points out, Chinese characters "have always been criticized and attacked in the name of an *obsolete writing system* which prevents modernization" (1999, p.17; emphasis in the original). Yet, it is pertinent to remember at this point that there is nothing "rational" about English spelling or the QWERTY keyboard which is, in fact, a rather inefficient input system that was standardized by a confluence of historical circumstances rather than deliberate policy based on efficiency (Choi 2013, p. 38; Zhang 2016). The view that character-based scripts are less optimal

is also dependent on a bias in Western linguistics that assumes written language should be made up of phonemes, that is, a restricted number of letters laid out horizontally to make up separate words (Nishigaki, 1999, p. 18). Han characters represent language differently. A single character is constructed as if to fill a square-shaped block of space from up to 30 different strokes that are written in a distinct order (top to bottom, left to right, character spanning or enclosing strokes last, et cetera). These strokes are organised according to 214 "radicals" that signify objects and abstract nouns – from basic elements like fire, earth, wind and water, through to more abstract ideas such as big and small, movement and speech, such that their combination gives a metaphorical sense of the character's overall meaning. For instance the radicals for "rice field" 田 and "strength" 力, when combined produce the character "man" 男 (that is, metaphorically, a person who provides strength in the fields). Given their condensed construction, a proficient reader can process the information in Han characters very quickly.

For Han characters, since meaning and pronunciation are separate issues, once a character's meaning is memorized it can be recognized and understood by anyone irrespective of the pronunciation assigned to it in different languages, just as a smiley face emoticon can be comprehended as signifying "happiness" despite the fact that the word happiness is different for speakers of different languages. For example, spoken Chinese is made up of numerous mutually unintelligible dialects – but the written characters can be understood by any literate person, irrespective of the pronunciation assigned in their specific region (Nishigaki, 1999, p. 18). Han characters can also be understood by readers across languages, such as those in Japan and Korea in a similar manner to how speakers of different languages can all use and comprehend emoticons. In the case of Japanese and also Korean, the large number of homophones (words with the same sound but different meanings) would make a Romanised version of a text difficult to comprehend – it is the Han characters, known respectively as *kanji/hanja,* that indicate the specific and unambiguous meaning, making *kanji* "indispensable for the reading of Japanese" (Shibamoto Smith, 1996, p. 210). In addition, since Japanese is written without word breaks, the *kanji* are critical "visual cues to morphological segmentation" that help the reader parse the lexemes essential to comprehension (Shibamoto Smith, 1996, p. 210). Although the transcription of Japanese via Roman letters or the local *hiragana* syllabary might seem more "logical" than the use of 2000

plus *kanji*, when written without *kanji* the language is extremely difficult to comprehend. Hence "language reform" that would have required the use of Roman or a similar alphabetic script, was never a viable option.

From monolingual to multilingual character coding

During the 1970s separate attempts were made in countries of the Norht-East Asian region by computer manufacturing companies and other bodies to develop "double byte character sets" that could cope with a non-phonetic ideograph-style writing system, but unlike in North America and Europe where the American system had become standard, there was no unifying system in place for the input and conversion of Han characters (which are used across all three languages). It was not until the 1980s that the needs for texts to move across national boundaries as well as the use of multiple scripts in the same document were clearly comprehended (Breen and Tokita, 2004, p. 2). Similarly, lack of agreement on how to handle the local *hiragana* and *katakana* syllabaries used in Japan and the Hangul alphabet in Korea also caused problems across different computer platforms. As Jo points out, attempts at standardization of the native Hangul script in Korea "repeatedly failed" and it was not until the roll-out of UNICODE in 1995 that there was broad agreement on its adoption as a national standard (2017: 199).

Further problems that delayed the development of a unified code were the facts that although Han characters originating in China are closely related across the region, they are not identical. Japan adapted existing characters and invented some new ones, as did Korea. Japanese usage of these characters also differs from the Chinese in that "[t]he same character may stand, as a homograph, for several different morphemes (each with its specific meanings and . . . pronunciation" (Shibamoto Smith, 1996, p. 209). In addition, other differences exist between Chinese speaking territories. For instance, the People's Republic of China rolled out a simplified set of characters in 1956 (known as Simplified Characters). These forms are also used in Singapore but not in Japan, Korea, Hong Kong or Taiwan (where Traditional Characters are used). It was not until 1993 when the Unicode/ISO 10646 code, a 16-bit system allowing a potential 65,536 characters was developed, that a system of inputting and displaying "unified CJK ideographs" across platforms and applications was made available. Previous locally

developed systems for the input and display of Han characters were not interoperable. Unicode's innovation was to give similar characters across the three languages the same code space meaning that any computer loaded with the required fonts was able to recognise and display them. Unicode has however been justly criticised for homogenising the characters and failing to provide code space for character variants (Auh 1998) that are important for the writing of personal and place names as well as characters that are seldom used in today's texts but are important in historical documents.

Moreover, the roll-out of Unicode did not address some fundamental biases concerning script built into the architecture of the internet itself. Given the very large number of Han characters in circulation (some estimates exceed 50,000), the pronunciation and meaning of unfamiliar characters is often indicated by an interlineal annotation known as ruby (or *furigana* in Japanese). However, Web browsers are inconsistent in their ability to recognise and display this important information. In addition, Japanese and Chinese texts are traditionally written from right to left in a vertical manner from top to bottom of the page (Shibamoto Smith, 1996, p. 214) and, unlike European languages which separate words on the page and screen, other than paragraph spaces, they are written without word breaks (Unger, 1987, p. 29-31). This distinctive writing style means there is no hyphenation for word breaks across lines, and consequently "word wrap" software must know where to break lexemes so as to maintain legibility. Similarly, the input and conversion software used to display text on the screen must recognise not just individual words but also lexemes since basic lexical components are not separated by spaces during input (W3C, 2012). Another fundamental bias concerning English is the reliance on the Roman alphabet plus numerals and common punctuation devices in the coding of internet protocols and Web domain addresses, "the technical legacy of the DNS's development and initial implementation in the United States" (National Research Council 2005, p. 197). An International Domain Name task force was established to address this issue in 2000 and in 2003 a standard conversion algorithm was developed supported by the Internet Corporation for Assigned Names and Numbers (ICAN n.d.; Torsen 2005). This allowed for a domain name to be written in a local script familiar to end users (deploying Unicode) and then converted into the international format familiar to Web browsers that utilised the limited ASCII character

set (Breen and Tokita, 2004, p. 3). It thus took a decade after the introduction of the Web before users in Japan and China were able to use their character-based scripts for the designation of local Web addresses.

The characteristics of written Chinese and Japanese also had implications for Web search engines. The development of search engines that enabled users to quickly locate a large amount of material, including image and sound files, by the use of key words or phrases was an important step. Yet, early pioneers in Web search technology such as Google and Yahoo deployed a standard framework for searching across all languages which served Japanese and Chinese rather poorly. As mentioned above, both languages are written without inter-word spaces or markers requiring a search engine to parse sections of text rather than pick up on isolated terms. Also, in Unicode, the Han characters used in Chinese, Japanese (and to a lesser extent, Korean) share the same code space, so when searched for using a global search engine, results can show up in each language despite the search terms having different connotations in the different languages (Auh 1998). A further issue with Japanese is that there is flexibility in the orthography for writing some terms (often depending on personal style and preference). This is a particular issue with the use of *katakana* for the transliteration of foreign loan-words, for which several variants may exist. In a trial of Google and Yahoo's capacity to search in Japanese and Chinese, Breen (2005, p. 6) noted that these issues could be mostly overcome by the use of the "exact phrase" option but that their effective use depended upon "being aware of the nature of the parsing and indexing in order to make full use of the engines." This of course required an extended period of acculturation prior to the efficient adoption of Web use by native speakers of these languages.

Conclusion

This brief discussion has looked at the influence of language on the early internet, particularly script input and retrieval methods. Given the complex input and display issues associated with non-Roman scripts such as Chinese, Japanese (and to a lesser extent, Korean), the introduction of computer mediated communication in North-East Asia was less straightforward than in North America and Europe (Unger, 1987; Gottlieb, 2000; McLelland, 2017). Computer users in these societies had first to become

familiar with Roman script, the QWERTY keyboard and the non-intuitive "conversion" input style required for local language display. Furthermore, the Roman script is part of the very architecture of the internet and early Web browsers were not configured to support domain names in non-Roman scripts or search for or display text in East-Asian scripts as accurately as they could in European languages.

I have argued that these problems largely arose due to the fact that advances in computer communication and the establishment of the internet took place in an English-speaking environment. However the calumny heaped upon North-East Asian scripts, particularly the continued use of Han characters, is due to a cultural bias that assumes the primacy of English and views other languages as problematic to the extent that they present challenges for software built with the use of English in mind. In particular, as Shibamoto Smith points out, "no writing system has been written about so pejoratively as Japanese" despite the fact that Japan is among the world's most literate societies (1996, p. 214) and was throughout the 1970s and 80s seen as an economic powerhouse and model for Western economies (Vogel, 1979). Nishigaki also points out, with some irony, that the increasing use of "icons," that is, non-phonetic pictograms (such as the "trash" bucket) that are used to negotiate today's graphic interfaces, have much in common with how Chinese characters function, stating "No visual symbols on earth have higher ability of representing abstract concepts than Han characters" (1999, p. 19). Indeed the very condensed nature of Han characters and the fact that they are not tied to specific pronunciations promote speed reading and the efficient processing of information that makes them better for communication in a multilingual environment than a phonetic system like English (Nishigaki, 1999, p. 19). It has even been argued that given advances in predictive software, it is now faster to input Chinese on a computer using the stoke-order method than it is to input English text and that the continued reliance on the QWERTY keyboard is a liability for users of European languages (Zhang 2016).

An interesting alternative history of language on the internet, rather than looking at the "problematic" nature of non-Roman, particularly character-based scripts, would instead problematize the manner in which one particular script became so embedded in computer software and internet infrastructure. What might the history of the internet

have looked like had it been pioneered in China? Would we today be scoffing at the deficiencies of the QWERTY keyboard and discussing the inefficiencies of "computing in English"? Indeed "challenging the West as default" (Zhang 2016) and imagining internet histories outside of Anglophone paradigms can be a productive process, one that we should all look forward to in future issues of this journal.

References

Auh, T. S. (1998). Promoting Multiculturalism on the Internet: Korean Experience. Paper presented at the Graduate School of Journalism and Mass Communication, Korea University, Republic of Korea. Downloaded from: http://www.unesco.org/webworld/infoethics_2/eng/papers/paper_8.htm

Breen, J. and A. Tokita. (2004). The WWW in Japan: A threat to Cultural Identity or a Domesticated System? Paper presented at the First International Conference on Cultures and Technologies in Asia, Mumbai, India. Downloaded from: http://www.edrdg.org/~jwb/paperdir/jwww.html

Breen, J. (2005). WWW Search Engines and Japanese Text. Paper presented at the Sixth Symposium on Natural Language Processing, Chiang Rai, Thailand. Downloaded from: http://www.edrdg.org/~jwb/paperdir/wwwjsrch.html

Carey, J. and Martin E. (2010). *When Media Are New: Understanding the Dynamics of New Media Adoption and Use*. Ann Arbor: MIT Press.

Castillo, M. (2011) "QWERTY, @, &, #," *American Journal of Neuroradiology*, 32, 613-14.

Contreras, J. (2014). Divergent Patterns of Engagement in Internet Standardization: Japan, Korea and China, Telecommunications Policy 38: 914-32.

Choi, Y. B. (2013). Path Dependence on the Korean Keyboard. *Journal of Economic Behaviour and Organization*, 88, 37-46.

Daoudi, A. (2017). Rethinking Arabic Linguistics: The History of the Internet in the Arabic-Speaking Region and the Rise of e-Arabic. In G. Goggin and M. McLelland (eds) *The Routledge Companion to Global Internet Histories*. New York: Routledge, pp. 227-43.

Fouser, R. (2001). 'Culture', Computer Literacy and the Media in Creating Public Attitudes to CMC in Japan and Korea. In C. Ess, (ed.), *Culture, Technology, Communication: Towards an Intercultural Global Village* (pp. 261-78). Albany: SUNY Press.

Goggin, G. and M. McLelland. (2007). *Internationalizing Internet Histories: Beyond Anglophone Paradigms*. Abingdon: Routledge.

Goggin, G. and M. McLelland. (2017). Global Coordinates of Internet Histories. In G. Goggin and M. McLelland (eds), *The Routledge Handbook to Global Internet Histories*. New York: Routledge, pp. 1-19.

Gottleib, N. (2000). *Word Processing Technology in Japan: Kanji and the Keyboard*. Richmond: Curzon.

Houissa, A. (2000). The Internet Predicament in the Middle East and North Africa: Connectivity, Access and Censorship. *Journal of Librarianship and Information Science*, 32(2), 56-63.

ICAN. (n.d.) Guidelines for the Implementation of Internationalized Domain Names/ Version 1.0. downloaded from: https://www.icann.org/resources/pages/idn-guidelines-2003-06-20-en

Jo, Dongwon. (2017). H-Mail and the Early Configuration of Online User Culture in Korea. In G. Goggin and M. McLelland (eds), *The Routledge Handbook of Global Internet Histories*. New York: Routledge, pp. 197-208.

Lunde, K. (1999). *CJKV Information Processing*. Sebastapol C.A.: O'Reilly & Associates.

McLelland, M. (2017). Early Computer Networks in Japan 1984-1994. In G. Goggin and M. McLelland (eds), *The Routledge Handbook of Global Internet Histories*. New York: Routledge, pp. 171-81.

National Research Council. (2005). *Signposts in Cyberspace: The Domain Name System and Internet Navigation*. Washington D.C: National Academies Press.

Nishigaki, T. (1997). A Multilingual Information Environment for a Multicultural World. *Social Science Japan.* August, 6-8.

Nishigaki T. (1998). Multilingualism on the Net. Paper presented at UNESCO INFOethics '98. Downloaded from:
http://www.unesco.org/webworld/infoethics_2/eng/papers/paper_5.htm

Nishigaki, T. (1999). What Can MT Do for Multilingualism on the Net? Proceedings of Machine Translation Summit VII. Downloaded from:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.493.1099&rep=rep1&type=pdf

Pargman, D. and J. Palme. (2009). ASCII Imperialism. In M. Lampland and S. L. Star (eds), *Standards and their Stories: How Quantifying, Classifying and Formalizing Practices Shape Everyday Life* (pp. 177-99). Ithaca: Cornell University Press.

Seo, D. B. (2013). *Evolution and Standardization of Mobile Communications Technology*. Hershey, P.A.: Information Science Reference.

Shibamoto Smith, J. (1996). Japanese Writing. In P. Daniels and W. Bright, (eds), *The World's Writing Systems* (pp. 209-14). Oxford: Oxford University Press.

Torsen, Molly. (2005). The Domination of the English Language in the Global Village: Efforts to Further Develop the Internet by Populating it with Non-Latin-Based Languages. *Richmond Journal of Law & Technology*, XII(I): 1-23.

UNESCO. (2005). *Measuring Linguistic Diversity on the Internet*. Paris: UNESCO.

Unger, J. M. (1987). *The Fifth Generation Fallacy: Why Japan Is Betting its future on Artificial Intelligence*. New York: Oxford University Press.

Vogel, E. (1979). *Japan and Number One: Lessons for America*. Cambridge: Harvard University Press.

W3C Working Group. (2012). *Requirements for Japanese Text Layout*. Downloaded from: https://www.w3.org/TR/2012/NOTE-jlreq-20120403/

Zhang, Sarah. (2016). Chinese Characters Are Futuristic and the Alphabet Is Old News. *The Atlantic*. November 1. Downloaded from: http://www.theatlantic.com/technology/archive/2016/11/chinese-computers/504851/