Rebuilding Babel: Copyright and the Future of Online Machine Translation

Erik Ketzan[*]

   Machine translation (MT) is computer technology that translates human languages. The legal implications of MT have yet to be considered by legal scholars or technologists because the technology, in its current state, is too far from functional to create actual legal problems. This Article predicts that if MT ever evolves to "good enough," it will create massive copyright infringement on an unprecedented global scale. The Article argues that MT, specifically online MT, needs to be protected from litigation because it is socially, politically, and commercially beneficial. Online MT has the power to revolutionize communication by eliminating language barriers, bridging the gap between different cultures, providing services to minority language speakers, and transforming global e-commerce by allowing even the smallest online vendor to serve the international market. I argue for protection through the creation of effective licenses and statutory clarification of online MT's noninfringing nature.

I.   INTRODUCTION

>   *"Mi Casa es Su Casa."* (traditional Spanish expression) *"My House is Its House."* (translated to English using MT)[1]

      Through technology, mankind has split the atom and cracked the human genome, but creating a computer program that reliably translates human languages has proven more stubborn than any of the last century's scientific challenges. MT is technology that analyzes the text in one language, a "source text," and attempts to produce an equivalent text, a "target text," in another language. The legal implications of MT have yet to be fully considered by legal scholars or technologists because the technology, in its current state, is too far from functional to create actual legal problems.

      What will happen, though, when online machine translators are able to produce accurate results? Suddenly even the most hopeless student of languages could easily check the news at *Le Monde*, comment on a blog written in Russian, or chat with a Japanese-language speaker in real-time. The makers and proponents of MT have been promising this fantastic future for decades, and have so far failed to deliver, but MT, like most impossible technological problems, may eventually be solved.

      The legal problem is that under American copyright law and our international treaties, a translation of any text is considered a derivative work and translators must obtain permission from the derivative right holder to create translations of the original text.[2] There is a danger, then, that when online translators finally become perfected, whether that happens next year or next decade, they will create massive copyright infringement on an unprecedented global scale through the creation of unauthorized derivative works.

      This Article argues that the law needs to pave the way for companies to develop online translators and forestall a chilling effect on innovation that may result from legal uncertainty; software companies may not pursue online translators because of the threat of litigation. Online MT needs to be protected because it is socially, politically, and commercially beneficial. Technology may have put man on the moon, but MT has the potential to take us farther, across the gulf of comprehension that lies between people from different places.

---

      1.   Using the machine translation of Google, Inc., http://www.google.com/language_tools?hl=en (last visited Aug. 6, 2006).
      2.   17 U.S.C. § 101 (2000).

II.   BACKGROUND

A.   *A (Brief) History of the Translation Derivative Right*

Unauthorized translation is a concept that predates the first copyright laws. The Protestant Reformation turned, in large part, upon translations of the Bible unauthorized by the Catholic Church: Martin Luther was excommunicated a year before publishing the New Testament in German in 1522, while the scholar William Tyndale was strangled and burned at the stake for his English translation of the Bible in 1536.[3] But such examples from the prehistory of copyright arose from censorship by church and state, not the property rights of authors.

The concept that authors have a right to control translations of their work is a historically recent one. The 1710 Statute of Anne, England's first copyright law, gave an author of a book the exclusive right to print *that book* for a fixed number of years, but did not explicitly grant authors any right or control over translations. The Statute of Anne was designed to protect only British authors and booksellers publishing in Britain.[4] Foreign authors and their works (usually written in other languages, obviously) were explicitly excluded from protection by the Statute of Anne and by most copyright systems that emerged in Europe in the centuries that followed.[5]

Because of this disregard for foreign authors, translation across Europe throughout the eighteenth and nineteenth centuries remained a chaotic free-for-all. Although translators often sought permission of authors out of "literary courtesy,"[6] authors maintained effectively zero control over foreign translations and the poorly paid translators (often needy scholars) who abridged, adapted, censored, and altered their works to suit the demands of censors, publishers, and the book-buying public.[7] The law regarding translation, meanwhile, was in equal disarray. In the dicta of two nineteenth century cases, English courts did not consider translations to be copies under the Statute of Anne, but no case addressed

---

    3.   *See* THE RADICAL REFORMATION (Michael G. Baylor ed. & trans., 1991); DAVID DANIELL, WILLIAM TYNDALE: A BIOGRAPHY (1994).

    4.   *See* AUGUSTINE BIRRELL, SEVEN LECTURES ON THE LAW AND HISTORY OF COPYRIGHT IN BOOKS 27 (August M. Kelly Publishers 1971) (1899) ("[Britain's] original copyright conceptions were wholly municipal. We wished to protect the British author, who published in Britain on British paper a book printed by British printers. . . . As for foreign books, we had no great opinion of them; but were they worth translating, Grub Street was full of needy scholars who for a few shillings would place their contents before the British reader, but the idea of paying for the privilege was quite alien to the mind of the trade . . . .").

    5.   JULIE E. COHEN ET AL., COPYRIGHT IN A GLOBAL INFORMATION ECONOMY 49 (2004).

    6.   BIRRELL, *supra* note 4, at 155.

    7.   TRANSLATORS THROUGH HISTORY 213 (Jean Delisle & Judith Woodsworth eds., 1995).

the issue head-on.[8]   The Act of 1842, which controlled copyright in England throughout the Victorian era, made no mention of translations, and although English legal commentators of the nineteenth century began to advocate for inclusion of a translation right, the right was not made concrete in the United Kingdom until the 1911 Copyright Act.[9]

American courts did not recognize unauthorized translations as copyright violations until the late nineteenth century.[10]   In *Stowe v. Thomas*, a Pennsylvania district court held that an unauthorized German translation of *Uncle Tom's Cabin* (German being commonly spoken in Pennsylvania) did not constitute a "copy" under copyright law.[11] Congress explicitly reversed this holding in the 1870 Copyright Act, which recognized a form of derivative right in translations by providing that "authors may reserve the right to dramatize or to translate their own works."[12]   The 1909 Act maintained and expanded this translation derivative right, granting authors the right to "translate the copyrighted work into other language or dialects."[13]

International copyright law on the translation right was unclear until the issue was considered and addressed by the Berne Convention of 1886.[14]   After considerable debate, the Convention reached a compromise by granting authors of "Berne works" the exclusive right to make or authorize translations of their work for ten years.[15]   The 1896 Berne Revision extended the translation right to the entire copyright term, provided the author published an authorized translation within ten years

---

8.   *See* Burnett v. Chetwood, (1720) 35 Eng. Rep. 1008, 1008-09 (Ch.) ("[A] translation might not be the same with the reprinting the original, on account that the translator has bestowed his care and pains upon it . . . ."); Millar v. Taylor, (1769) 98 Eng. Rep. 201, 205 (K.B.) (1769) ("Certainly bonâ fide imitations, translations, and abridgments are different [from copies]; and, in respect of the property, may be considered as new works: but colourable and fraudulent variations will not do.").

9.   Lionel Bently, Copyright, Translations and Relations Between Britain and India, 1880-1914, at 8, 22 (July 2005), unpublished draft, *available at* http://cipil.law.cam.ac.uk/ activities/events.php.

10.   Naomi Abe Voegtli, *Rethinking Derivative Rights*, 63 BROOK. L. REV. 1213, 1233 (1997).

11.   Stowe v. Thomas, 23 F. Cas. 201, 207 (C.C.E.D. Pa. 1853) ("A 'copy' of a book must, therefore, be a transcript of the language in which the conceptions of the author are clothed; of something printed and embodied in a tangible shape. The same conceptions clothed in another language cannot constitute the same composition, nor can it be called a transcript or 'copy' of the same 'book.'").

12.   Act of July 8, 1870, ch. 230, § 86, 41st Cong., 2d Sess., 16 Stat. 198.

13.   Act of March 4, 1909, ch. 320, § 1(b), 60th Cong., 2d Sess., 35 Stat. 1075.

14.   Bently, *supra* note 9, at 13.

15.   *Id.* at 14; Berne Convention for the Protection of Literary and Artistic Works, art. 5, Sept. 9, 1886, 102 Stat. 2853, 1161 U.N.T.S. 3.

of publication of the original work,[16] and the 1908 Berlin Revision to the Convention finally assimilated the translation right into general reproduction rights.[17]

Today, derivative works are accorded copyright protection in the United States under section 102 of the Copyright Act, in which a "derivative work" is defined as "a work based upon one or more preexisting works, such as a *translation*, musical arrangement, dramatization, fictionalization."[18]

B.   A (Brief) History of Machine Translation

MT research began in earnest in the years after World War II, although a few earlier thinkers and scientists had foreseen its creation.[19] In 1949, Warren Weaver, the director of the Rockefeller Foundation's natural sciences division, sent a short but visionary memorandum outlining the possibility of translation by machine to 200 of the best minds of his generation.[20] Weaver's memo brought MT to general notice among the government and scientific community, and within a handful of years, MT programs had been launched at MIT, UCLA, the National Bureau of Standards, the University of Washington, and the Rand Corporation.[21] In 1951, Yehoshua Bar-Hillel became the nation's first full-time researcher of MT at MIT, where the first MT conference was held a year later.[22]

Throughout the fifties, MT research proliferated in the United States, the Soviet Union, Japan, and Europe.[23] Research in the United States focused on Russian-English translation, while Soviet research focused on the reverse. The period was marked by extremes of expectation and disillusion. Some scientists predicted that MT would be perfected within a generation, while others dismissed MT as an insurmountable challenge.[24] Bar-Hillel himself dramatically quit

---

16.   BIRRELL, *supra* note 4, at 33.
17.   Bently, *supra* note 9, at 22.
18.   17 U.S.C. §§ 101-102 (2000) (emphasis added).
19.   For the prehistory of MT, see W. JOHN HUTCHINS, MACHINE TRANSLATION: PAST, PRESENT, FUTURE (1986).
20.   MAKOTO NAGAO, MACHINE TRANSLATION, HOW FAR CAN IT GO? 19 (Norman D. Cook trans., 1989).
21.   Steve Silberman, *Talking to Strangers*, WIRED, May 2000, at 228.
22.   W. John Hutchins, *Machine Translation: A Brief History*, *in* CONCISE HISTORY OF THE LANGUAGE SCIENCES 431-45 (E.F.K. Koerner & R.E. Asher eds., 1995).
23.   NAGAO, *supra* note 20, at 20-21.
24.   *See* Silberman, *supra* note 21, at 225.

research in the MT field after declaring in 1960 that fully automated, high-quality MT was categorically impossible.[25]

      In 1964, the U.S. Academy of Science created an Automatic Language Processing Advisory Committee (ALPAC) to assess the state and potential of MT research. The Committee delivered the *ALPAC Report* a year later, concluding that MT was and would remain inferior to human translation in quality and cost in the near future. As a direct result of the *ALPAC Report*, funding for MT research in the United States dwindled to almost nothing, and the period from 1965 to 1975 has been called the "Dark Ages of machine translation."[26]

      MT research in the United States remained stagnant in the decade after 1975, but revived in Europe, because of the administrative and legal translation needs of the European Community, and in Japan, due to the technical translation needs of Japan's rapidly growing electronics and engineering industries.[27] Beginning in 1976, the European Community employed the private company Systran to assist in the translation of documents across all its member nations' languages. Systran and competitors like Logos Corporation built their business on general application MT systems, while other companies like the Smart Corporation thrived on tailor-made systems for large companies like Citicorp and Ford.[28] Throughout the 1980s, MT research advanced rapidly and the commercial market for MT expanded in new directions.

      In the 1990s, the commercial use of MT accelerated, primarily due to the needs of commercial agencies, governments, and multinational companies, especially large computer software makers, which sold to international markets and demanded speedy translation of software and documentation. In 1993, over 300 million words (about 30,000 documents the length of the Article you are reading) were translated by the major MT companies.[29] MT research has revived and thrived in the United States as the commercial potential of MT has grown since the 1980s, and a number of universities and companies have major research groups devoted to computers and natural languages.[30]

---

      25.   NAGAO, *supra* note 20, at 25.
      26.   *Id.* at 26-28.
      27.   *Id.* at 43-46.
      28.   *See* Hutchins, *supra* note 22, at 437.
      29.   *Id.* at 443.
      30.   Notable academic centers of MT research in the United States include Carnegie Mellon University's Language Technologies Institute, the Center for Language and Speech Processing at the Johns Hopkins University, the University of Southern California's Information Sciences Institute, and Computational Linguistics and Information Processing Laboratory (CLIPS) at the University of Maryland. For corporate and international MT research leaders, see Carl Zimmer, *Universal Translators*, WIRED, May 2000, at 234-35.

The rise of the Internet created a new market for MT services: the translation of Web pages, e-mail, and chat messages. A number of established players vied with newly formed (including many short-lived) startups, and Systran has emerged as the market leader of online MT.[31] Systran currently licenses its technology to power online translation provided by Google, Apple, AOL, Yahoo, and many others.[32]

Today, MT is used by a variety of governments, businesses, and organizations, but despite quantum leaps in computer technology and frequent predictions that usable MT is just around the corner, MT is still largely unusable for Internet users looking to translate most Web pages or documents on the fly. In his 2000 State of the Union address, President Clinton promised that, "[s]oon researchers will bring us devices that can translate foreign languages as fast as you can talk," to which *Wired* magazine succinctly replied, "Why does the future of MT never seem to arrive?"[33]

MT may be an impossible problem that will never be solved, but rather than sit around idly "waiting for Godot," the law should set the stage in anticipation of this technology's possible arrival.

C.   The Technology and How It Works

A galaxy of research and literature exists on the science of MT, but the present study is less interested with *how* MT works as long as it eventually does. With that in mind, a brief summary is beneficial to understanding how recent advances may soon bring about the long-promised Holy Grail: usable MT.

MT systems from their inception until the 1970s were largely "rule-based." Programmers create rule-based systems through a complex set of algorithms which rely on linguistic theory and bilingual dictionaries.[34]

The first type of rule-based MT developed employed the "direct translation" approach: the MT system was designed specifically for one pair of languages, e.g., French and English, and the vocabulary and syntax of texts were analyzed to resolve ambiguities between the

---

    31.   Some now-defunct companies that tested the waters of online MT include Wholetree, Lernout & Hauspie, e-lingo, and Logos. Mary Flanagan & Steve McCure, *SYSTRAN and the Reinvention of MT*, INT'L DATA CORP. BULL., Jan. 2002, *available at* http://www.SYSTRANsoft. com/IDC/26459.html.

    32.   Although Google is developing its own MT in 2006, it appears to still license from SYSTRAN. *See* About SYSTRAN, http://www.SYSTRANsoft.com/company/ (last visited Aug. 6, 2006).

    33.   Address Before a Joint Session of the Congress on the State of the Union, 36 WEEKLY COMP. PRES. DOC. 160 (Jan. 27, 2000); Silberman, *supra* note 21, at 226.

    34.   This part is based largely on Hutchins, *supra* note 22, at 431-35.

language.   For instance, one of a thousand rules added directly by programmers specified that "you" could be translated into French as "vous" (formal/plural) or "tu" (informal singular).

The second type of rule-based system is the "interlingua" approach, which creates an additional step in the analysis with the aim of translation between many languages.[35]   This approach first converts the source language text into a universal language-independent vocabulary, an interlingua, which can then theoretically be converted to any other language.   This interlingua, although not yet achieved by MT researchers, would theoretically function as a kind of perfect language, an intermediary from which any language could be translated into all others.

The third rule-based system is the "transfer" approach.   Whereas the interlingua approach attempts to resolve all ambiguities in a source text so that translation to any language is possible, the transfer approach only tackles those ambiguities inherent in the source language; no universal interlingua, a daunting technological problem, is required.

Since the late 1980s, great strides have been made in corpus-based, as opposed to rule-based, methods.[36]   Corpus-based methods are so called because they use statistical methods and huge libraries of translated texts (for instance from the European Union) to perform translation.   In the first corpus-based research of the late 1980s and early 1990s, no linguistic rules were applied; translation was accomplished solely through statistical analysis.   The increasing speed of computer processors and the rise of the Internet aided the development of corpus-based MT research, which has emerged as the most promising hope for achieving usable MT.[37]

Since 2001, the National Institute of Science and Technology (NIST) has conducted a comparative test of MT systems, the NIST Machine Translation Evaluation.[38]   In 2005, the evaluation tested the ability of systems by most of the world's leading MT developers to translate newswire documents from Arabic and Chinese into English, many employing corpus-based MT.   In both language tests, Google, Inc. scored highest by a significant margin.   Lead engineer Franz Och stated that feeding Google's software with text that equated to one million

---

      35.   *Id.*
      36.   *Id.*
      37.   *Tongues of the Web*, THE ECONOMIST, Mar. 16, 2002, at 26.
      38.   *NIST 2005 Machine Translation Evaluation Official Results*, http://www.nist.gov/ speech/tests/mt/mt05eval_official_results_release_20050801_v3.html (last visited Aug. 6, 2006).

books was key to their performance.[39] Corpus-based research, therefore, may hold the key to finally unlocking usable MT.

    In addition to translating text, researchers have been making inroads on MT for spoken language, literally the stuff of science fiction.[40] The realistic prospect of spoken language translators began in the early 1990s, and the technology continues to evolve, although it faces the technological challenges of both MT and speech recognition.[41] This Article does not speak to the legal consequences of spoken MT, although the legal problems would be related to those discussed here.

D.   How Good Is "Good Enough"?

    The present study predicts that if MT ever evolves to "good enough," legal problems will be there waiting. What do I mean by "good enough"?

    The evaluation of MT systems is neither completely art nor completely science.[42] On the scientific end are automatic metrics like BLEU, recently developed by IBM to quickly and cheaply measure the accuracy of MT output.[43] But since BLEU scores can vary widely, depending on the number of reference translations used in the evaluation,[44] and some researchers doubt whether automatic evaluators are perfectly or even significantly reliable,[45] it is unwise to suggest a solid evaluation number to indicate "good enough" MT for purposes of the present study.

    The MT literature has coined three terms of art to indicate levels of MT output quality, based on usability. The first, "fully automatic high-quality translation" (FAHQT), is a term originating in the early days of

---

    39.   Gary Stix, *The Elusive Goal of Machine Translation*, SCI. AM., Mar. 2006, at 95.
    40.   Two notable examples of translation technology in science fiction are the Universal Translator in *Star Trek*, which translates speech on the fly and serves as a convenient explanation for why the alien races encountered by the crew of the Enterprise speak perfect English, and the "Babel fish" in *The Hitchhiker's Guide to the Galaxy*. *See* MICHAEL OKUDA, STAR TREK ENCYCLOPEDIA 361 (1994); DOUGLAS ADAMS, THE ULTIMATE HITCHHIKER'S GUIDE TO THE GALAXY 42 (2002) ("[I]f you stick a Babel fish in your ear you can instantly understand anything said to you in any form of language.").
    41.   Silberman, *supra* note 21, at 288.
    42.   Derek Lewis, *MT Evaluation: Science or Art?*, *in* TRANSLATING AND THE COMPUTER 19 (1997).
    43.   KISHORE PAPINENI ET AL., BLEU: A METHOD FOR AUTOMATIC EVALUATION OF MACHINE TRANSLATION 2 (2001), *available at* http://domino.watson.ibm.com/library/CyberDig. nsf/home (keyword = RC22176).
    44.   *Id.* at 6.
    45.   Yasuhiro Akiba et al., *Experimental Comparison of MT Evaluation Methods: RED vs. BLEU*, *in* MT SUMMIT IX: PROCEEDINGS OF THE NINTH MACHINE TRANSLATION SUMMIT (2003).

MT research and remains the technology's unfulfilled promise.[46]   The second, "gisting," describes the current state of MT online; MT produces a rough text that allows users to "get the gist" of the source text, but words and phrases are frequently mistranslated.   The reader may not understand that Oswald shot Kennedy and not vice-versa.[47]   Gisting has real-world applications so long as its inherent unreliability is understood, but MT will have to evolve far past gisting to be truly useful on the Internet.

      "Near Human Quality" (NHQ) is a much higher standard than gisting and refers to MT that is grammatically correct, technically accurate, and without major semantic ambiguities, but may be monotonous or stylistically clumsy.[48]   Put another way, no words or idioms are translated *incorrectly*, but the output is not as polished or complete as that of an experienced human translator.[49]   NHQ has been likened to the robotic voice on telephone directories: correct and comprehensible, but clearly artificial.[50]   NHQ is currently available for certain uses and language pairs.

      Translation, by definition, is never perfect.   The full range of nuance and connotation in one language can never be completely replicated in another, but despite this, translation happens.[51]   Someone reading *War & Peace* in Chinese will have a different experience than another reading it in Russian, but both will agree that Prince Andrei dies at the end.   To become a truly useful tool for everyday online users, MT similarly does not need to be perfect.   Merely *viable*.   The present study predicates its predictions on the future arrival of "viable," which this Article defines as better than the current gisting, not quite NHQ, and not improved by pre- or post-editing by humans.

---

      46.   Stix, *supra* note 39, at 95.
      47.   *Id.*
      48.   The term "Near Human Quality" was coined and described in Terence Lewis, *Can We Make Do with Near Human Quality?*, *in* TRANSLATING AND THE COMPUTER 17 (1995).
      49.   *Id.*
      50.   *Id.*
      51.   Philosophers and writers have grappled with the impossibility of translation for centuries. *See, e.g.*, UMBERTO ECO, EXPERIENCES IN TRANSLATION, at ix (Alastair McEwen trans., 2001) ("Every sensible and rigorous theory of language shows that a perfect translation is an impossible dream.   In spite of this, people translate.   It is like the paradox of Achilles and the turtle.   Theoretically speaking, Achilles should never reach the turtle.   But in reality, he does.   No rigorous philosophical approach to that paradox can underestimate the fact that, not just Achilles, but any one of us, could beat a turtle at the Olympic Games."); DOUGLAS HOFSTADTER, LE TON BEAU DE MAROT: IN PRAISE OF THE MUSIC OF LANGUAGE 291 (1998) ("Despite the grain of truth that resides in the thesis that cross-language, cross-culture substitutability of words is shaky, translation is still possible.").

III. CLAIM

    Online MT may be on the horizon and could lead to copyright infringement on a global scale. This Article argues, however, that the social, political, and commercial benefits of online MT outweigh the cost to society and that creators of MT software must be protected in advance from litigation.

    At the same time, creators of copyrighted material should be able to protect their works. In other words, the law should strive to shield the creators of online MT from lawsuits *and* allow creators who wish to prevent their works from being translated to do so.

A.   *Social and Cultural Benefits of MT*

    Although about eighty percent of Web pages are currently in English, less than a tenth of the world speaks English.[52] MT has the potential to bring one to the other and complete the unifying promise of the Internet, eliminating the barriers of language just as geographical borders disappear online.

    Online MT is social software, or software that facilitates communication.[53] Social software is far more than Harry Potter message boards and teenage chat rooms; it is big business. Venture capitalists invested tens of millions of dollars in the online social networking segment following the new millennium, public companies have followed suit, and the past few years have seen a string of high-profile acquisitions of companies like Myspace.com and Craigslist by media giants such as Fox Interactive and eBay.[54]

    The most straightforward social use of viable MT would be the translation of e-mail and static documents. Software that integrates MT to translate e-mail is already on the market, although naturally limited by the current state of the art. It is possible to imagine an e-mail application or word processor that allows users to change language just as easily as choosing a font.[55] Since MT is neutral social software, such an application could (and already may) be used for unwelcome

    52.   U.N. DEV. PROGRAMME [UNDP], HUMAN DEVELOPMENT REPORT 1999, at 6 (1999).
    53.   For a thorough analysis of social software and law, see generally Michael J. Madison, *Social Software, Groups, and Governance*, 2006 MICH. ST. L. REV. 153 (arguing that as matters of law and policy, society should look for ways to enhance—and, where appropriate, to limit—the social character of computing).
    54.   Scott H. Kessler, *Computers: Consumer Services & The Internet*, *in* 174 STANDARD & POOR'S INDUSTRY SURVEYS 3-5 (Eileen M. Bossong-Martines ed., 2006).
    55.   *See* HOFSTADTER, *supra* note 51, at 119-20 (imagining a word processor with integrated MT as a "dreadful, nightmarish vision").

communications such as spam, but has the potential to foster positive social relationships of any kind online, from social networking and pen pals to (who knows?) cross-lingual online dating.

   Viable MT could also power instant chat on the fly. Chat rooms powered by MT may resemble an online United Nations General Assembly, with dozens of people engaging in mutually comprehensible conversation. Machine-translated chat could also be integrated into online video games, including Massively Multiplayer Online Role Playing Games (MMORPGs), virtual worlds in which players interact to trade in-game items and form teams.[56] In MMORPGs, as in real life, difference in language has led to hate crimes incited by mutual incomprehension.[57] But communication can lead to cooperation, and just as Wellington's army spoke six different languages at Waterloo, players from different countries could use viable MT to create massively multinational virtual armies.[58]

   A natural consequence of increased communication (such as the messaging, chat, and play described above) is the spread of information, or, in the jargon of the dot-com era, "content." The content available for translation online, copyrighted or not, is virtually infinite. Ignoring the legal impact, which is discussed at length *infra*, what would happen to online content if viable MT was achieved?

   In order to understand the changes that viable online MT could bring about, it is necessary to survey the *current* solutions used to translate content online. First, there are Web sites which legally offer translations produced by professional human translators: the Web sites of multinational corporations like Kodak (localized in dozens of languages) and multilingual content providers like the BBC (which offers news stories in over thirty languages).[59]

   Next, there is online MT at today's technological level, available free from numerous Web sites and for sale from others as desktop

---

   56.   There is an extensive body of literature on the social character of online gaming. *See, e.g.*, EDWARD CASTRONOVA, SYNTHETIC WORLDS: THE BUSINESS AND CULTURE OF ONLINE GAMES (2005).
   57.   Most notable is the recent virtual massacre of Chinese gamers by Korean players of the MMORPG, *Lineage*, incited by an alleged breach of online etiquette. *See* Posting of James Ransom-Wiley to Joystiq, http://www.joystiq.com/2006/02/21/korean-gamers-massacre-chinese-over-etiquette-dispute/ (Feb. 21, 2006, 15:00 EST) ("One Korean player claims, 'you can tell they are Chinese because they can't speak Korean.'").
   58.   Until recently, many modern MMORPGs segregated users from the major video game market regions—North America, Japan, and Europe—into discrete virtual worlds, but games are beginning to place users worldwide in the same virtual world.
   59.   *See* Kodak, http://www.kodak.com (last visited Aug. 6, 2006); BBC World Service, http://www.bbc.co.uk/worldservice/languages/index.shtml (last visited Aug. 6, 2006).

software.[60]   Despite the crude level of current MT output, Internet users apparently prefer an imperfect solution to none; as early as 1999, Altavista alone, using software licensed from Systran, processed 500,000 translations daily.[61]   Searches for "online translation" often rank among the top ten most popular reported queries at Google's Chinese[62] and Japanese[63] Web sites, and even this author's modest Web sites and blogs frequently receive visitors using the aid of online translators. A number of monolingual Web sites incorporate links to online MT to provide their content in a variety of languages,[64] a functionality now made simple through easy-to-use online tools.[65]

     Finally, vast networks of amateur human translators have assembled to fill the demand for translated material. Articles on the open-source encyclopedia, Wikipedia, are solicited for translation by volunteers.[66] Global Voices Online, launched by Harvard's Berkman Center, features "roundups" of foreign blogs by bilingual "bridge bloggers," who sum up foreign blogs and translate selections.[67]   Groups of amateurs known as "scanlators" translate and illegally distribute images of Japanese comic books (known as manga), often years before officially licensed copies are sold in the United States.[68]   The "fansub" community does the same with Japanese animation, or anime, posting subtitled anime episodes online mere days or even hours after they air in Japan.[69]   Similar fansub

---

        60.   Flanagan & McCure, *supra* note 31.
        61.   Penelope Patsuris, *Lost in the Translation*, FORBES.COM, Aug. 27, 1999, http://www. forbes.com/1999/08/27/feat.html.
        62.   Searches for "online translation" in Mandarin Chinese ranked among the top ten searches at Google China in August 2004 and September 2005. *See* Google Zeitgeist Archive, http://www.google.com/press/zeitgeist/archive.html (last visited Aug. 6, 2006).
        63.   The Japanese search term is actually *hon'yaku*, meaning simply "translation," but Japanese users were probably looking for online MT given that the first ten search results for this query contain or link to online MT. Searches for *hon'yaku* ranked among the top ten searches at Google Japan seven months in the year 2004 and January and April 2005. *See id.*
        64.   For a survey of how monolingual Web sites integrate online MT, see Federico Gaspari, *Integrating On-Line MT Services into Monolingual Web-Sites for Dissemination Purposes: An Evaluation Perspective* (2004), http://www.mt-archive.info/EAMT-2004-Gaspari. pdf.
        65.   *See* Yahoo! Babel Fish, http://babelfish.yahoo.com/ (last visited Aug. 6, 2006).
        66.   *See* Wikipedia: Translation into English, http://en.wikipedia.org/wiki/Wikipedia: Translation_into_English (last visited Aug. 6, 2006).
        67.   *See* Global Voices Online, http://www.globalvoicesonline.org/top/about-global-voices/ (last visited Aug. 6, 2006).
        68.   *See* Sean Kirkpatrick, *Like Holding a Bird: What the Prevalence of Fansubbing Can Teach Us About the Use of Strategic Selective Copyright Enforcement*, 21 TEMP. ENVTL. L. & TECH. J. 131, 135 n.25 (2003).
        69.   *Id.* at 134-37.

communities exist for Chinese, Korean, and Japanese television shows,[70] Asian cinema,[71] and Japanese video games.[72]

How would better MT supplement or replace the current methods for translating online content: professional localization, MT, and amateur groups? Professional localization could be supplemented by viable MT and perhaps replaced outright in certain situations; an international corporation may want to ensure that its legal terms of service are accurately translated for its best markets, but may offer less important information via MT to save on localization costs. Viable MT would naturally replace the crude MT of today, and MT could aid or replace amateur translation groups online, depending on the proficiency of the current human translators.[73]

In sum, viable MT would definitely increase social interaction across cultures and languages, possibly dramatically. It is easy to imagine a future in which we seamlessly surf the entirety of the Internet in our native language, with occasional (or frequent) linguistic mistakes being the only indication that we are reading machine-translated text.

There are potentially negative social effects of viable MT, mainly that viable MT may replace the desire to actually *learn* other languages. America already lags behind much of the world in language skills; at least 10% of Americans are bilingual,[74] but 40% of people in Luxembourg are *quadrilingual*.[75] With viable MT serving as an adequate gateway to foreign content online, users may forgo experiencing foreign

---

70.   Japanese television dramas in particular command a loyal niche fan base online, who refer to such shows as "dorama," after the Japanese pronunciation of "drama." *See* D-Addicts, http://www.d-addicts.com/forum/ (last visited Aug. 6, 2006).

71.   Fansubs of Asian and other foreign cinema generally appeal to film enthusiasts who import foreign DVDs never (or not yet) licensed in the United States. *See* Kloofy, http://www.kloofy.net/ (last visited Aug. 6, 2006).

72.   As fan translation of video games requires hacking (or altering) video game files known as ROMs, this type of fansub is sometimes called "romhacking." *See* Romhacking.net, http://www.romhacking.net (last visited Aug. 6, 2006).

73.   MT could be used to translate film scripts to create movie subtitles on the fly, with or without post-editing by human translators. On fansubbing, it is worth noting that although some fansubbers pride themselves on releasing better translations than those released commercially by the actual licensees of anime, some fan "translations" are performed by fansubbers with rudimentary or no actual language skills.

74.   According to the 2000 census, 9.8% of U.S. residents speak a minority language at home but also speak English "very well." HYON B. SHIN & ROSALIND BRUNO, U.S. CENSUS BUREAU, LANGUAGE USE AND ENGLISH-SPEAKING ABILITY: 2000, at 4, *available at* http://www.census.gov/prod/2003puls/ckbr_29.pdf (last visited Aug. 6, 2006). The problem is that this does not take into account people who speak a second language outside of the home (i.e., one learned in school or abroad).

75.   John Freivalds, *Self-Study Programs Aid Language Learning*, HR MAG., Jan. 1997, at 58.

travel, music, and literature in the original language. If the result is mutual comprehension, however, does it matter how it is achieved? The novelist and philosopher Umberto Eco has conceived of an ideal linguistic future in which everyone is polyglot, speaking their own language but *understanding* everyone else's.[76] So, Pierre would speak French to Ivan, who would understand and answer in Russian. Although the erudite Eco no doubt dreamed that endless education would bring this future about, MT could ironically achieve much the same result for even the laziest language student.

Although MT will never be perfect, viable MT could bring about what can be described without hyperbole as a revolution in global social interaction.

B.   *Political Benefit of MT*[77]

Language is political in a very real way.[78] Borders and jurisdictions are drawn according to language. Language is often the most contested issue in multicultural states.[79] Minority languages have been violently suppressed throughout history, and even in well-established democracies, monolingual policies are controversial.[80] The debate over bilingual education in the United States is a battlefield.[81] Over twenty of the fifty states have recently created statutes or state constitutional provisions declaring English their official language[82] and there is even debate over whether "The Star-Spangled Banner" should be sung in Spanish.[83]

Politics and international relations in general have driven MT research since its inception. Research in the United States and the

---

76.   Erik Ketzan, Eco at the Folger (Oct. 18, 2002), http://www.themodernword.com/eco/eco_folger02.html.

77.   This Article's distinction between social and political implications of MT is necessarily arbitrary.

78.   A small sampling of the voluminous literature on language and politics includes GEORGE ORWELL, POLITICS AND THE ENGLISH LANGUAGE (1945), *reprinted in* GEORGE ORWELL, A COLLECTION OF ESSAYS 162 (1970); LANGUAGE, A RIGHT AND A RESOURCE: APPROACHING LINGUISTIC HUMAN RIGHTS (Miklós Kontra et al. eds., 1999); AFRICAN LANGUAGES, DEVELOPMENT AND THE STATE (Richard Fardon & Graham Furniss eds., 1994); and NOAM CHOMSKY, LANGUAGE AND POLITICS (2004).

79.   U.N. DEV. PROGRAMME [UNDP], HUMAN DEVELOPMENT REPORT 2004, at 9 (2004).

80.   *Id.*

81.   In the United States, the term "bilingual education" normally refers to programs that help children whose first language is not English learn English through the medium of their native languages. *See* Bilingual Education Act, 20 U.S.C. §§ 7401-7434 (2000) (providing federal financial support for such programs and not, for example, for programs to develop competence in two languages); Lau v. Nichols, 414 U.S. 563 (1974).

82.   Peter W. Schroth, *Language and Law*, 46 AM. J. COMP. L. 17, 17-18 (1998).

83.   *See, e.g.*, Peter Baker, *Administration Is Singing More Than One Tune on Spanish Version of Anthem*, WASH. POST, May 3, 2006, at A6.

U.S.S.R. focused on the English and Russian language pair throughout the Cold War.[84] The CIA, Department of Defense, and Air Force poured funding into early MT research.[85] Throughout the 1970s and 1980s, researchers in Saarbrücken, Germany focused on Russian and German.[86] The first MT project of the Logos Corporation, still in business, was an English-Vietnamese system for translating aircraft manuals during the 1970s.[87] One of the best-known projects of the 1980s was the European Communities' Eurotra project, which aimed to translate among all the Community languages.[88] Systran developed Serbo-Croatian-to-English MT for U.S. forces sent to the former Yugoslavia.[89]

       Research on the military and intelligence applications of MT continues this trend through today. Researchers from Carnegie Mellon and Lockheed Martin collaborated to design wearable translation computers for field testing by the United States Army.[90] In 2005, the NIST Evaluation tested MT on only two strategically important languages: Mandarin Chinese and Arabic.[91] The recent MT startup, Language Weaver, supplies MT to the U.S. government to screen reams of foreign language news broadcasts, chat rooms, and Web sites.[92] Much as we have rapid-response military forces, the Department of Defense is experimenting in "rapid response machine translation" to prepare for a sudden need to translate a "surprise language" (presumably in response to an attack from a previously unsuspected enemy).[93]

       In sum, many of the same companies and institutions that drive MT research for American and European government and intelligence are also serving the private sector. Forestalling a chilling effect on MT research by preempting litigation against these MT companies therefore has the political benefit of better MT available to the military and government.

       Although born and developed in large part as military technology, MT for peace is also not a hypothetical. For example, Wired for Peace is

---

       84.   *See* Silberman, *supra* note 21, at 228-29.
       85.   *Id.* at 229.
       86.   *See* Hutchins, *supra* note 22, at 440.
       87.   *Id.* at 437.
       88.   *Id.* at 438.
       89.   Zimmer, *supra* note 30, at 235.
       90.   *Id.* at 234.
       91.   *See NIST 2005 Machine Translation Evaluation Results*, *supra* note 38.
       92.   Gregory T. Huang, *Translation by Numbers*, TECH. REV., Oct. 2005, at 26.
       93.   Douglas W. Oard & Franz Josef Och, *Rapid-Response Machine Translation for Unexpected Languages*, *in* MT SUMMIT IX, PROCEEDINGS OF THE NINTH MACHINE TRANSLATION SUMMIT (2003), *available at* http://www.amtaWeb.org/summit/MTSummit/FinalPapers/108-Oard-final.pdf.

an international project on "virtual diplomacy." Its goal is to organize regular communication between policymakers and researchers of the United States, China, Russia, Japan, North Korea, and South Korea with online tools to allow for continuous communication on issues regarding regional security and cooperation. To that end, the Wired for Peace Web site offers a library of documents translated via MT and chat with MT built in.[94]

Just as MT promises to bring basic translation needs to unserved markets, MT has the potential to supplement the translation needs of minority language speakers worldwide. Although non-English speakers are usually assured translation and language accommodation in the criminal courtroom, classroom, and voting booth under American law, minority language speakers could benefit from increased access to translation in the areas of transactions and consumer protection.[95] Two-way translation devices could improve communication with minority language speakers in public safety functions like law enforcement and providing health care.[96] More generally, MT could prevent the disenfranchising of linguistic minorities in economically developing nations from full participation in the information age.

MT for minority languages sounds like an altruistic daydream but is actually in development.[97] Researchers at Spanish universities are developing MT for Galician and Basque, stressing the importance of MT to the survival of these languages.[98] A daily Spanish newspaper, *El Periódico*, published a Catalan version online, employing MT developed at the Universitat Politècnica de Catalunya in Barcelona.[99] AVENUE is a

---

   94.   Vladimir Obronko, Wired for Peace and Multi-Language Communication (2000), http://www.mt-archive.info/AMTA-2000-Oboronko.pdf; *see also* Wired for Peace Home Page, http://www.wiredforpeace.org/ (last visited Aug. 6, 2006).
   95.   Steven Bender, *Consumer Protection for Latinos: Overcoming Language Fraud and English-Only in the Marketplace*, 45 AM. U.L. REV. 1027, 1029-30 (1996).
   96.   David Belluomini, *Translation by Machine*, FUTURIST, Mar./Apr. 2006, at 57 (discussing the possible impact of language technology on law enforcement).
   97.   For MT and minority languages in general, see Harold Somers, *Machine Translation and Minority Languages*, *in* TRANSLATING AND THE COMPUTER, *supra* note 42; Vincent Berment, *Several Directions for Minority Languages Computerization* (2002), http://www.cs.mu.oz.au/acl/C/C02/C02-2016.pdf.
   98.   Arantxa Diaz de Ilarraza et al., Building a Lexicon for an English-Basque Machine Translation System from Heterogeneous Wide Coverage Dictionaries, Paper Presented at the British Computer Society's Conference, MT2000: Machine Translation and Multilingual Applications in the New Millenium, University of Exeter (Nov. 20-22, 2000), http://www.mt-archive.info/BCS-2000-Diaz-1.pdf; Inés Diz Gamallo, *The Importance of MT for the Survival of Minority Languages: Spanish-Galician MT System* (2001), http://www.eamt.org/summitVIII/papers.html.
   99.   *See* El Periódico de Catalunya, http://www.elperiodico.com/default.asp?idoma=CAT (last visited Aug. 6, 2006).

project by researchers at Carnegie Mellon's Language Technologies Institute to create MT for indigenous languages of South America, beginning with the Mapundundan language, which is spoken by almost a million people (Mapuche) in Chile and Argentina.[100]   Although, as mentioned above, a danger of viable MT is its negative effect on language education, AVENUE's researchers see that danger as a blessing in disguise to minority language speakers; by reducing the need to speak a common language (such as Spanish or English), MT may allow local languages to flourish.[101]

     MT, although historically a technology of military intelligence, has the potential to act as a bridge between people and nations with different languages and different world views.

C.   *Commercial Benefit of Viable MT*

     Viable MT would have two commercial consequences. First, there would be an effect on the current translation market. Second, viable MT would eliminate the language barriers between sellers and buyers online and transform international e-commerce.

     The worldwide translation market is estimated at $11.5 billion by the end of 2007, with software and Web localization markets to account for over half that amount.[102]   The translation labor market consists of translators in the public and private sector, employed either as freelancers in translation agencies, company departments, organizations, or localization firms, which have been consolidated recently in the hands of a few large players.[103]   MT is estimated to earn only $134 million in revenues by 2007, or less than 1% of the total translation market.[104]   The world's largest MT company, Systran, posted revenues of only $13 million in 2004.[105]   MT, in other words, is a market still in its infancy. It is possible that legal uncertainty and litigation at this juncture could nip the technology in the bud.

     Viable MT's effect on the current translation market depends on three factors: how much it can improve productivity, its automation potential (i.e., what human labor it can replace), and the supply/demand

---

     100.   Ariadna Font Llitjós et al., *Building Machine Translation Systems for Indigenous Languages* (2005), http://sidecar.sp.cs.cmu.edu/ari/papers/FontLlitjos-CILLAII-2005.pdf.
     101.   *See* AVENUE, http://www.cs.cmu.edu/~avenue/ (last visited Aug. 6, 2006).
     102.   Press Release, ABI Research, Language Translation, Localization and Globalization (Sept. 12, 2002), *available at* http://www.abiresearch.com/abiprdisplay.jsp?pressid=181.
     103.   Ed Frauenheim, *Translation Firms Grow Beyond Tech*, CNET, Mar. 28, 2003, http://news.com.com/2100-1011_3-994415.html.
     104.   *See* Press Release, *supra* note 102.
     105.   Stix, *supra* note 39, at 93.

of human translators. So far, MT has replaced humans outright only in areas of "restricted-source translation," in which the subject matter and format of the source text are extremely limited and therefore easy to translate.[106] The best example is the Météo system, which has been translating Canada's weather bulletins since 1977.[107] Before Météo, the Canadian government had a hard time finding translators because the work was so mind-numbingly dull. Viable MT would probably expand possibilities for restricted-source translation.

For some types of technical documents, MT can replace human translators, though not the need for human labor entirely, through "pre-edited translation," in which a human editor prepares a text for MT by removing ambiguities and limiting the language to certain vocabulary and grammar.[108] This "restricted language" approach is particularly cost effective when there is a need to translate a document into many languages, for instance legal documents in the European Community or companies that need the same manual translated for many markets.[109] It is also cost effective because such editors can be paid less than bilingual translators.

Restricted languages are sometimes called "Caterpillar English" because Caterpillar Corp. was the first to try writing its manuals in a form of radically restricted English, called "Caterpillar Fundamental English."[110] Like most experiments in invented language, from Esperanto to Ebonics, Caterpillar Fundamental English failed miserably, because Caterpillar's employees refused to learn it.[111] But Caterpillar eventually got restricted language rights, as have companies like Xerox, which translates some 50,000 pages of manuals each year written in restricted English.[112] Viable MT will continue this process and expand the types of documents that restricted language MT can translate.

Viable MT will not have a uniform effect on human translation labor. MT will replace human translators for documents that can be translated outright or through pre-editing, but advanced gisting may *increase* demand for human translators; for example, a Reuters editor reads a Chinese blog via MT, decides that it sounds newsworthy, and fires it off to a human translator to verify the facts and make the news

---

106. STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 851 (2d ed. 2003).

107. *See* Silberman, *supra* note 21, at 228.

108. *See* RUSSELL & NORVIG, *supra* note 106.

109. *Id.*

110. *Id.*

111. Silberman, *supra* note 21, at 233.

112. HOFSTADTER, *supra* note 51, at 503.

print-worthy. The demand for translators of rare and high-demand language pairs may remain completely unaffected by better MT. Accordingly, legislation that would protect MT from litigation will not create a marked increase or decrease in jobs for human translators, a factor important to policy makers.

Amidst all the guesswork and crystal ball gazing involved when predicting the future of MT, one thing at least is absolutely certain: literary translators would be least affected by better MT. These professional translators of novels, articles, and poetry do make use of computers, including online dictionaries for definitions and search engines to analyze common usage and context of words.[113] But MT, even excellent MT, has little to offer the fully bilingual human translator. The reason is that MT has a "tin ear," so to speak; any output by MT would have to be re-worked to make it readable from a literary standpoint, eliminating any marginal productivity gains that MT may accomplish. Viable MT, sadly, will therefore not create libraries of translated fiction and magazines.

Generally, improved MT will have the same impact on the translation market as other computer technologies have changed existing markets: some human labor will be automated, other jobs (such as Caterpillar English editors) may be created, and the market for the technology will grow. The more tantalizing potential of online MT is the impact it may have on global e-commerce as a whole. Quite simply, if a customer cannot understand what you are selling, he will not buy.[114] By allowing buyer and seller to speak the same language, viable MT could transform virtually all types of e-commerce: advertising, content, auctions, business-to-consumer, and business-to-business. Currently, the translation needs of e-commerce are handled by the translation industry. Web site localization, the process of making the same information available in numerous languages, is unsurprisingly the fastest growing

---

113. E-mail from Geoffrey Brock, translator of Italian literature, to Author (Jan. 6, 2006, 3:32:46 PM EST) (on file with author) ("Whenever I'm translating . . . I have a browser open with eight tabs: two tabs display monolingual Italian dictionaries, two tabs display Italian/English dictionaries, three tabs display English dictionaries, and the eighth tab is for Google. Google is actually the ultimate global dictionary, because I can search not just for individual words but also for whole phrases, and I can see how they are used in a variety of contexts, from formal (edited, professional Web sites) to informal (blogs, message boards, etc), and I can see other relevant things, too, such as how common a word or phrase is or how commonly a certain adjective is paired with a certain noun—things regular dictionaries don't tell you.").

114. A management consultancy, Aberdeen Group, conducted a study to verify this obvious conclusion. *Tongues of the Web, supra* note 37 ("On average, users spend up to twice as long at a site, and are four times more likely to buy something from it, if it is presented to them in their own language.").

segment of the translation market, increasing from $499 million in global revenues in 2001 to an estimated $3.1 billion in 2007.[115] This growth in localization spending shows that companies are eager to communicate with customers around the world and that the demand for a solution such as viable MT is already strong.

Through online MT, every e-commerce site on the Web will suddenly serve an international market, increasing the potential sales for every product sold online, from Pez dispensers to real estate. The low marginal cost of viable MT will bring the technology to buyers and sellers who previously had no access to translation, human or otherwise. Localization is difficult for small players and a massive cost for large corporations; it recently took Philips Electronics *four months* to add the same update to all fifty of its localized sites.[116] Viable online MT could shorten such lead times for larger companies that employ localization companies and commercially benefit small e-commerce outfits by adding functionality. In theory, MT could revolutionize online commerce as profoundly as online communication.

IV. MACHINE TRANSLATION UNDER CURRENT LAW

Translating content online does not always create copyright problems. Users can translate their own expressions, and anyone could legally translate copyrighted works whose authors expressly allow it, such as Creative Commons licensees that allow the creation of derivative works;[117] works not subject to copyright, including factual material such as locations, directions, forecasts, departure times, and recipe lists, which have a limited scope of protection under copyright;[118] and texts that have fallen into the public domain.

Copyrighted material, however, makes up much of the best information online: news reports, blogs, song lyrics, music, movies, and images. To analyze viable MT under current copyright law, assume the following hypothetical. Babeltron.com operates much the same as

---

115. Press Release, *supra* note 102.

116. Maximizing Reuse in Product Communications: Philips Case Study, *available at* http://software.emc.com/microsites/brand_mgmt/mrm_brand_mgmt/collateral/Case_Study_Phili ps.pdf (last visited Aug. 6, 2006).

117. *See* Creative Commons Licenses, http://creativecommons.org/about/licenses/meet-the-licenses (last visited Aug. 6, 2006).

118. Section 102(b) of the Copyright Act is universally understood to prohibit any copyright in facts. Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 356 (1991). Factual compilations, on the other hand, may possess the requisite originality so long as the selection and arrangement of facts are made independently by the compiler and entail a minimal degree of creativity. *Id.* at 358.

current free online MT offerings.[119] Once a user accesses Babeltron.com, she can either input text or the URL of a site she wants translated into empty text boxes. The user then selects a language pair (e.g., from Urdu to English) and clicks the "Translate!" button. Either the translated text or a copy of the original Web site, with language translated inserted, is delivered from Babeltron's servers. Babeltron derives revenue from the site by selling banner advertisements on its home page. Katie is a young American scholar who placed her historical essay, "Napoleon in Vienna," online.[120] Katie finds out that Jacques, a French-speaking history buff who is a resident of Maine, translated "Napoleon in Vienna" from its Web page into French using Babeltron.com. Katie sues Babeltron and Jacques for copyright infringement.

A.   *Direct Infringement by Babeltron or User*

    No federal court has yet addressed the copyright implications of MT. To demonstrate copyright infringement, a plaintiff must show ownership of the copyright and copying by the defendant.[121] The owner of a copyright has the exclusive right to create and authorize derivative works.[122] It is beyond question that translations constitute derivative works, which are actionable if not authorized by the copyright holder of the original.[123]

    But is MT output actually a "translation," and therefore a derivative work, under the Copyright Act? Just as the copyright laws do not expressly require "human" authorship, title 17 does not explicitly require translation, or any other derivative work, to be performed by a human.[124] Sound recordings and art reproductions, like MT, can be created at the touch of a single button and create derivative works under section 101 of the Copyright Act. While an argument can be made that, theoretically, MT is not "translation," a plain language reading suggests that MT performs what it says: translation.[125] As such, machine translation of a

---

    119.   *See* Yahoo! Babel Fish, *supra* note 65.

    120.   If more details are desired by the reader, Katie is an attractive brunette known for wearing glasses and light blue oxford shirts, while "Napoleon in Vienna" is a historical study of the French occupation of Vienna in 1809.

    121.   4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 13.01 (2005).

    122.   17 U.S.C. § 106(2) (2000).

    123.   *Id.* § 101 (defining "derivative work" to include "translation"); *see also* 1-2 NIMMER & NIMMER, *supra* note 121, §§ 2.04, 3.01.

    124.   Urantia Found. v. Maaherra, 114 F.3d 955, 958 (9th Cir. 1997) (addressing the bizarre question of whether a book purportedly authored by celestial beings may be copyrighted). "The copyright laws, of course, do not expressly require 'human' authorship." *Id.*

    125.   *See* HOFSTADTER, *supra* note 51, at 515-18 (discussing whether MT can actually be termed "translation" by analogizing whether a computer controlling a car is actually "driving").

text creates a derivative work under the Copyright Act and Babeltron may be liable for copyright infringement if that translation is unauthorized.

Direct infringement is a strict liability violation and does not require a showing of intent.[126] But a number of courts have held that a finding of direct infringement still requires a "volitional act"—in other words, you still must *do something*—and that "automated copying by machine occasioned by others" therefore cannot support direct infringement.[127] Similarly, MT responds automatically to the information a user types in, from their own expressions to an entire novel, and it is likely that jurisdictions following the volitional act requirement would find no direct infringement by Babeltron. Jacques, on the other hand, may still be liable for direct infringement.

B. Secondary Liability

Contributory copyright infringement is established where, in addition to direct infringement by a user, a defendant has "knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another."[128] Under the landmark case *Sony Corp. of America v. Universal City Studios, Inc.*, the United States Supreme Court held in a five-four decision that "the sale of copying equipment, like the sale of other articles of commerce, does not constitute contributory infringement if the product is widely used for legitimate, unobjectionable purposes. Indeed, it need merely be capable of substantial noninfringing uses."[129] MT is an "article of commerce" under *Sony* and clearly capable of both infringing and noninfringing uses, as detailed above.[130] The Supreme Court in *Sony* held that "time shifting" (i.e., recording onto videotape) of programs broadcast on television

---

126. Although direct infringement does not require a showing of intent, willfulness is relevant to the award of statutory damages. 17 U.S.C. § 504(c).

127. *See* Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1370 (N.D. Cal. 1995) (stating that direct infringement requires a volitional act by defendant; automated copying by machines occasioned by others is not sufficient); Co-Star Group, Inc. v. LoopNet, Inc., 373 F.3d 544, 555 (4th Cir. 2004) ("Agreeing with the analysis in Netcom, we hold that the automatic copying, storage, and transmission of copyrighted materials, when instigated by others, does not render an ISP strictly liable for copyright infringement under §§ 501 and 106 of the Copyright Act."); Newborn v. Yahoo!, Inc., 391 F. Supp. 2d 181, 186 (D.D.C. 2005); Field v. Google Inc., 412 F. Supp. 2d 1106, 1115 (D. Nev. 2006); Parker v. Google, Inc., 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006).

128. Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc., 443 F.2d 1159, 1162 (2d Cir. 1971).

129. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442 (1984).

130. *See* Vault Corp. v. Quaid Software, Ltd., 847 F.2d 255, 267 (5th Cir. 1988) (suggesting that a single noninfringing use is sufficient to establish that a technology is a staple article of commerce).

constituted substantial noninfringing use because a significant number of broadcasters had no objections to such home taping.[131] As noted above, there are also a number of Web sites that do not object to translation of their content through third-party MT, and in fact actively link to them.

    The Supreme Court supplemented the rule in *Sony* in the landmark *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster Ltd.*, holding that "one who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties."[132] In *Grokster*, the Supreme Court noted a number of "affirmative steps" which led to a finding of inducement: (1) targeting users of Napster, a "notorious" P2P application which was successfully sued for copyright infringement; (2) planning to flaunt the illegal uses of the application through marketing; (3) business models which turned on high volume use of their application, which the P2P companies knew were infringing; and (4) no effort to filter copyrighted material from users' downloads or take other preventive measures.[133] The Supreme Court noted, however, that failure to take affirmative steps to prevent infringement could not lead to a finding of contributory infringement unless other evidence of intent to infringe was present and the device was capable of substantial noninfringing uses.[134]

    Generally, Babeltron is protected from secondary liability under *Grokster* unless it induces users to infringe copyrighted works. What additional steps should Babeltron take to prevent liability under the *Grokster* inducement standard? If a notoriously infringing MT site were to come along, Babeltron would be prudent to avoid targeting its users. More to the point, Babeltron should not market its infringing uses, for instance with advertising slogans like, "Read Harry Potter in any language," or even, perhaps, "Surf the entire Web in English." Babeltron's business model may venture too close to *Grokster* territory; it is likely that many users will use Babeltron to translate copyrighted material, and advertising revenue naturally increases with high-volume use. For efforts to filter copyrighted material, Babeltron may take preemptive steps to block translation of major information or news

---

    131. *Sony*, 464 U.S. at 446 ("[T]here are many important producers of national and local television programs who find nothing objectionable about the enlargement in the size of the television audience that results from the practice of time-shifting for private home use.").
    132. Metro-Goldwyn-Mayer Studios, Inc. v. Grokster Ltd., 545 U.S. 913, 919 (2005).
    133. *Id.* at 922-27.
    134. *Id.* at 939 n.12.

sources, although a preferred solution is to honor certain metatags as advocated below.

C.   *Implied License*

Babeltron may also raise the affirmative defense of implied license. Although a transfer of copyright ownership must be in writing under section 204(a) of the Copyright Act, a nonexclusive license may be implied from conduct.[135] Uses of the copyrighted work that stay within the scope of a nonexclusive license are immunized from infringement suits.[136] The burden of proving the existence of such a license is on the party claiming its protection, the licensee.[137]

Some commentators argue that implied license may be a way for courts to protect potentially infringing, but ubiquitous, online activities like forwarding e-mail or downloading documents from the Internet.[138] Even simple Web browsing may be interpreted to function through implied license.

> We can presume that the copyright owner has granted an implied license to allow people to copy a Web page to a local machine and display it there; after all, if they did not want people to be able to read a page (which . . . means making a temporary copy on your [computer]), they would not have put the document up on the Web.[139]

This argument assumes that Web site owners, by placing sites online, imply a license to *do things* with their content: search engine robots crawl and cache the site, users may select to view only certain RSS feeds, while other users may use sophisticated browsers to remove the ads, view the site in text or links only, or view the site in different sizes or with a different cascading style sheet. This argument also correctly assumes that Web site owners maintain some control over such activities through the use of HTML tags and other standards.

---

135.   Effects Assocs., Inc. v. Cohen, 908 F.2d 555, 558-59 (9th Cir. 1990). In recognizing that nonexclusive licenses in copyright may be created orally or by conduct in *Cohen*, the Ninth Circuit relied in part upon 3 NIMMER & NIMMER, *supra* note 121, § 10.03[A] (1989). *Cohen*, 908 F.2d at 558-59. For a critique of this reliance on Nimmer, see Ann Bartow, *The Hegemony of the Copyright Treatise*, 73 U. CIN. L. REV. 581, 632-35 (2004).

136.   *See* John G. Danielson, Inc. v. Winchester-Conant Props., Inc., 322 F.3d 26, 40 (1st Cir. 2003) (citing Graham v. James, 144 F.3d 229, 236 (2d Cir. 1998)).

137.   Bourne Co. v. Walt Disney Co., 68 F.3d 621, 631 (2d Cir. 1995).

138.   COHEN ET AL., *supra* note 5, at 194.

139.   Peter B. Hirtle, Digital Preservation and Copyright, http://fairuse.stanford.edu/ commentary_and_analysis/2003_11_hirtle.html (last visited May 5, 2006).

Although no single test has been adopted, a number of courts have held that the touchstone for finding an implied license is intent.[140] A recognition of standards and tags may help courts recognize implied licenses online because they tell us more about Katie's intent than the mere fact that she placed her work online. In *Field v. Google Inc.*, the United States District Court for the District of Nevada held that an author granted an implied license to a search engine to cache the copyrighted works he had placed online because (1) he had knowledge of a well-known industry standard "no-archive" metatag and (2) chose not to include the "no-archive" tag in his code.[141] If courts follow the holding in *Field*, Katie and Babeltron's problems could be prevented through the use of a "do not translate" metatag, which would inform online translators that an author does not wish her Web pages to be directly translated online.

Currently there is not a widely used "do not translate" metatag in existence. Where might one come from? Many of the most widely spread Internet standards, including Cascading Style Sheets (CSS) and Extensible Markup Language (XML), were promulgated by the World Wide Web Consortium (W3C), the leading international Internet standards body. The W3C, however, tends to create large sets of standards, not individual tags. At least one expert also believes that due to its international nature, the W3C would actually *oppose* a "do not translate" tag on principle.[142] The most effective implementation of a "do not translate" standard would probably come from MT providers themselves. If courts give weight to the decision in *Field*, Babeltron could shield itself from liability by creating such a metatag and informing content providers about it.

Code-based solutions, like any solutions, may be good or bad. The leading theorist on code and law, Professor Lawrence Lessig, opposes

---

    140.   *See Danielson*, 322 F.3d at 40-41 ("The touchstone for finding an implied license . . . is intent."); Nelson-Salabes, Inc. v. Morningside Dev., LLC, 284 F.3d 505, 515 (4th Cir. 2002) (calling intent the "determinative question"); Johnson v. Jones, 149 F.3d 494, 502 (6th Cir. 1998) ("Without intent, there can be no implied license."); Data Gen. Corp. v. Grumman Sys. Support Corp., 36 F.3d 1147, 1167 n.35 (1st Cir. 1994) (stating that a license is found from the copyright owner's grant of "permission to use").
    141.   Field v. Google Inc., 412 F. Supp. 2d 1106, 1116 (D. Nev. 2006).
    142.   E-mail from Professor Lorrie Cranor, Director, Carnegie Mellon University Usable Privacy and Security Laboratory (CUPS), to Author (Feb. 6, 2006, 10:43:03 AM EST) (on file with author) ("I suspect they [W3C] wouldn't want to take on something so narrow as a do-not-translate tag. Instead they might be interested in doing a larger meta-data framework. Also, they are really into internationalization and they would probably be opposed to a do-not-translate tag on principle.").

some code-based solutions but champions others.[143]   Recognition of a code-based test for an implied license is attractive because it is simple. Liability would hinge on code: whether Katie uses a "do-not-translate" tag and whether Babeltron honors it are easy concepts for lawyers and Webmasters alike. An implied license theory based on code also makes economic sense; given the potential benefits of online MT, society should place the burden of preventing infringing use on those most able to do it, content providers.

     There are two potential problems with finding an implied license to translate online in the absence of a "do not translate" tag. Practically, a tag would not stop someone determined to translate Katie's work. Although Babeltron could honor a "do not translate" tag by not allowing Katie's URL to be entered into Babeltron's "translate a Web page" text box, someone could simply cut and paste her text into Babeltron's text translation box. But the aim of the "do not translate" tag is not to prevent online translation, which, as noted, is impossible, but to reduce legal risk for creators of MT. By accepting the "do not translate" tag, Babeltron can argue that it does all it can in honoring Katie's intent.

     Legally, finding implied license *unless* Katie takes the active step of adding the "do not translate" tag to her code is inconsistent with traditional copyright law.   Karl-Friedrich Lenz argues we should not support opt-out solutions for practices that "violate copyright as the default," like the way Google caches all Web content that does not contain "no archive" code.[144] Lenz believes that the default should be the reverse: users who wish their content to be cached by Google should affirmatively make that decision, also through code.[145]

     A better solution, therefore, may be the creation and promulgation of an affirmative "you may translate" tag to serve as an explicit license. Such a tag would allow courts to gauge Katie's intent and allows users to benefit from content online while protecting the rights of authors. The downside is that if the "you may translate" tag fails to become widely adopted, surfing the Web with Babeltron may still constitute illegal activity. In sum, courts, MT companies, and legislators should create and promulgate either a "do not translate" or "you may translate" metatag for

---

     143. *See, e.g.*, Lawrence Lessig, *Tyranny in the Infrastructure*, WIRED, July 1997 (discussing PICS, an HTML standard that makes it possible to filter material online, Lessig states that "PICS is the devil"); *cf.* Lawrence Lessig, *Porn Free*, WIRED, Sept. 2004, at 104 (arguing for the adoption of a mandatory HTML tag such as <porn> to regulate pornography online).
     144. Karl Lenz's criticism of Google's opt-out procedure for default Internet caching would probably apply to a "do not translate" tag. *See* Posting of Karl-Friedrich Lenz to Lenz Blog, http://k.lenz.name/LB/archives/001083.html (Aug. 9, 2005, 9:24 EST).
     145. *Id.*

users like Katie who place content online. Although neither is perfect, such a code-based solution will create more legal certainty for online MT.

D. Fair Use

Is it possible that translating a Web page using Babeltron constitutes fair use, an affirmative defense to copyright infringement?[146] Fair use is a limitation on the exclusive rights under section 106, and includes criticism, comment, news reporting, teaching, and research.[147] These exceptions, like MT, further cultural interchange, and courts have recognized fair use in culturally beneficial types of works that otherwise resemble derivative works, most notably parody.[148]

Fair use is determined by four factors under section 107. The first factor is "the purpose and character of the use, including whether such use is of a commercial nature."[149] Babeltron sells advertisements on its site, and is therefore commercial in nature, but commercial use of copyrighted material does not end the inquiry under this factor.[150] The central purpose of the investigation is whether and to what extent the new work is "transformative."[151] The more transformative the new work, the less important the other factors, including commercialism, become.[152] Translation is certainly transformative, insofar as it transforms words from one language to another. Some courts have been inclined to find fair use for useful transformative online technologies, like image search engines and search engine caching.[153]

The second statutory factor looks to "the nature of the copyrighted work."[154] The Supreme Court has held that creative works fall within the core of copyright's protective purposes, and although Katie's essay is historical in nature, it is a work of prose, and therefore expression, which may weigh against Babeltron.[155] The third factor looks to "the amount and substantiality of the portion used in relation to the copyrighted work

---

146.   Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 590 (1994).
147.   17 U.S.C. § 107 (2000).
148.   *Campbell*, 510 U.S. at 594 (holding that the parody of a song constitutes fair use).
149.   17 U.S.C. § 107(1).
150.   *Campbell*, 510 U.S. at 579.
151.   *Id.*
152.   *Id.*
153.   Kelly v. Arriba Soft Corp., 77 F. Supp. 2d 1116, 1121 (C.D. Cal. 1999), *aff'd in part and rev'd in part*, 280 F.3d 934 (9th Cir. 2002); Field v. Google Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).
154.   17 U.S.C. § 107(2) (2000).
155.   *Campbell*, 510 U.S. at 586.

as a whole," and here, Babeltron copied the entirety of "Napoleon in Vienna."[156]

The fourth fair use factor is "the effect of the use upon the potential market for or value of the copyrighted work."[157] Here, the inherent limitations of MT weigh in favor of Babeltron. Since automatic translation of online works will never attain the quality of a translation created by humans, the market for translations of "Napoleon in Vienna" is not affected by Babeltron. It may even help Katie, if MT piques the curiosity of non-English speakers like Jacques, who then create demand for a professionally translated copy of her work.

Overall, Babeltron would fare well under the fair use analysis were it not for the fact that it translates entire Web pages. If Babeltron were used to translate only a few sentences of works, it probably would qualify as fair use.

E.   Digital Millennium Copyright Act

The Digital Millennium Copyright Act (DMCA) may affect Babeltron in two ways. First, it could strengthen the efficacy of code solutions like a "do not translate" or "you may translate" tag. The DMCA prohibits the circumvention of technological measures that effectively control access to protected works.[158] Although a translation tag is more of a rights-control (it creates an implied or express right to translate) than an access-control measure, at least one court has applied the DMCA's anticircumvention measures to enforce rights-control measures.[159] It remains to be seen whether other courts will follow suit.

Second, it is possible that Babeltron may be sued for caching the content that its users input for translation, storing translations of copyrighted material on its systems, or for linking to infringing content translated using its services. To prevent liability for these or future related claims, Babeltron may qualify for one of the safe harbors of the DMCA, which protect certain common activities of Internet Service Providers (ISPs).[160] The DMCA defines "service provider" in two different ways, depending upon which safe harbor is at issue.[161] Online MT like Babeltron can probably qualify as a service provider under

---

156.   17 U.S.C. § 107(3).
157.   *Id.* § 107(4).
158.   *Id.* § 1201.
159.   *See* R. Anthony Reese, *Will Merging Access Controls and Rights Controls Undermine the Structure of Anticircumvention Law?*, 18 BERKELEY TECH. L.J. 619, 631 (2003).
160.   17 U.S.C. § 512(a)-(d).
161.   *Id.* § 512(k)(1).

section 512(k)(1)(B), "a provider of online services or network access." Babeltron meets a plain language reading of the statutory language—it provides the online service of translation—and courts have emphasized the broad scope of this definition.[162]

V. CONCLUSION

Although viable MT is still on the horizon, it may arrive sooner than we think. As this Article should make clear, a variety of overlapping claims lie waiting for online MT and threaten to shoot this socially, politically, and commercially beneficial technology out of the sky. We should pave the way for online MT through statutory recognition of its noninfringing nature and the creation of standards such as translation metatags.

In *Genesis*, "the whole earth was of one language, and of one speech" when a united mankind began to build a tower that reached to the Heavens, the Tower of Babel.[163] According to *The Bible*, God saw what mankind was up to and said,

> Behold, the people is one, and they have all one language; and this they begin to do: and now nothing will be restrained from them, which they have imagined to do. Go to, let us go down, and there confound their language, that they may not understand one another's speech.[164]

The people thereupon abandoned work on the Tower and God scattered them across the face of the Earth.

Could MT be the tool that reunites mankind, bringing about an age when nothing we can imagine will be restrained from us? MT researchers have toiled for decades to bring about this future, but recent advances and the advent of the Internet may make viable MT possible, after all. MT, even if far from perfect, may allow us to rebuild a shaky sort of Babel, not quite reaching the heavens, but pointed in the right direction.

---

    162.   *See In re* Aimster Copyright Litig., 334 F.3d 643, 655 (7th Cir. 2003); ALS Scan, Inc. v. RemarQ Cmtys., Inc., 239 F.3d 619, 623 (4th Cir. 2001); Corbis Corp. v. Amazon.com, Inc., 351 F. Supp. 2d 1090, 1100 (W.D. Wash. 2004) ("This definition [of § 512(k)(1)(B)] encompasses a broad variety of Internet activities.").
    163.   *Genesis* 11:1 (King James).
    164.   *Genesis* 11:6-7 (King James).