10 Is machine language translation a viable tool for health communication?

Xuewei Chen and Sandra Acosta

Section I: language barriers, miscommunications, and health care

Language barriers and miscommunications may lead to negative consequences in health care. For example, when patients and health care providers speak different languages, medical misdiagnoses likely follow. Patients often misunderstand or have limited access to quality health information, resulting in medication dosage errors, inappropriate health decisions, extra time and money for health care services, and ineffective experiences with health care providers (Jacobs, Karavolos, Rathouz, Ferris, & Powell, 2005; Nair & Cienkowski, 2010; Ponce, Hays, & Cunningham, 2006; Thomson & Hoffman-Goetz, 2011; Toci et al., 2014; Wilson, Chen, Grumbach, Wang, & Fernandez, 2005). In brief, language barriers between patients who are speakers of other languages and health providers result in less effective health care. For vulnerable populations, language barriers contribute to health disparities and poor health outcomes.

Given the potential health impacts of language barriers, adequate language translation assistance is crucial. Linguistically vulnerable populations include immigrants whose primary language is different from the host country, international travelers needing medical care or disease management, and limited or non-English speaking health professionals seeking medical research articles in English (Anazawa, Ishikawa, & Takahiro, 2013; F. Liu, Ackerman, & Fontelo, 2006; Wu, Xia, Deleger, & Solti, 2011). By way of illustration, more than 21 percent of the U.S. population speaks a language other than English at home, and over 41 percent of them speak English "less than very well" (Ryan, 2013). The U.S. Census Bureau defines this group as individuals with limited English proficiency (LEP). This group comprises individuals ages 5 and older who self-report speaking English "less than very well" as LEP.

Other examples, of populations include English speakers with diet management requirements travelling in non-English-speaking countries, such as Spain, who need language translation assistance for restaurant food menus (Pozo, Haddad, Boutin, & Delp, 2011). Others include the English-speaking populations in South Korea and Chinese-speaking population in Japan. These populations are increasing and need language translation assistance when seeking health care services (Fukushima, Yoshino, & Shigeno, 2011; Ozaki, Matsunobe, Yoshino, & Shigeno, 2011; Shin et al., 2015). Thus, language translation and health communication across languages touches all countries and languages groups.

Section II: machine translation tools

Language machine translation technology can support patient-clinician communication when human interpreters are not available. This section introduces language translation devices including computer software, websites, and mobile translation applications used in clinical and public health settings. Most devices are text-based and/or speech-based. We present the research findings in peer-reviewed papers with regard to accessing the accuracy of language translation devices in health-related contexts. One research study is a recent systematic literature review (Dew, Turner, Choi, Bosold, & Kirchhoff, 2018). This synthesis contains detailed information about machine translation tools developed for overcoming language barriers in health settings.

This section is comprised of two categories of machine language tools followed by an evaluation of each tool: (1) *Text-based tools* – Google Translate, Babel Fish, Babylon, NoteAid$_{\text{Spanish}}$, Moses, xprompt, language grid, BioMT, Micro Merchant, OPUS-ISM, and BestRx; (2) *Speech-based Systems* – Speaking Multilingual Interactive Natural Dialog Systems (S-MINDS), Petit Translator, MedSLT, DARwIN-OP robot, and iTranslate, Converser for Health care.

Text-based systems

Google Translate

Google Translate is a commonly used machine translation tool that instantly translates text and Webpages. Its mobile application also translates text into images (instant camera translation), photos, conversations, and handwriting (draw text characters instead of typing).

Evaluation of Google Translate

Google Translate has been evaluated as a tool for translating health education documents for patients and public. Kirchhoff, Turner, Axelrod, and Saavedra (2011) evaluated the accuracy and adequacy of Google Translate when translating health-promotion materials from local and national public health Websites from English to Spanish. Two native speakers of Spanish with fluent knowledge of English rated the translated document. They found common error categories were morphologic errors, word sense errors, and other grammatical errors. The study concluded that although the translation quality was imperfect, it held great promise for assisting health agencies to provide quality translations to individuals with LEP (Kirchhoff et al., 2011). Khanna and colleagues (2011) evaluated Google Translate when translating sentences from the instruction manual regarding warfarin use from English to Spanish. The authors (Khanna et al., 2011) evaluated four domains: fluency (grammatical correctness), adequacy (information preservation), meaning (connotation maintenance), and severity (perceived dangerousness of an error if present). They found that Google Translate, when compared to professional human translators, made more errors. However, Google Translate was not statistically more likely to make a severe error (Khanna et al., 2011).

Chen, Acosta, and Barry (2016) used Khanna and colleagues' (2011) evaluation rubric to assess Google Translate when translating a diabetes patient education pamphlet from English to Spanish and English to Simplified Chinese. The researchers (Chen et al., 2016) invited a group of professional medical translators to evaluate and score the translations based on fluency, adequacy, meaning, and severity. They found that Google Translate produced accurate translations for simple sentences but not complex ones. Also, Google Translate produced more accurate translations from English to Spanish than from English to Chinese. Some sentences translated by Google Translate from English to Chinese potentially might result in delays of patient care (Chen et al., 2016).

Turner, Dew, Desai, Martin, and Kirchhoff (2015) evaluated Google Translate when translating health promotion documents from English to Traditional Chinese. Two professional public health translators and five native Chinese speakers rated the translations. The study notes that the most common translation errors were errors of word sense and word order. The researchers concluded that translation quality was problematic and more work was needed to improve the tool before it might be used routinely in public health practice (Turner et al., 2015).

Patil and Davies (2014) evaluated the accuracy and usefulness of Google Translate when translating common medical statements from English to 26 languages. Their study found that among all the translated phrases, 57.7 percent were correct and 42.3 percent were wrong. African languages scored the lowest, followed by Asian languages and Eastern European languages. Western European language translations were the most accurate. The researchers concluded that Google Translate should be used with caution because many translations were completely wrong (Patil & Davies, 2014).

Google Translate has also been evaluated when translating restaurant menus for patients with diet management, screening questionnaires, and research articles/abstracts for medical professionals. Pozo and colleagues (2011) evaluated Google Translate for helping English-speakers, following a medical diet, maintain this diet while visiting non-English speaking countries. Their sample was English speakers travelling in Spain. The researchers compared the content of food offered on restaurant menus for the English-speaking travelers on a special diet. A Spanish speaker evaluated the translation accuracy (correct vs incorrect). The study found that Google Translate yielded the correct translation 73 percent of the time (Pozo et al., 2011).

Taylor, Crichton, Moult, and Gibson (2015) accessed Google Translate when translating an English information sheet (the Strengths & Difficulty

Questionnaire – a behavioral screening tool for children) into Spanish and Chinese. Sixteen people, fluent in Spanish or Chinese and written/spoken English, evaluated the translations based on four key categories: target language, textual and functional adequacy, non-specialized content, and specialized content. The study found that no score was in the acceptable range. The researchers concluded that Google Translate was not sufficiently accurate at this stage (Taylor et al., 2015).

Anazawa and colleagues evaluated the quality of Google Translate when translating nursing research article abstracts from English into Japanese. Two hundred and fifty nurses in Japan (Japanese native speakers) finished the survey and were asked to rate the translation products. The study indicated that Japanese nurses did not perceive Google Translate as having adequate quality (Anazawa et al., 2013). They also compared the translation from Google Translate with other three machine translation devices (Cross Language, Bing Translator, and BizLingo). Two researchers rated the accuracy of these four devices. The study found that Cross Language performed best for accuracy, followed by BizLingo, Google Translate, and Bing Translator (Anazawa et al., 2013).

Babel Fish

Babel Fish (www.babelfish.com) is a free online translator for translating phrases and sentences into different languages. This tool supports 75 languages.

Evaluation of Babel Fish

Four reviewers evaluated the understandability and correctness of medical record sentences translated from English into Spanish, Chinese, Russian, and Korean (Zeng-Treitler, Kim, Rosemblat, & Keselman, 2010) by Babel Fish. They found that in each language, the majority of the translations were incomprehensible and/or incorrect. Spanish had higher accuracy than other languages. They contend that "one possible explanation for this may well lie in the fact that English and Spanish are more similar (e.g., word order, inflections) than English and Chinese, Korean or Russian" (p. 76). The researchers concluded that Babel Fish was not adequate for translating medical records.

Babylon

Babylon.com offers translations in 77 languages with the free version or the commercial version (upgrades). This program was developed by the Israeli company Babylon Software Ltd.

Evaluation of Babylon

Taylor and colleagues (2015) evaluated and compared Google Translate and Babylon 9, a commercial dictionary and translation software, when translating an

English information sheet (the Strengths & Difficulty Questionnaire – a behavioral screening tool for children) into Spanish and Chinese. Sixteen people, who were fluent in Spanish or Chinese and fluent in written/spoken English, evaluated the translation based on four key categories: target language, textual and functional adequacy, non-specialized content, and specialized content. The study found that Google Translate received better scores than Babylon 9. Nonetheless, no score was in the acceptable range. The researchers concluded that these two machine translation tools were not sufficiently accurate at this stage (Taylor et al., 2015).

NoteAid$_{Spanish}$

NoteAid$_{\text{Spanish}}$ is an English-Spanish machine translation system for electronic health records.

Evaluation of NoteAid$_{Spanish}$

W. Liu and Cai (2015) used BLEU score and a bilingual human expert to evaluate three systems: NoteAid$_{\text{Spanish}}$, Google Translate, and Bing Translator when translating electronic health records. BLEU is a method for automatic evaluation of machine translation quality (Papineni, Roukos, Ward, & Zhu, 2002). W. Liu and Cai (2015) found that Google Translator outperformed NoteAid$_{\text{Spanish}}$. The authors also developed hybrid machine translation systems to replace medical jargon with lay terms and then perform the translation. They concluded that this step (decreasing medical jargon) improved the translation quality scores (W. Liu & Cai, 2015).

Moses

Moses is a statistics-based system that uses a phrase-table built from a given parallel corpora for translation.

Evaluation of Moses

Pozo and colleagues (2011) evaluated and compared Moses and Google Translate for helping individuals, following medical diets, maintain their diet plan while visiting non-English speaking countries. In their study (Pozo et al., 2011), the authors selected English speakers travelling in Spain (the content of food offered on restaurant menus). A Spanish speaker evaluated the translation accuracy (correct vs incorrect). The study found that Google Translate yielded the correct translation 73 percent of the time; Moses yielded the correct translation 83.2 percent of the time (Pozo et al., 2011).

xprompt

Xprompt is an iPad/iPhone application used in medical care to provide multilingual assistance.

Evaluation of xprompt

A study examined 39 health care staff's experiences of using xprompt. The authors found that health care staff perceived the tool as helpful for communicating with foreign language patients (Albrecht, Behrends, Schmeer, Matthies, & von Jan, 2013).

Language Grid

The Language Grid is a multilingual service platform with online dictionaries, bilingual corpora, and machine translators. A Japanese non-profit organization, Language Grid Association, operates this service.

Evaluation of Language Grid

Fukushima and colleagues (2011) evaluated the translation accuracy of Langue Grid when translating hospital interview sheets (e.g., Where is the pain?) from Japanese to Chinese. They found that the machine-translated text was inaccurate 14 percent of the time. Moreover, some inaccurate translations involved various regions of the human body, which could lead to a misdiagnosis (Fukushima et al., 2011).

BioMT

Wu and colleagues (2011) developed BioMT, an in-house machine translation system built for translating PubMed titles.

Evaluation of BioMT

Wu and colleagues (2011) evaluated the translation quality of Google Translate and BioMT to translate PubMed titles for patients in six foreign language-English pairs (bi-directionally): French, Spanish, German, Hungarian, Turkish, and Polish. Their judges evaluated the translation based on content (How well the main message of the source sentence is communicated in the translation even if the translation's fluency is terrible?) and fluency (How human like is the translation as a sentence in the target language?). Besides human evaluation, this study also implemented automated evaluations (BLEU scores) to measure translation quality. They reported high performance for German, French, and Spanish – English bi-directional translation pairs for both Google Translate and BioMT, but not for other languages (Wu et al., 2011).

Micro Merchant, OPUS-ISM, and BestRx

These computer programs are used for translating prescription medical labels from English into Spanish. A study found that the quality of the translations was inconsistent and potentially hazardous (Sharif & Tse, 2010).

Other machine translation tools to translate written text

New machine translation technologies have been developed to address specific communication needs. An English-Hindi machine translation system was developed translating texts in homoeopathy (Dwivedi & Sukhadeve, 2013). The BLEU score indicated that the translation accuracy rate was 82.23 percent. Another research team (Muhaxov, Tayila, & Yedemucao, 2016) developed a machine translation system for supporting/improving patient-physician communications in Xinjiang province China, where a large percentage of people living in the countryside of that province do not speak Mandarin. This system translates medical written sentences focusing on long-distance medical and outpatient services among Chinese, Uyghur and Kazakh languages.

BabelMeSH is a cross-language tool for searching Medline/PubMed articles in the user's native language (Spanish, French, or Portuguese). Although it provides an alternative resource for searching PubMed to health care personnel whose native language is not English, it produces translation errors. Therefore, it needs further development (F. Liu et al., 2006).

The Global Public Health Intelligence Network (GPHIN) is an Internet-based "early warning" system that disseminates information about disease outbreaks and other public health issues in multiple languages with machine translation (English, Simplified Chinese, Traditional Chinese, Farsi, French, Russian, Portuguese, Spanish, Russian, and Arabic). The GPHIN employs a "best-of-breed" approach to select the best machine translation device. To date, these technologies have not been evaluated for translation accuracy.

Speech-based systems

Speaking Multilingual Interactive Natural Dialog System (S-MINDS)

S-MINDS is a speech translation system developed by Fluential Inc., to help communication between English and LEP speakers.

Evaluation of S-MINDS

Ehsani, Kimzey, Zuber, Master, and Sudre (2008) evaluated translation accuracy in communications between English-speaking nurses and Spanish-speaking patients in Kaiser Permanente Hospital in San Francisco, CA with 95 nurse-patient encounters (500 conversation segments). The translations were rated on a 5-point scale (good, fair, poor, mistranslated, or not translated). This speech translation system had an overall accuracy rate of 93 percent. However, some sentences were mistranslated or not translated (Ehsani et al., 2008).

Another study evaluated the accuracy of S-MINDS for Spanish-speaking LEP diabetes patients in clinical settings (Soller, Chan, & Higa, 2012). Two researchers who were fluent in medial English and Spanish rated the verbal and audio translation accuracy (correct, conceptually correct, partially correct, and incorrect). The study found that S-MINDS demonstrated high accuracy.

Petit Translator

Petit Translator is a speech translation helping people speaking different languages in the hospital settings. It supports five languages: Japanese, English, Chinese, Korean, and Portuguese. It can operate on smart phones (Ozaki et al., 2011).

Evaluation of Petit Translator

Ozaki and colleagues (2011) conducted an experiment to evaluate the Petit Translator when helping Japanese medical staff to understand Chinese patients. For example, the staff members might ask whether the patient ate breakfast, and then ask the patient to describe a symptom that he or she has suffered. The experiment subjects (9 Japanese and 9 Chinese) rated the efficiency and accuracy of the Petit Translator. The study found that "the Chinese subjects could understand the intentions of the medical staff easily" and "the Japanese subjects understood the patients' intentions well" (Ozaki et al., 2011, p. 395).

MedSLT

MedSLT is a speech-to-speech translation system that helps health care providers communicate with patients speaking a different language, by translating diagnosis questions and answers to English, French, Japanese, Spanish, and Catalan.

Evaluation of MedSLT

In a study (Starlander & Estrella, 2009), eight physicians and sixteen patients participated in the experiments, where physicians determined whether patients had a bacterial infection, and patients simulated a history of symptoms using scripted scenarios. A human and an automatic evaluation scale rated the translation quality of MedSLT V.1 (the constrained version, allows only "yes/no" and ellipsis) and V.2 (the less restricted version, covers full sentences).

Starlander and Estrella (2009) found that MedSLT had certain translation problems that might cause diagnosis errors. Another study (Bouillon, Flores, et al., 2007) evaluated MedSLT V.2 in a simulated medical examination scenario between English-speaking physicians and Spanish-speaking patients, where physicians needed to determine whether the patient had a viral throat infection. Among 42 utterances, 36 were completely correct; 3 were correct in meaning but lacked fluency; and 3 were badly translated. This system was also adapted for Arabic (Bouillon, Halimi, Rayner, & Hockey, 2007). Four translators rated the English-Arabic translations using a three-point scale (good, ok, bad). The authors concluded that although the Arabic translations were promising, they needed further development.

DARwIn-OP robot

DARwIn-OP robot is an English-Korean speech-to-speech translation humanoid robot manufactured by ROBOTIS Inc. in South Korea for assisting

English-speaking patients to describe their symptoms to Korean doctors and nurses. Humanoids can move around the hospital and approach patients in need.

Evaluation of DARwIn-OP robot

Shin and colleagues (2015) conducted an experiment to evaluate DARwIn-OP robot's translation accuracy when translating ten simple sentences. The success rate was 84 percent with native English speakers and 64 percent with non-native English speakers. Among the accurate translation results, the expression was sometimes unnatural. The authors concluded that this technology needs further development (Shin et al., 2015).

iTranslate

iTranslate is a machine language translation mobile application with a voice recognition feature.

Evaluation of iTranslate

Chen and colleagues (2017) evaluated the accuracy of iTranslate when translating three recommended questions for diabetes patients to ask their clinicians (e.g., What are my blood sugar, blood pressure, and cholesterol numbers?). They found that iTranslate generally provided translation accuracy comparable to human translators on simple sentences; however, iTranslate made more errors when translating difficult sentences (Chen et al., 2017).

Converser for Healthcare

Converser for Healthcare is a speech translation system between English and Spanish. Seligman and Dillinger (2015) tested Converser for Healthcare in a large hospital complex in the U.S. by interviewing patients and staff members with regard to their user experience. The study found that the staff members perceived the translation quality as "good enough" (Seligman & Dillinger, 2015). The translation accuracy of this device has not been evaluated in research papers yet.

Overall, the machine translation devices used in a health context at this stage can accurately translate simple sentences but not complex texts or speech. The evaluation studies indicated that the error rates of machine translations were currently unacceptable in actual health settings (Dew et al., 2018).

Section III: implications for future research and practice

Our review of the research on language translation tools currently employed for health communication provides implications for future research and development of tools, as well as for health care providers and practitioners. Further, the current corpus of research suggests protocols for continuing to improve patient-clinician communication when two sides speak different languages. Suggested protocol procedures include the following: (a) performing pre-translation and post-editing, (b) providing health information in multiple languages, (c) increasing the access to professional translators/interpreters, and (d) developing health literacy/English language interventions for LEP populations.

For instance, Dew and colleagues' systematic review study synthesized multiple methods for improving translation results such as making corrections to the translations (post-editing) and providing trainings for machine translation users (Dew et al., 2018). Other methods include adding images to sentences as visual supports (Solano-Flores, Wang, Kachchaf, Soltero-Gonzalez, & Nguyen-Le, 2014) and replacing medical jargon with lay terms before performing the translation (Fukushima et al., 2011; W. Liu & Cai, 2015).

Further, in addition to efforts for improving the quality of machine language translation, providing health information in multiple languages might also increase communication effectiveness (Chen et al., 2016). Take the United States for instance, only 10 percent of the health information websites provide language versions other than English (Becker, 2004). Immigrants who are non-native-English speakers prefer reading health information in their native languages rather than in the language of their adopted country (Singh et al., 2007). Thus, providing health information in multiple languages is critical.

Another strategy for decreasing language barriers in clinical settings is to increase the quality and use of professional translators/interpreters. A study conducted in Norway found that professional language assistance remained underutilized in medical encounters (Kale & Syed, 2010). However, it is important to note that professional human translators also make translation errors that might lead to delayed necessary patient care (Chen et al., 2016). Therefore, we call for more rigorous, evidence-based standards and credentialing for professional human translators and continuing professional development to enhance human translators' skills and knowledge about medical terms and oral as well as written communications between health care professionals and their patients.

Finally, health literacy interventions and tools are essential for decreasing health disparities among linguistically vulnerable populations (e.g., immigrants). Chen and colleagues identified strategies for developing effective health literacy interventions among individuals with LEP. These strategies include (a) grounding efforts in theory and evidence; (b) collaborating with the government, community organizations, English instructors, health professionals, and university researchers; (c) conducting a needs assessment before developing a curriculum; (d) receiving feedback from all curriculum stakeholders; (e) incorporating instructional approaches that are learner centered and interactive; (f) using a variety of classroom activities (e.g., role playing); and (g) integrating technology for improving health literacy skills.

Chen, Goodson, and Acosta point out three limitations of current interventions. First, the interventions' coverage of the health topics is limited. Second, current programs are designed for immigrant adults, not native-born populations or children/ adolescents from bi- or multi-lingual backgrounds. Third, current interventions predominantly focus on Mexican and Chinese immigrants with little or no consideration of other rapidly growing populations such as immigrants from India and Arabic-speaking countries.

Summary

This chapter introduces language translation devices – computer software, websites, and mobile translations applications – used in clinical and public health settings. These devices, at this stage, when used in health contexts can accurately translate simple sentences but not complex texts or speech. The error rates of machine translations that we reported were unacceptable in actual health settings. Performing pre-translation and post-editing, providing health information in multiple languages, increasing the quality and use of professional translators/ interpreters, and developing health literacy/English language interventions for LEP populations are strategies that we have suggested for addressing language barriers in health communication.

References

Albrecht, U.V., Behrends, M., Schmeer, R., Matthies, H.K., & von Jan, U. (2013). Usage of multilingual mobile translation applications in clinical settings. *JMIR Mhealth Uhealth, 1*(1), e4. doi:10.2196/mhealth.2268.

Anazawa, R., Ishikawa, H., Park, M. J., & Kiuchi, T. (2013). Online machine translation use with nursing literature: evaluation method and usability. *Comput Inform Nurs, 31*(2), 59–65.

Anazawa, R., Ishikawa, H., & Takahiro, K. (2013). Evaluation of online machine translation by nursing users. *Comput Inform Nurs, 31*(8), 382–7. doi:10.1097/ NXN.0b013e3182999dc2.

Becker, S.A. (2004). A study of web usability for older adults seeking online health resources. *ACM Trans Comput Hum Interact, 11*(4), 387–406.

Bouillon, P., Flores, G., Starlander, M., Chatzichrisafis, N., Santaholma, M., Tsourakis, N., ..., Hockey, B.A. (2007). *A bidirectional grammar-based medical speech translator.* Paper presented at the Proceedings of the Workshop on Grammar-based approaches to spoken language processing.

Bouillon, P., Halimi, S., Rayner, M., & Hockey, B. A. (2007). *Adapting a medical speech to speech translation system (MedSLT) to Arabic.* Paper presented at the Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources.

Chen, X., Acosta, S., & Barry, A.E. (2016). Evaluating the accuracy of Google translate for diabetes education material. *JMIR Diabetes, 1*(1), e3.

Chen, X., Acosta, S., & Barry, A.E. (2017). Machine or human? Evaluating the quality of a language translation mobile app for diabetes education material. *JMIR Diabetes, 2*(1), e13.

Dew, K.N., Turner, A.M., Choi, Y.K., Bosold, A., & Kirchhoff, K. (2018). Development of machine translation technology for assisting health communication: A systematic review. *J Biomed Inform, 85*, 56–67. doi:10.1016/j.jbi.2018.07.018.

Dwivedi, S.K., & Sukhadeve, P.P. (2013). *Comparative structure of Homoeopathy language with other medical languages in machine translation system.* Paper presented at the International Conference on Advances in Computing, Communications and Informatics.

Ehsani, F., Kimzey, J., Zuber, E., Master, D., & Sudre, K. (2008). *Speech to speech translation for nurse patient interaction*. Paper presented at the Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications.

Fukushima, T., Yoshino, T., & Shigeno, A. (2011). *Development of multilingual interview-sheet composition system to support multilingual communication in medical field*. Paper presented at the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems.

Jacobs, E.A., Karavolos, K., Rathouz, P.J., Ferris, T.G., & Powell, L.H. (2005). Limited English proficiency and breast and cervical cancer screening in a multiethnic population. *Am J Public Health, 95*(8), 1410–16. doi:10.2105/AJPH.2004.041418.

Kale, E., & Syed, H. R. (2010). Language barriers and the use of interpreters in the public health services. A questionnaire-based survey. *Patient Educ Couns, 81*(2), 187–91.

Khanna, R.R., Karliner, L.S., Eck, M., Vittinghoff, E., Koenig, C.J., & Fang, M.C. (2011). Performance of an online translation tool when applied to patient educational material. *J Hosp Med, 6*(9), 519–25.

Kirchhoff, K., Turner, A.M., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: a feasibility study. *J Am Med Inform Assoc, 18*(4), 473–8. doi:10.1136/amiajnl-2011-000176.

Liu, F., Ackerman, M., & Fontelo, P. (2006). BabelMeSH: Development of a cross-language tool for MEDLINE/PubMed. *AMIA Annu Symp Proc, 2006*, 1012.

Liu, W., & Cai, S. (2015). Translating electronic health record notes from English to Spanish: A preliminary study. *Proc BioNLP, 15*, 134–40.

Muhaxov, H., Tayila, S., & Yedemucao, D. (2016). *Multiple-language translation system focusing on long-distance medical and outpatient services*. Paper presented at the IEEE Second International Conference on Multimedia Big Data.

Nair, E.L., & Cienkowski, K.M. (2010). The impact of health literacy on patient understanding of counseling and education materials. *Int J Audiol, 49*(2), 71–5. doi:10.3109/14992020903280161.

Ozaki, S., Matsunobe, T., Yoshino, T., & Shigeno, A. (2011). *Design of a face-to-face multilingual communication system for a handheld device in the medical field*. Paper presented at the International Conference on Human-Computer Interaction.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *BLEU: A method for automatic evaluation of machine translation*. Paper presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics.

Patil, S., & Davies, P. (2014). Use of Google Translate in medical communication: Evaluation of accuracy. *BMJ, 349*, g7392.

Ponce, N.A., Hays, R.D., & Cunningham, W.E. (2006). Linguistic disparities in health care access and health status among older adults. *J Gen Intern Med, 21*(7), 786–91.

Pozo, A.P., Haddad, A.W., Boutin, M., & Delp, E.J. (2011). *A hand-held multimedia translation and interpretation system for diet management*. Paper presented at the Proceedings of the 2011 IEEE International Conference on Multimedia and Expo.

Ryan, C. (2013). Language use in the United States: 2011. *ACS Rep, 22*, 1–16.

Seligman, M., & Dillinger, M. (2015). *Evaluation and Revision of a Speech Translation System for Health Care*. Spoken Translation, Inc. Retrieved from http://workshop2015.iwslt.org/downloads/IWSLT_2015_RP_4.pdf.

Sharif, I., & Tse, J. (2010). Accuracy of computer-generated, Spanish-language medicine labels. *Pediatrics, 125*(5), 960–5. doi:10.1542/peds.2009–2530.

Shin, S., Matson, E.T., Park, J., Yang, B., Lee, J., & Jung, J.-W. (2015). *Speech-to-speech translation humanoid robot in doctor's office.* Paper presented at the 6th International Conference on Automation, Robotics and Applications.

Singh, P.M., Wight, C.A., Sercinoglu, O., Wilson, D.C., Boytsov, A., & Raizada, M.N. (2007). Language preferences on websites and in Google searches for human health and food information. *J Med Internet Res, 9*(2), e18. doi:10.2196/jmir.9.2.e18.

Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., & Nguyen-Le, K. (2014). Developing testing accommodations for English language learners: Illustrations as visual supports for item accessibility. *Educ Assess, 19*(4), 267–83.

Soller, R.W., Chan, P., & Higa, A. (2012). Performance of a new speech translation device in translating verbal recommendations of medication action plans for patients with diabetes. *J Diabetes Sci Technol, 6*(4), 927–37. doi:10.1177/193229681200600426.

Starlander, M., & Estrella, P.S. (2009). *Relating recognition and translation quality with usability of two different versions of MedSLT.* Paper presented at the Proceedings of Machine Translation Summit XII.

Taylor, R.M., Crichton, N., Moult, B., & Gibson, F. (2015). A prospective observational study of machine translation software to overcome the challenge of including ethnic diversity in health care research. *Nurs Open, 2*(1), 14–23. doi:10.1002/nop2.13.

Thomson, M.D., & Hoffman-Goetz, L. (2011). Challenges of recruiting ESL immigrants into cancer education studies: reflections from practice notes. *J Cancer Educ, 26*(1), 170–4.

Toci, E., Burazeri, G., Jerliu, N., Sørensen, K., Ramadani, N., Hysa, B., & Brand, H. (2014). Health literacy, self-perceived health and self-reported chronic morbidity among older people in Kosovo. *Health Promot Int, 30*(3), 667–74.

Turner, A.M., Dew, K.N., Desai, L., Martin, N., & Kirchhoff, K. (2015). Machine translation of public health materials from English to Chinese: A feasibility study. *JMIR Public Health Surveill, 1*(2), e17.

Wilson, E., Chen, A.H., Grumbach, K., Wang, F., & Fernandez, A. (2005). Effects of limited English proficiency and physician language on health care comprehension. *J Gen Intern Med, 20*(9), 800–6. doi:10.1111/j.1525-1497.2005.0174.x.

Wu, C., Xia, F., Deleger, L., & Solti, I. (2011). *Statistical machine translation for biomedical text: Are we there yet?* Paper presented at the AMIA Annual Symposium Proceedings.

Zeng-Treitler, Q., Kim, H., Rosemblat, G., & Keselman, A. (2010). Can multilingual machine translation help make medical record content more comprehensible to patients? *Stud Health Technol Inform, 160*(Pt 1), 73–7.