

# Needing to Have A Voice: Linguistic Grouping in the Digital Networked Environment

HAN-TENG LIAO | JULY 2011

INSTITUTE FOR THE STUDY OF DIPLOMACY

“[T]here isn’t an economic internet and a social internet and a political internet; there’s just the Internet,” U.S. Secretary of State Hillary Clinton said about the U.S.’s Internet freedom agenda, claiming that there should not be “walls that divide the Internet.”

Indeed, as the Internet is embedded in all aspects of everyday life, it is difficult for states to contain Internet and Internet-enabled activities in one or a few selected domains. However, what really divides the Internet may not be “walls” erected by governments that do not like the U.S.’s Internet freedom agenda but the linguistic barriers that have been used to draw the boundaries in which voices can be articulated, heard, and exchanged. Though the Internet may be global and universal in its reach, any Internet user can only access certain segments of the Internet that are determined by his or her linguistic capacity. There may still be a Chinese-language Internet and an Urdu-language Internet.

The link between language and media is not new. Consider the following quotes regarding the role of languages, first in the print media two hundred years ago and second in social media today:

These [mechanically reproduced] print-languages laid the bases for national consciousnesses. . . . These fellow-readers . . . formed . . . the embryo of the nationally imagined community. . . . Second, print-capitalism gave a new fixity to language. . . . Third, print-capitalism created languages-of-power.<sup>1</sup>

Benedict Anderson, *Imagined Communities*

The contours of diplomatic engagement are changing rapidly, as are the environments in which diplomacy is crafted, honed, and practiced. New media have changed the pace and content of political awareness and provided new tools for diplomacy.

Every global issue now tests the assumptions and practices of traditional diplomacy. Non-state actors—whether benign or malign, constructive or disruptive—now play increasingly important roles in the conduct of international politics and lead us to think differently about global development, conflict, and reconciliation.

These issues, conditions, and actors are helping to refine, and perhaps redefine, what diplomacy means, how it is conducted, and how we examine the new terrain of diplomacy.

Han-Teng Liao was the Yahoo! Fellow in Residence (2010–2011) at the Institute for the Study of Diplomacy.

Social media is all about connections. . . . There's no stronger connection than literally speaking the same language.<sup>2</sup>

Quote from a digital marketing specialist

Benedict Anderson argues that print technologies, print-languages, and print-capitalism have laid “the bases for national consciousnesses”<sup>3</sup> when he examines the creation and global spread of nationalism from the late eighteenth century onwards. When the world's media content and technologies converge in the digital networked environment,<sup>4</sup> what are the shaping roles of languages? If Anderson is right that “capitalism and print created monoglot mass reading publics,”<sup>5</sup> what kind of language users have the Internet technologies created?

One does not have to agree that speaking the same language is the strongest connection for users of social media in order to appreciate the important role that language may play in the way Internet users are grouped, which is the subject of this essay. How has the role of languages shaped and influenced the grouping of Internet users?

This paper begins by discussing an often-overlooked point in understanding the digital and networked environment: Before one has a voice, one has to speak a language of some kind. To be heard, one has to speak the language of an audience. Thus, the idea of providing more information and faster communication is not that simple and straightforward for all the languages in the world. Though the general trend of digital and computer network development is to become faster and bigger in content production and circulation, not all languages suffer from the problem of overabundance or information overload as English does.

At a minimum, without proper standards and technologies for encoding and decoding human-readable texts in digital codes, a language can never enjoy the volume and speed that digital technologies can provide. “All human-readable text has a language,”<sup>6</sup> but not all languages are ready and available for the latest information, communication, and telecommunication (ICT) technologies.

In other words, what are the dynamics of linguistic grouping in the digital networked environment? A series of questions

can be asked again about Internet technologies. First, what kind of “digit-coded languages” are mechanically reproduced that may reconfigure the bases for a national (or other kind of) consciousness? Under this overarching question, three sub-questions can be raised following Anderson’s formulation in the opening quote: (1) In the digital networked environment, what kinds of fellow-readers are formed, and what are their implications for the idea of an imagined community? (2) Are languages more fixed or fluid? (3) What kind of languages-of-power are reshaped or created?

The above questions, centering on the idea of having a voice online, are fundamental for the future of Internet. If properly answered, the questions have implications for linguistic human rights online and ICT for development. It all comes down to the double meaning of the term “voice.”

On one hand, a voice is human utterance articulated in a certain language. On the other hand, it is about “a right to express a preference or opinion,” which can be traced back to casting votes in a deliberative assembly.<sup>7</sup> Thus, the movement of linguistic human rights, or language human rights, which emerged mainly in the 1980s mostly in Europe, should be regarded as an effort to remedy the assumption that monolingual speech is ideal for the national or public sphere.<sup>8</sup> The expansion of the human rights movement from a political and civic rights movement to the promulgation of economic, social, and cultural rights thus has linguistic components in international documents. These include the 1948 Universal Declaration of Human Rights, the 1966 International Covenant on Civil and Political Rights, and the 1996 Universal Declaration of Linguistic Rights.<sup>9</sup> The idea of multiple voices in multiple languages has also been linked to the provision of universal access to cyberspace when the United Nations Educational, Scientific, and Cultural Organization (UNESCO) recommended “the promotion and use of multilingualism” in 2003. Since then, UN member states have been requested to submit a report once every four years on such action taken.<sup>10</sup>

To implement the multilingual digital networked environment thus embodies two interrelated desires. One is to expand human rights from the political sphere to the economic, social,

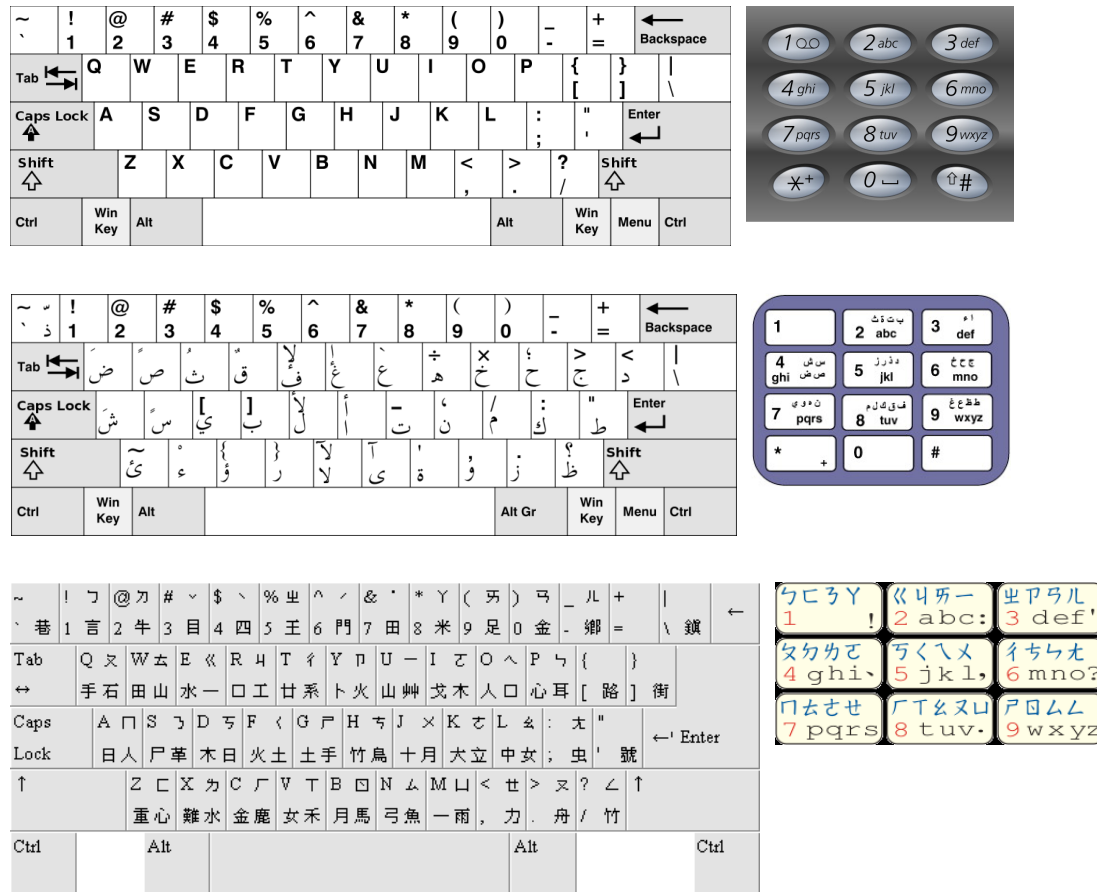
and cultural domains. Another is to diffuse the Internet technologies to different language users in the world. Needing to have a voice online is thus an endeavor to achieve the double meaning of the term “voice” for a group of language users: their expressions and their languages should be accommodated online.

### **LINGUISTIC INFRASTRUCTURE: NOT JUST DESIGNED FOR ENGLISH**

“Will the Internet always speak English?” an American linguist asked in 2002, recognizing the fact that the “Internet was basically an American development” spreading across the English-speaking world at a fast speed.<sup>11</sup> In 2002, the English-speaking world generated nearly 80 percent of Internet traffic. Quoting then French President Jacques Chirac, the American linguist pointed out why the prevalence of English on the Internet had been seen as a “major risk for humanity” that might eventually lead to linguistic and cultural uniformity. A director of a Russian Internet provider, who noticed the distinction between English-speaking elites and non-English-speakers among Russian Internet users, refuted the idea of an “open” Web as “the ultimate act of intellectual colonialism” because it was not yet “open” enough to Russian language and content.<sup>12</sup> These reactions reflected the traditional concerns of print-languages or media-languages as the bases both for non-English voices and a national consciousness.

Indeed, to have a voice in the digital networked environment requires infrastructure that can mechanically support the language in question. To be “mechanically reproduced” in the digital networked environment, a digital language/writing system must be developed to address issues such as encoding (representing language symbols in computer digit codes), linguistic and orthographic factors, and text processing (input, rendering, ordering, etc.). Take keyboards and keypads as examples of text input: Figure 1 shows the respective layouts of American English, Arabic, and Chinese. Languages other than American English have to cram their language symbols in a way

Figure 1. Selected Keypad and Keyboard Layouts for American English, Arabic, and Chinese



isd working papers in new dipomacy

similar to cramming English alphabets onto keypads of twelve keys on mobile devices. Users need to be trained to type the language of their choice. Users without proper language support are effectively voiceless or speechless. Some languages may face many more digital difficulties if the symbol and writing system they require is very different from the Latin-alphabet-based American English.

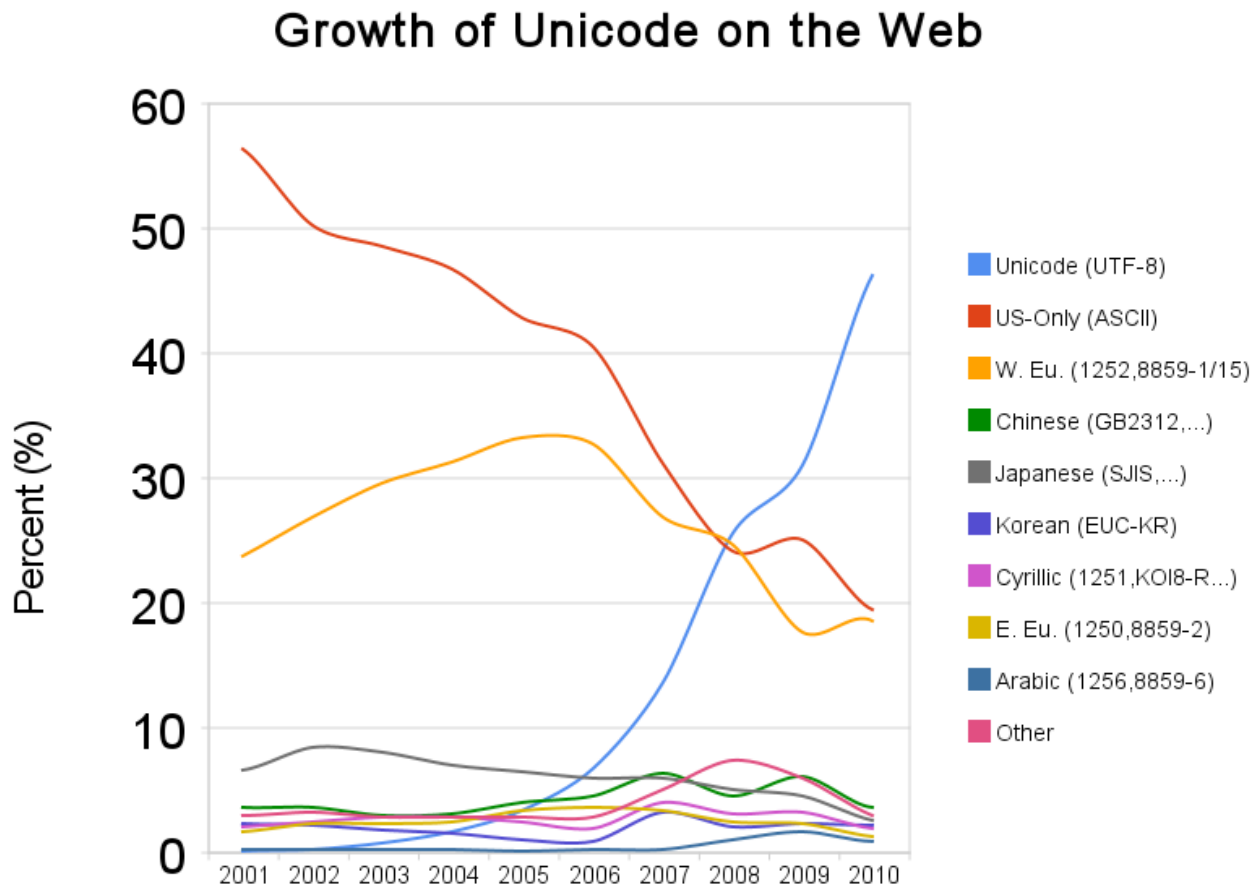
A succinct historical summary of the last decade of linguistic infrastructure is thus needed to demonstrate how languages can be “mechanically reproduced” for all languages in the

world. The basic story is to adopt the American English-only platform to a global one where all languages can coexist.

Targeting all the languages in the world, several international bodies have contributed to efforts in making the digital networked environment ready for world languages to be mechanically reproduced. Recognizing the issue as early as in the late 1990s, the Internet Engineering Task Force (IETF), the main international technical community concerned with Internet architecture and related standards, has declared that the “Internet is international,” and thus it is “an absolute requirement to interchange data in a multiplicity of languages, which in turn utilize a bewildering number of characters.”<sup>13</sup> The actual encoding standardization efforts for world languages, mainly coordinated by a nonprofit organization called the Unicode Consortium,<sup>14</sup> with full members mostly from major computing and Internet companies,<sup>15</sup> have resulted in an international industry standard called the Unicode Standard with the aim of handling all the world languages.<sup>16</sup> On the basis of providing a consistent encoding standard for all world languages, the Unicode Consortium also maintains a repository that provides key building software components for “locale” parameters. These define not only the language settings of users but also the country, date, currency, and other settings.<sup>17</sup> Furthermore, with all these specifications, major software and Internet companies can thus provide software and Web services that meet the different language, regional, and technical requirements of a target market.<sup>18</sup>

Thus it can be argued that the linguistic infrastructure of the digital networked environment has shifted from the monolingual to the multilingual. One example of such a shift is the adoption of the encoding standards on the Web. According to the global search engine company Google, the proportion of the textual materials that uses Unicode on the Web has already surpassed all other encoding standards<sup>19</sup> and is expected to reach half of the total Web in 2010 (see Figure 2).<sup>20</sup> As shown in Figure 2, the other major encoding standards such as ASCII (for American English only), Windows-1252, or ISO/IEC 885901 (for western European languages mostly) are in clear

Figure 2. Growth of Unicode on the Web (by Google), 2001–2010



Source: Mark Davis, 2008.

decline in terms of proportion of Web usage.<sup>21</sup> What Figure 2 demonstrates is a clear trend of the Web adopting a universal standard that can potentially accommodate all the languages in the world, not just English or Latin alphabet-based languages. The progress in adopting such a multilingual standard in the past decade has so far exceeded the IETF’s original expectation, which estimated that it might take at least fifty years.<sup>22</sup>

However, the shift from the monolingual to the multilingual in the basic encoding standard does not automatically

suggest the web's readiness for supporting all the languages in the world. Further empirical work is required to see which languages are supported and to what extent: these questions are beyond the scope of this essay. We can ask, however, who should adopt the technologies for different language users? Of course, one can develop interfaces and applications that are designed for just one individual user; however, it would be too costly and inefficient to serve such an individual's unique language needs. It is, then, no surprise that developers of digital networked technologies will have to put target and potential users in a box, essentially grouping them into a category so that they can be served in bulk. Users belonging to the same category thus would share the same linguistic (and sometimes also regional) interface with the digital networked environment. How can these linguistic aspects of technologies produce linguistic groupings in the same way in which the print-languages have laid the bases for "fellow-readers," "a new fixity to language," and "languages-of-power," as mentioned in an opening quote in this essay?

#### **LINGUISTIC GROUPING: CITIZENS AND/OR CONSUMERS?**

In the digital networked environment, what kinds of "fellow-users" are formed, and what are their implications for the idea of an imagined community? The way in which languages are embedded in media technologies may continue to have an impact on our everyday lives, as print-languages have previously laid the bases for a national consciousnesses. Fellow-readers may have become fellow-citizens who usually shared the same or increasingly the same language. The way in which languages are currently embedded in the digital networked environment may be shaping and being shaped by political and economic units. By far the most systematic technological implementation of such language support is what is called "locale" in computing. Locale is a set of parameters that define not only the language settings of users but also the country, date, currency, and other settings.<sup>23</sup> One example of this is the IETF language tags,



**Table 1.** Examples of the IETF Language Tags.

<i>Code</i>	<i>Language</i>	<i>Subtags</i>
Zh	Chinese	Language
zh-Hans	Chinese written with simplified script	language+script
zh-Hant	Chinese written with traditional script	language+script
zh-CN	Chinese used in People’s Republic of China (PRC)	language+region
zh-SG	Chinese used in Republic of Singapore	language+region
zh-HK	Chinese used in Hong Kong	language+region
zh-TW	Chinese used in Taiwan	language+region

maintained by the Internet Engineering Task Force, so that the naming convention of the locale parameters can be shared and used consistently in the digital networked environment.<sup>24</sup>

Using the Chinese language as an example, the linguistic grouping reflects the complications of the political and economic units in Chinese politics. Table 1 shows some of the possible IETF language tags (on which different locales can be built) that are used for the Chinese language. Although regions such as China, Hong Kong, and Taiwan use Chinese as their official language, different IETF language tags have been developed to identify the various kinds of Chinese used. The language identifier “zh” refers to Chinese, without any specification about the choice of script and the preference of region. The IETF tags “zh-Hans” and “zh-Hant” are language tags with subtags “Hans” and “Hant” to identify the target script (Chinese written with simplified script and Chinese written with traditional script, respectively). The IETF tags “zh-CN,” “zh-SG,” “zh-HK,” and “zh-TW” are language tags with subtags “CN,” “SG,” “HK,” and “TW” to identify the target region (Chinese used in China, Singapore, Hong Kong, and Taiwan, respectively).

Is one closer to her/his fellow-users when using the same or a similar set of “locale settings” that usually include a language identifier and region identifier, such as “pt-BR” for Portuguese used in Brazil or “zh-TW” for Chinese used in Taiwan? These are empirical questions that need empirical research for

specific locales, which is beyond the scope of this essay. Still, without unwarranted fears that these linguistic and geographic parameters will constitute the “filter bubble,”<sup>25</sup> it is instructive to note that these locale settings or parameters are the key for certain language and language preferences to be “mechanically reproduced.”<sup>26</sup> For software and Internet companies, these language preferences are essential to providing a basic environment that is acceptable, if not friendly, for a group of consumers. For some other websites, these language preferences reflect their own specific grouping of users based on languages and sometimes regions. For example, in the Chinese-language online space, Baidu Baike, the equivalent product of Chinese Wikipedia, hosted by China’s major search company, supports only simplified script. Chinese Wikipedia, however, has managed to provide the same content for four different categories of users and editors: simplified Chinese used in China, simplified Chinese used in Singapore (and Malaysia), traditional Chinese used in Hong Kong, and orthodox Chinese used in Taiwan.<sup>27</sup> Thus, the motivations and concerns for different kinds of linguistic groupings online seem to reflect the grouping of political and economic units, and the “fellow-users” are thus shaped by the implementation of different “locale parameters” embedded in the digital networked environment.

Major operating systems and websites today have supports for various language and region settings, and different levels of support by different players can show that certain linguistic groups have more presence or significance than others. Wikipedia, with its open and do-it-yourself approach to language versions and interfaces, has 276 languages in 2011, while Facebook only has more than one hundred language interfaces.<sup>28</sup> Google allows users to change their linguistic and geographic preferences, including interface language, search language, location, etc.,<sup>29</sup> with nearly two hundred language interfaces and nearly two hundred country domain versions.<sup>30</sup> It is worth noting that Wikipedia’s do-it-yourself or serve-yourself approach, meaning that users can build their own interfaces on the platforms provided, has also been used by companies such as Facebook<sup>31</sup> and Google.<sup>32</sup>

The fact that users of different languages are encouraged to serve themselves on a global platform leads to the next question: Does the digital networked environment give “a new fixity to language” as in the print era, where print-languages are standardized enough to create a common basis for exchange and communication?

### LINGUISTIC (AND REGIONAL) DIGITIZABILITY

In contrast to the print era, in which the mechanically reproducible print languages are those that can be standardized in a nation-state for a national market, the digital networked environment of the universal kind seems to be more flexible and fluid. If anything has to be “fixed” in this new environment, it may well be the fact that the language in question must be digitized or “fixed” in the digital forms of zeroes and ones.

Digitizability thus opens up space for all kinds of language and regional variations of the digital networked technologies. Partly because of the international efforts to digitize and network all world languages, partly because of the malleability of using digit codes of zeroes and ones to represent all languages, any language in the world seems to be able to be mechanically reproduced in the digital networked environment. It is then not necessary to argue that major Internet companies, even if they are hosted in the United States, can never respect the need to have “national” voices online. It is difficult to discount these international efforts in enabling the Internet to “speak” the language where the voice can be articulated and heard. The multilingual development of the world’s Internet infrastructure should be seen in the overall context of the “internationalization and localization” efforts of the software industry, which is often abbreviated to the numeronyms of “i18n and L10n.”<sup>33</sup>

Going beyond mere character encoding, the efforts in i18n and L10n cover many aspects of software usage that are related to language and regional differences, which include the design and layout of user interface, input, display, time zone, currency, and language preferences. Internationalization of a certain

piece of software suggests that the linguistic and regional aspects of the software are designed and developed in a way that is not limited or specific to any “locale” (the parameters that define the language and regional aspects of computing). In other words, internationalization prepares a piece of software to be independent from configurations of a certain language and/or region so that it can be repurposed to serve other languages and/or regions. Localization of a piece of software refers to repurposing the already internationalized software for various locales. Another way to look at this i18n and L10n process is as follows: The i18n process first makes sure the digital networked environment is built language neutral (and thus not biased toward any specific language), and the L10n process then allows for different kinds of language and region support actually to be implemented. The i18n process aims to accommodate digital bits so that the digital networked environment is flexible enough to provide support for different locales.

Are locales the new “fixities” that are reintroduced back to the digital networked environment? Does the use of locales segment or even fragment the digital networked environment into different worlds that have few connections? For example, does the use of locales such as “zh-CN,” “zh-SG,” “zh-HK,” and “zh-TW” (Chinese used in China, Singapore, Hong Kong, and Taiwan, respectively) create compartments of different Chinese-speaking worlds? A similar question can also be asked about “en-US,” “en-UK,” and “en-IN” (English used in United States, the United Kingdom, and India, respectively). One can even ask further if the locales with a smaller population can thus avoid the print-era fate of being subsumed by the standardization of a major nation-state language. All these questions may require further empirical research, but some educated guesses are provided by understanding how Web searches work linguistically. The linguistic dimension of Internet development is vividly described by linguist Geoffrey Nunberg, who captures the mechanism of search engines as the ultimate voice collector and conversation monitor:<sup>34</sup>

Seen from a Google’s eye view, in fact, the Web is less like a piazza than a souk—a jumble of separate spaces, each with its own isolated chatter.

The search engines cruise the alleyways to listen in on all of these conversations, locate the people who are talking about the subject we're interested in, and tell us which of them has earned the most nods from the other confabulators in the room.

What brings the "isolated chatter" together and accessible is the computer power to match strings of text that articulates human voices. The way search engines manage to listen to and monitor human voices online depends on the shared keywords and thus certain shared language tokens. Even social media such as Twitter and Facebook use hashtags, tags, or simply keywords in certain languages as the underlying mechanism to exchange and aggregate voices online. Thus, it is very likely that linguistic (and regional) digitizability may allow for more diverse usage of major language scripts, such as Chinese characters, English Roman alphabets, standard Arabic script, Russian Cyrillic alphabets, etc., for all kinds of appropriation and adoption by languages such as written Cantonese, Singlish, Egyptian Arabic, modern Mongolian, and so forth.

Bringing fixity to languages, be it in the print era or the current digital networked environment, is thus about power and the politics of voice. If Nunberg is correct in his description of the Web as "less like a piazza than a souk," the potential to connect disparate voices from different corners of the Internet requires people to use the same keyword in certain languages in order to aggregate souks into a temporary piazza-like public forum on the very topic embodied in the language-dependent keyword. For voices to travel from one locale to another, these voices need to be articulated, digitized, aggregated, searched and listened to, with the help of keywords that are language dependent. Since it is probably the first time in human history that multiple languages can coexist in a universal platform, it is more important to listen to the global voices in their various own isolated chatter (e.g., the GlobalVoicesOnline.org) before entering various ongoing conversations. There are alternative methods of listening to Chinese or American voices online then going to the official diplomatic website, as there are multiple new ways to engage existing conversations with the relevant keywords in their own vocabularies.

So far we have answered some of the questions regarding the dynamics of linguistic grouping in the digital networked environment. All kinds of “digit-coded languages” potentially can be potentially mechanically reproduced in the digital networked environment; however, it may be too early to generalize their implications for building a national (or any other kind of) consciousness. Still, in the increasingly multilingual digital networked environment, new groups of fellow-users are formed and potentially can change the idea of an imagined community online in a more fluid fashion.

The remaining question is what kind of languages-of-power are reshaped or created? This essay ends with a discussion of the notion of a Chinese voice versus an American voice to highlight the centrality of needing a voice online. It is not only a language or technical support issue but also a much broader issue about the future of the Internet.

### **LANGUAGES OF POWER: A CHINESE VERSUS AN AMERICAN VOICE?**

Whether a voice can or cannot be articulated online shows the double meaning of the central concept of “voice”: it is about expressions and languages. The very concept can highlight the implications for the current disagreement between the U.S. and Chinese governments on the role of the Internet. The most crucial background information is that, in terms of Internet users, the two countries are seen as the major representatives of the top two languages: English and Chinese.<sup>35</sup> Although the Internet was originally designed for American English,<sup>36</sup> China has the highest number of Internet users in the world,<sup>37</sup> surpassing that of the United States in 2008.<sup>38</sup> Google’s chief executive officer envisioned that before 2014, the Internet will be dominated by Chinese-language content. Despite having an authoritarian regime that can control and censor the democratizing potential of the Internet,<sup>39</sup> China has the highest proportion of users who produce online content most frequently and are the most socially playful online, more so than countries such as the United States in a 2011 survey on global Internet values.<sup>40</sup> These

Chinese Internet users may employ the Internet mostly, if not exclusively, in Chinese.

With such background information in mind, it is then instructive to see the different interpretations of the “voice” online by both governments. For the U.S. government, the focus has been the human right of free expression for those who encounter censorship by their governments, which includes the Chinese government.<sup>41</sup> For Washington, DC, it is about the political and civic rights of the voices online, with the U.S. government as the major agent for Internet freedom.<sup>42</sup> For the Chinese government, the issue has been framed to reassert the collective “Chinese voice” online to exercise state sovereignty on behalf of China. For Beijing, it is about the extended version of economic, social, and cultural rights with the Chinese government as the major agent.

These arguments made by the Chinese government, perhaps more readily accepted by the majority of Chinese citizens and less known to outsiders, can be summarized as follows. First, according to Beijing, the concept of human rights should not be limited to the (sometimes irresponsible) rights of individuals but rather should include “the basic rights of survival and development,” which is almost synonymous with the “national and sovereign rights of a country.”<sup>43</sup> This line of argument strikes a Japanese and Chinese modern interpretation of western liberal terms such as “freedom,” “individualism,” and “rights,” as retaining the negative connotations of “willful irresponsibility” or “a selfish power play” for profits and privilege, embodied in the Chinese characters used for translating these terms.<sup>44</sup> Second, by arguing that “the national language is the most basic linguistic right for the Chinese people,” Beijing emphasizes the collective perspective of linguistic human rights that “the right to learn and use the common spoken and written Chinese” is the most fundamental and important factor for the Chinese people.<sup>45</sup> This elevates the status of “putonghua,” the standardized Chinese with pronunciation of the Beijing dialect and written scripts of simplified Chinese characters. Third, claiming to know the Chinese (language) better, Chinese Internet companies have portrayed foreign companies

such as Google as clueless Caucasians who cannot even understand the sentence structure of the Chinese language.<sup>46</sup> This reinforces the belief that foreign forces care little about the collective benefits of the Chinese people, in contrast to “national champions” who understand and care. Finally, the Chinese government and national companies are the major guardians of national sovereignty or Chinese “Internet sovereignty” to fend off information that “contains contents subverting state power, undermining national unity, infringing upon national honor and interests.”<sup>47</sup> As reported by Chinese official media, the existence of threats to Chinese “Internet sovereignty” is the basis on which well-known foreign social networking sites, such as Facebook, are blocked in China.<sup>48</sup> Almost all international or American social media platforms have Chinese equivalents served mostly in Chinese-language interfaces.<sup>49</sup> These arguments combined present a powerful version of the collective “Chinese voice” in developing and ruling its Internet: expressions of a unified Chinese voice against foreign influence, using the Beijing-defined national language.

Is it fair to argue that major Internet companies, mostly hosted in the United States with a majority of English-speaking users, cannot understand the need for national voices and respect national consciousness? Is it fair to argue that the Chinese voice will be best served when voiced in the Beijing-defined national language? If so, does the Chinese voice include the voices of Tibetans, Uyghurs, and Mongolians, etc? Can non-American English (for example, British English or Indian English) and non-Mandarin Chinese (for example, Cantonese) prosper in the digital networked environment where American English and Mandarin Chinese constitute the majority of content production in respective language scripts online? Who are the new language supporters and authorities when the digital networked environment is our environment? These kinds of questions are expected to be raised across different languages and regions as the global Internet reaches to more people in the world. It is then an absolute requirement that policymakers understand the values and norms for the digital and networked environment to accommodate a multiplicity of languages, in



a way that respects preexisting conversations and the need to have a voice. In particular, the U.S. government and Internet companies should reexamine their roles in accommodating and aggregating local voices, by including much needed cultural and linguistic sensibilities that have been the expert domains of traditional diplomacy and area studies. Languages should not be barriers that constitute a “filter bubble”<sup>49</sup> but valuable resources for mutual respect and mutual understanding, making the digital networked environment fruitful grounds.

## NOTES

1. Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London, New York: Verso, 2006), pp. 45–46.
2. Mark Anderson, “Parlez-vous Facebook?” *IEEE Spectrum* (June 2011), <http://spectrum.ieee.org/at-work/innovation/parlezvous-facebook>.
3. Anderson, *Imagined Communities*, p. 45.
4. Henry Jenkins, *Convergence culture: where old and new media collide* (New York: NYU Press, 2006). <http://books.google.com/books?id=RIRVNikT06YC>; Tim Dwyer, *Media Convergence* (McGraw-Hill International, 2009). [http://books.google.com/books?id=rbovMF7\\_Ft8C](http://books.google.com/books?id=rbovMF7_Ft8C).
5. Anderson, *Imagined Communities*, p. 45.
6. Harald Tveit Alvestrand, “IETF [Internet Engineering Task Force] Policy on Character Sets and Languages,” *IETF* (Best Current Practice) (January 1998), p. 4. <http://tools.ietf.org/html/rfc2277>.
7. *Oxford English Dictionary*, “voice, n.” London: Oxford University Press, 2011. <http://www.oed.com:80/Entry/224334>.
8. Domna C. Stanton, “On Linguistic Human Rights and the United States ‘Foreign’ Language Crisis,” *Profession* (January 1, 2005): 64-79.
9. *Ibid.*
10. UNESCO, Recommendation concerning the promotion and use of multilingualism and universal access to cyberspace, (Paris: UNESCO, General Conference of UNESCO, 32nd session, October 2003). <http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=151952&gp=1&mode=e&lin=1>.
11. Geoffrey Nunberg, “Will the Internet Always Speak English?” *The American Prospect* (November 30, 2002). [http://www.prospect.org/cs/articles?article=will\\_the\\_internet\\_always\\_speak\\_english](http://www.prospect.org/cs/articles?article=will_the_internet_always_speak_english).
12. Michael Specter, “World, Wide, Web - 3 English Words,” *New York Times*, April 14, 1996. <http://query.nytimes.com/gst/fullpage.html?res=9E0DE2DA1139F937A25757C0A960958260&pagewanted=all>.
13. Alvestrand, “IETF Policy,” p. 1.
14. Unicode Inc, “Unicode Standard,” 2011. <http://unicode.org/standard/standard.html>.
15. Unicode Inc, “The Unicode Consortium Members,” April 27, 2011.

<http://www.unicode.org/consortium/memblogo.html>.

16. Unicode Inc, "Unicode Standard."
17. Unicode Inc, "CLDR Project - Unicode Common Locale Data Repository," *Unicode Official Website* (2011). <http://cldr.unicode.org/>.
18. Donald A. DePalma, "Internationalization and Localization," in *Business without borders: a strategic guide to global marketing* (New York: John Wiley and Sons, 2002); Sarmad Hussain and Ram Mohan, "Localization in Asia Pacific," in *Digital Review of Asia Pacific 2007/2008*, ed. Felix Librero and Patricia B. Arinto (Orbicom and the International Development Research Centre (IDRC), 2008). <http://www.idrc.ca/openebooks/377-5/>; Byung-Keun Kim, "Internationalizing the Internet," in *Internationalizing the Internet: the co-evolution of influence and technology* (Northampton, MA: Edward Elgar Publishing, 2005), pp. 1-9; Pan Sorasak and Chriv Kosona, "Cambodia," in *Digital Review of Asia Pacific 2009/2010*, ed. Shahid Akhtar and Patricia Arinto (IDRC, 2010); John Yunker, *Beyond borders: web globalization strategies* (Berkeley: New Riders, 2002).
19. Mark Davis, "Moving to Unicode 5.1," *Official Google Blog* (2008). <http://googleblog.blogspot.com/2008/05/moving-to-unicode-51.html>.
20. Mark Davis, "Unicode nearing 50% of the web," 2010. <http://googleblog.blogspot.com/2010/01/unicode-nearing-50-of-web.html>.
21. It should be noted that the decline of all these non-Unicode encoding standards does not suggest the decline of the language usage. Those Web documents that were originally written only in English, Chinese, or Arabic can be converted from their old language-specific encoding standards to the universal standard of Unicode.
22. Ivestrand, "IETF Policy."
23. Unicode Inc, "CLDR Project."
24. *Wikipedia*, "IETF language tag - Wikipedia, the free encyclopedia," *English Wikipedia*, May 10, 2011. [http://en.wikipedia.org/wiki/IETF\\_language\\_tag](http://en.wikipedia.org/wiki/IETF_language_tag); A. Phillips and M. Davis, "RFC 5646 - Tags for Identifying Languages," The Internet Engineering Task Force (IETF) (September 2009). <http://tools.ietf.org/html/rfc5646>; IANA, "IETF Language Subtag Registry," *The Internet Assigned Numbers Authority (IANA) official Website* (January 11, 2011). <http://www.iana.org/assignments/language-subtag-registry>.
25. Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press HC, 2011.
26. DePalma, "Internationalization and Localization"; Kok Yong Leong, Hai Liu, and Oliver P Wu, "Web Internationalization and Java Keyboard Input Methods," 1998. [http://www.isoc.org/inet98/proceedings/5f/5f\\_2.htm](http://www.isoc.org/inet98/proceedings/5f/5f_2.htm).
27. Han-Teng Liao, "Conflict and Consensus in the Chinese Version of Wikipedia," *IEEE Technology and Society* (2009). [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5072382](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5072382).
28. Anderson, "Parlez-vous Facebook?"
29. Google Support, "Search preferences : Search history and settings— Web Search Help," 2011. <http://www.google.com/support/websearch/bin/static.py?hl=en&page=guide.cs&guide=1224171&answer=35892&rd=1>.
30. Google, "Language Tools," 2011. [http://www.google.com/language\\_tools?hl=en](http://www.google.com/language_tools?hl=en).

31. Anderson, "Parlez-vous Facebook?"
32. Google, "Google in Your Language," 2011. [www.google.com/transconsole](http://www.google.com/transconsole).
33. Hussain and Mohan, "Localization in Asia Pacific."
34. Geoffrey Nunberg, "As Google Goes, So Goes the Nation," *New York Times*, May 18, 2003. <http://www.nytimes.com/2003/05/18/weekinreview/18NUNB.html>.
35. Miniwatts Marketing Group, "Top Ten Internet Languages," *World Internet Statistics* (June 30, 2010). <http://www.internetworldstats.com/stats7.htm>.
36. Nunberg, "Will the Internet Always Speak English?"
37. Central Intelligence Agency, "CIA World Factbook: China" (McLean, VA: CIA, May 17, 2011). <https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html>.
38. David Barboza, "China Surpasses U.S. in Number of Internet Users," *New York Times*, July 26, 2008. <http://www.nytimes.com/2008/07/26/business/worldbusiness/26internet.html>.
39. Evgeny Morozov, *The Net Delusion: The Dark Side of Internet Freedom* (New York, Public Affairs, 2011); Rebecca MacKinnon, "China's 'Networked Authoritarianism,'" *Journal of Democracy* (April 25, 2011).
40. Soumitra Dutta, William H. Dutton, and Ginette Law, "The New Internet World: A Global Perspective on Freedom of Expression, Privacy, Trust and Security Online," *SSRN eLibrary* (2011). [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1810005](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1810005).
41. Hillary Rodham Clinton, "Remarks on Internet Freedom" (Washington, DC: The Newseum, 2010). <http://www.state.gov/secretary/rm/2010/01/135519.htm>; Hillary Rodham Clinton, "Internet Rights and Wrongs: Choices & Challenges in a Networked World" (Washington, DC: George Washington University, 2011). <http://www.state.gov/secretary/rm/2011/02/156619.htm>.
42. U.S. Department of State, "Internet Freedom," *U.S. Department of State Official Website*, 2011. <http://www.state.gov/e/eeb/cip/netfreedom/index.htm>.
43. Zaibang Wang, "The Characteristics of Human Rights and the Challenges for World Human Rights in the 21st Century," *China Society For Human Rights Studies* (2005). <http://www.humanrights.cn/zt/magazine/200402004826120616.htm>.
44. John King Fairbank and Denis Crispin Twitchett, *The Cambridge History of China* [electronic resource] (United Kingdom: Cambridge University Press, 1978), p. 5.
45. Jinzhi Su, "Protection of Linguistic Rights in China," *China Society for Human Rights Studies* (2004). <http://www.humanrights.cn/zt/magazine/200402004826120616.htm>.
46. Brad Stone and Bruce Einhorn, "How Baidu Won China," *BusinessWeek: Online Magazine*, November 11, 2010. [http://www.businessweek.com/magazine/content/10\\_47/b4204060242597\\_page\\_6.htm](http://www.businessweek.com/magazine/content/10_47/b4204060242597_page_6.htm).
47. People's Republic of China, "China's Internet White Paper: The

Internet in China” (Beijing: Information Office of the State Council of the People’s Republic of China, 2010). [http://china.org.cn/government/whitepaper/node\\_7093508.htm](http://china.org.cn/government/whitepaper/node_7093508.htm).

48. PeopleDaily.com.cn, “White paper explains ‘Internet sovereignty,’” *People’s Daily Online*, June 9, 2010. <http://english.peopledaily.com.cn/90001/90776/90785/7018637.html>.

49. Thomas Crampton, “Infographic: China’s Social Media Evolution,” March 2, 2011. <http://www.thomascrampton.com/china/china-social-media-evolution/>.

50. Pariser, Eli. *The Filter Bubble*.

## REFERENCES

- Alvestrand, Harald Tveit. “IETF Policy on Character Sets and Languages.” *IETF (Best Current Practice) (January 1998)*. <http://tools.ietf.org/html/rfc2277>.
- . “IETF Policy on Character Sets and Languages.” *IETF (Best Current Practice) (January 1998)*. <http://tools.ietf.org/html/rfc2277>.
- Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London, New York: Verso, 2006.
- Anderson, Mark. “Parlez-vous Facebook?” *IEEE Spectrum* (June 2011). <http://spectrum.ieee.org/at-work/innovation/parlezvous-facebook>.
- Barboza, David. “China Surpasses U.S. in Number of Internet Users.” *New York Times*, July 26, 2008. <http://www.nytimes.com/2008/07/26/business/worldbusiness/26internet.html>.
- Central Intelligence Agency. “CIA World Factbook: China.” McLean, VA: CIA, May 17, 2011. <https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html>.
- Clinton, Hillary Rodham. “Internet Rights and Wrongs: Choices & Challenges in a Networked World.” Washington, DC: George Washington University, 2011. <http://www.state.gov/secretary/rm/2011/02/156619.htm>.
- . “Remarks on Internet Freedom.” Washington, DC: The Newseum, 2010. <http://www.state.gov/secretary/rm/2010/01/135519.htm>.
- Crampton, Thomas. “Infographic: China’s Social Media Evolution—Thomas Crampton.” March 2, 2011. <http://www.thomascrampton.com/china/china-social-media-evolution/>.
- Davis, Mark. “Moving to Unicode 5.1.” *Official Google Blog*, 2008. <http://googleblog.blogspot.com/2008/05/moving-to-unicode-51.html>.
- . “Unicode nearing 50% of the web,” 2010. <http://googleblog.blogspot.com/2010/01/unicode-nearing-50-of-web.html>.
- DePalma, Donald A. “Internationalization and Localization.” In *Business without borders: a strategic guide to global marketing*. New York: John Wiley and Sons, 2002.
- Dutta, Soumitra, William H. Dutton, and Ginette Law. “The New Internet World: A Global Perspective on Freedom of Expression, Privacy, Trust and Security Online.” *SSRN eLibrary* (2011). [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1810005](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1810005).
- Dwyer, Tim. *Media Convergence*. McGraw-Hill International, 2009. <http://>

- books.google.com/books?id=rbovMF7\_Ft8C.
- Fairbank, John King, and Denis Crispin Twitchett. *The Cambridge history of China* [electronic resource]. United Kingdom: Cambridge University Press, 1978.
- Google. "Google in Your Language," 2011. [www.google.com/transconsole](http://www.google.com/transconsole).
- . "Language Tools," 2011. [http://www.google.com/language\\_tools?hl=en](http://www.google.com/language_tools?hl=en).
- Google Support. "Search preferences : Search history and settings - Web Search Help," 2011. <http://www.google.com/support/websearch/bin/static.py?hl=en&page=guide.cs&guide=1224171&answer=35892&rd=1>.
- Hussain, Sarmad, and Ram Mohan. "Localization in Asia Pacific." In *Digital Review of Asia Pacific 2007/2008*. Edited by Felix Librero and Patricia B. Arinto. Orbicom and the International Development Research Centre (IDRC), 2008. <http://www.idrc.ca/openebooks/377-5/>.
- IANA. "IETF Language Subtag Registry." The Internet Assigned Numbers Authority (IANA) official Website, January 11, 2011. <http://www.iana.org/assignments/language-subtag-registry>.
- Jenkins, Henry. *Convergence culture: where old and new media collide*. New York: NYU Press, 2006. <http://books.google.com/books?id=RIRVNikT06YC>.
- Kim, Byung-Keun. "Internationalizing the Internet." In *Internationalizing the Internet: the co-evolution of influence and technology*, 1-9. Northampton, MA: Edward Elgar Publishing, 2005.
- Leong, Kok Yong, Hai Liu, and Oliver P Wu. "Web Internationalization and Java Keyboard Input Methods," 1998. [http://www.isoc.org/inet98/proceedings/5f/5f\\_2.htm](http://www.isoc.org/inet98/proceedings/5f/5f_2.htm).
- Liao, Han-Teng. "Conflict and Consensus in the Chinese Version of Wikipedia." *IEEE Technology and Society* (2009). [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5072382](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5072382).
- Miniwatts Marketing Group. "Top Ten Internet Languages." *World Internet Statistics*, June 30, 2010. <http://www.internetworldstats.com/stats7.htm>.
- Nunberg, Geoffrey. "As Google Goes, So Goes the Nation." *New York Times*, May 18, 2003. <http://www.nytimes.com/2003/05/18/weekinreview/18NUNB.html>.
- . "Will the Internet Always Speak English?" *The American Prospect* (November 30, 2002). [http://www.prospect.org/cs/articles?article=will\\_the\\_internet\\_always\\_speak\\_english](http://www.prospect.org/cs/articles?article=will_the_internet_always_speak_english).
- Oxford English Dictionary. "voice, n." London: Oxford University Press, 2011. <http://www.oed.com:80/Entry/224334>.
- PeopleDaily.com.cn. "White paper explains 'Internet sovereignty'." *People's Daily Online*, June 9, 2010. <http://english.peopledaily.com.cn/90001/90776/90785/7018637.html>.
- People's Republic of China. "China's Internet White Paper: The Internet in China." Beijing: Information Office of the State Council of the People's Republic of China, 2010. [http://china.org.cn/government/whitepaper/node\\_7093508.htm](http://china.org.cn/government/whitepaper/node_7093508.htm).
- Phillips, A., and M. Davis. "RFC 5646 - Tags for Identifying Languages." *The Internet Engineering Task Force (IETF)*, September 2009. <http://tools.ietf.org/html/rfc5646>.
- Sorasak, Pan, and Chriv Kosona. "Cambodia." In *Digital Review of Asia Pacific*

- 2009/2010. Edited by Shahid Akhtar and Patricia Arinto. IDRC, 2010.
- Specter, Michael. "World, Wide, Web - 3 English Words." *New York Times*, April 14, 1996. <http://query.nytimes.com/gst/fullpage.html?res=9E0DE2DA1139F937A25757C0A960958260&pagewanted=all>.
- Stanton, Domna C. "On Linguistic Human Rights and the United States 'Foreign' Language Crisis." *Profession* (January 1, 2005): 64-79.
- Stone, Brad, and Bruce Einhorn. "How Baidu Won China." *BusinessWeek: Online Magazine*, November 11, 2010. [http://www.businessweek.com/magazine/content/10\\_47/b4204060242597\\_page\\_6.htm](http://www.businessweek.com/magazine/content/10_47/b4204060242597_page_6.htm).
- Su, Jinzhi. "Protection of Linguistic Rights in China." *China Society for Human Rights Studies* (2004). <http://www.humanrights.cn/zt/magazine/200402004826120616.htm>.
- U.S. Department of State. "Internet Freedom." U.S. Department of State Official Website, 2011. <http://www.state.gov/e/eeb/cip/netfreedom/index.htm>.
- UNESCO. *Recommendation concerning the promotion and use of multilingualism and universal access to cyberspace*. Paris: UNESCO, General Conference of UNESCO, 32nd session, October 2003. <http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=151952&gp=1&mode=e&lin=1>.
- Unicode Inc. "CLDR Project - Unicode Common Locale Data Repository." Unicode Official Website, 2011. <http://cldr.unicode.org/>.
- . "The Unicode Consortium Members," April 27, 2011. <http://www.unicode.org/consortium/memblogo.html>.
- . "Unicode Standard," 2011. <http://unicode.org/standard/standard.html>.
- Wang, Zaibang. "The Characteristics of Human Rights and the Challenges for World Human Rights in the 21st Century." *China Society for Human Rights Studies* (2005). <http://www.humanrights.cn/zt/magazine/200402004826120616.htm>.
- Wikipedia. "IETF language tag—Wikipedia, the free encyclopedia." *English Wikipedia*, May 10, 2011. [http://en.wikipedia.org/wiki/IETF\\_language\\_tag](http://en.wikipedia.org/wiki/IETF_language_tag).
- Yunker, John. *Beyond borders: web globalization strategies*. Berkeley: New Riders, 2002.

Institute for the Study of Diplomacy  
Edmund A. Walsh School of Foreign Service  
Georgetown University  
Washington, DC 20057

telephone 202-965-5735

fax 202-965-5652

Web site <http://isd.georgetown.edu>