Application of statistical machine translation to public health information: a feasibility study

Katrin Kirchhoff,[1] Anne M Turner,[2,3] Amittai Axelrod,[1] Francisco Saavedra[3]

[1]Department of Electrical Engineering, University of Washington, Seattle, Washington, USA
[2]Northwest Center for Public Health Practice, University of Washington, Seattle, Washington, USA
[3]Department of Medical Education and Biomedical Informatics, University of Washington, Seattle, Washington, USA

Correspondence to
Professor Katrin Kirchhoff, Department of Electrical Engineering, University of Washington, Box 352500, Seattle, WA 98195, USA; katrin@ee.washington.edu

Received 11 February 2011
Accepted 24 March 2011
Published Online First 15 April 2011

ABSTRACT

Objective Accurate, understandable public health information is important for ensuring the health of the nation. The large portion of the US population with Limited English Proficiency is best served by translations of public-health information into other languages. However, a large number of health departments and primary care clinics face significant barriers to fulfilling federal mandates to provide multilingual materials to Limited English Proficiency individuals. This article presents a pilot study on the feasibility of using freely available statistical machine translation technology to translate health promotion materials.

Design The authors gathered health-promotion materials in English from local and national public-health websites. Spanish versions were created by translating the documents using a freely available machine-translation website. Translations were rated for adequacy and fluency, analyzed for errors, manually corrected by a human posteditor, and compared with exclusively manual translations.

Results Machine translation plus postediting took 15—53 min per document, compared to the reported days or even weeks for the standard translation process. A blind comparison of machine-assisted and human translations of six documents revealed overall equivalency between machine-translated and manually translated materials. The analysis of translation errors indicated that the most important errors were word-sense errors.

Conclusion The results indicate that machine translation plus postediting may be an effective method of producing multilingual health materials with equivalent quality but lower cost compared to manual translations.

INTRODUCTION

Effective communication of health-related information is a key component of health promotion. Over 46 million people in the USA have Limited English Proficiency (LEP), defined as having a primary language other than English and a limited ability to read, speak, write, or understand English.

For these individuals, obtaining accurate and up-to-date health information can be very challenging as a result of language barriers, cultural barriers and low health literacy.[1 2] Despite federal and state regulations mandating improved access to health information for LEP individuals, public-health materials in languages other than English remain scarce.

This article reports the results of a feasibility study investigating freely available statistical machine translation technology as a step in the multilingual document production process. We review requirements for multilingual health materials, provide an overview of current machine translation technology, and report on a pilot study investigating the accuracy, time, and cost associated with a machine-translation-based document production process compared to the standard process of using exclusively manual translations.

BACKGROUND

Information materials for limited English proficiency populations

The ability to access health information in the USA depends greatly on the ability to speak English. Yet there are approximately 300 different languages spoken in the USA. According to the 2009 American Community Survey,[3] 19.6% of the US population over 5 years speak a language other than English at home, and 43.8% of these have LEP. This percentage varies depending on age and the native language, and can constitute as much as 64.1% (Spanish speakers aged 65 years and older).

Studies have found that LEP populations have less access to health education and less preventive health screening, and report a poorer health status than English-speaking minority groups.[4–6] One causal factor is that the vast majority of up-to-date, high-quality healthcare information is published in English. There are few comprehensive health websites, even in Spanish, the second most common language in the USA.[1] Spanish translations of health information available on many websites have been reported to be of poor quality and inconsistent.[7 8] This situation persists despite federal requirements to provide equal access to health services for LEP communities: The 1964 Civil Rights act mandated that no individual should be denied access to services provided by a program receiving federal financial assistance on the grounds of race, color, or national origin. The Supreme Court has subsequently treated native language as equivalent to national origin, resulting in Executive Order 13166 issued in 2000, which requires all federal agencies providing assistance to non-federal entities to issue guidelines on making their services accessible to LEP individuals. The most recent guidelines issued by the Department of Health and Human Services (DHHS)[9 10] specify that recipients of federal funds from DHHS must take 'reasonable steps to provide meaningful access to LEP persons.' This mandate includes making available written translations of vital documents for LEP groups.[9] The Institute of Medicine and the National Library of Medicine have underscored the importance of access to language-specific health information in the fight to reduce health disparities.[11–13] In sum, we need efficient, low-cost ways to convert health information to a variety of languages if we are going to narrow, not widen, the health-disparity gap.

Despite the governmental mandate, there continues to be a lack of multilingual materials for many LEP groups, particularly at the state and community level. Factors contributing to this situation include the lack of standardized processes as well as the time and funds required for producing multilingual documents. An assessment of translation practices at 14 local health departments in Washington State conducted by Public Health-Seattle & King County revealed a wide variation in the procedures and standards used to create multilingual materials.[14] Agencies reported using a mix of in-house bilingual staff and external commercial translators for translation. The typical process of publishing multilingual material consists of the steps shown in figure 1A. Public-health agencies are also subject to marked financial constraints. For example, a medium-sized health department in Washington State reported having a mere $50/month to spend on translations for their county. At an average translation cost of 30 cents/word this allows for translation of 166 words/month or fewer than 2000 words/year. These costs do not yet account for staff time for reviewing and selecting source materials or assuring the quality and cultural appropriateness of the translation.

One obstacle to the faster and more widespread production of multilingual materials is the failure of many health departments to exploit state-of-the-art human-language technology to streamline their services. In particular, machine translation (MT), which in the past has often been regarded as too inaccurate to be useful, has recently made substantial progress and is now widely being used in both the commercial and the non-profit sector. Regional and local health departments could similarly benefit from an MT-supported document production process. Our goal is to investigate a procedure that replaces the steps associated with outsourcing documents for translation (steps B through E in figure 1A) with a freely available machine translation engine that generates quality translations at the click of a mouse button. The translation output is then reviewed and corrected by a human reader (cf figure 1B). This process could reduce both the turnaround time and the cost of translating public health materials and thus, in the long run, enable better access to health information for LEP individuals.

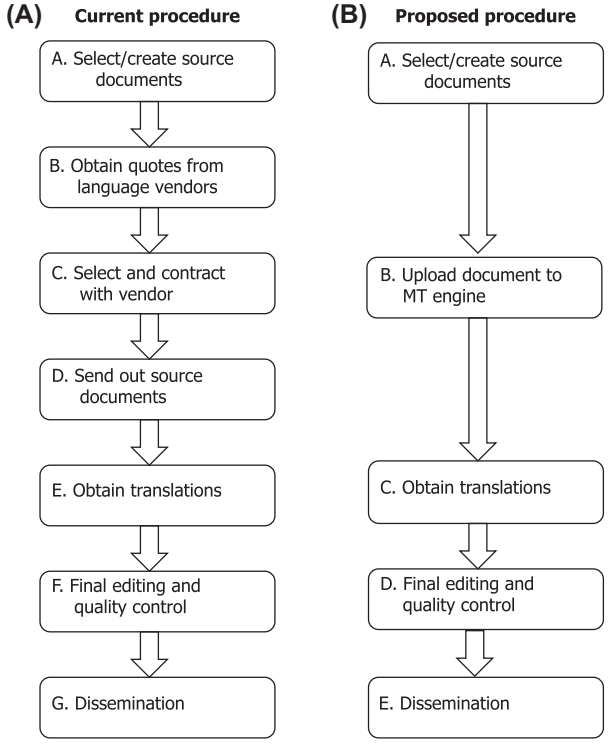


Figure 1 Current and proposed procedures for generating multilingual health materials.

Machine translation

Machine translation (MT), that is, the fully automatic translation of text or speech in one language (the source language) into a different language (the target language), has been an active field of research since the 1940s and has made remarkable progress over the last two decades. Among the various existing MT approaches, statistical MT (SMT) is currently considered the most promising. Under the SMT approach, statistical translation models are learned automatically from large corpora of parallel data, that is, text in the source language paired with its translation in the target language. Translation systems can thus be bootstrapped rapidly for new language pairs (provided that sufficient parallel data are available) without the need for laborious handcrafting of linguistic rules. A detailed overview of SMT technology can be found in Cancedda *et al*.[15] Here, we briefly describe its main features: SMT commonly models a sentence as a concatenation of smaller subsentential chunks (phrases).[16] Phrasal translations are learned automatically by first word-aligning the parallel training data, that is, finding correspondences among individual words in the source sentences and their translations. Alignment information is then used to extract larger matching chunks (phrasal translations, or phrase pairs), whose probabilities are estimated from their relative frequencies in the training data.[16 17] Several other models can be learned concomitantly, such as a reordering model that provides probabilities for reordering phrases relative to their original position in the sentence, and a lexicon model that provides individual word-translation probabilities. Finally, a decoding engine uses these models in combination with a statistical language model trained from a large amount of monolingual data that scores word sequences with respect to their probability of co-occurrence. Out of all possible phrase combinations hypothesized by the decoding engine, the combination with the highest score is chosen as the best translation. The key feature of this approach is that it does not explicitly model discourse, context, or domain information, that is, linguistic and extra-linguistic knowledge sources that are characteristic of the human translation process. It essentially stores a representation of the training corpus and has few ways of generalizing to novel sentences or phrases not present in the training data. For this reason, unfamiliar domains or divergent test data represent a challenge for SMT. One important question thus is whether a generic SMT system performs sufficiently well on texts characteristic of the public-health domain, which may contain specialized vocabulary (eg, medical terminology).

Machine translation output can be evaluated by automatic procedures (eg, BLEU[18]) or human judgments, which are considered more reliable. During human evaluation, evaluators rate individual sentences or paragraphs along two dimensions: adequacy and fluency. Adequacy measures to what extent the information provided in the original document is preserved in the translation output. Fluency measures whether the output conforms to the grammatical rules of the target language. Both are typically rated on a five-point scale. Additional insight into MT performance can be obtained by studies that manually categorize and quantify particular types of errors.[19] In international MT benchmark evaluations, SMT systems have generally outperformed systems based on complex linguistic rules; however, the performance of current systems still varies in

quality from perfect to unintelligible and is strongly dependent on the degree of overlap between training and test data.

Several SMT systems have been made available to the general public on the internet; the most well-known of these is Google Translate (http://www.google.com/translate), which currently comprises MT engines for close to 60 languages. The wide coverage of languages is an advantage over many off-the-shelf commercial systems that often only handle a small number of mainstream languages. In addition, freely available online engines eliminate the need for software purchase, installation, maintenance, and user training.

The current consensus in the MT user community is that in most situations, MT output is not adequate in its raw state, but it is perfectly adequate when postprocessed by a human editor.[20] It has been demonstrated that MT followed by such postediting can lead to substantial time and cost savings over the standard human-only translation process (eg, [21 22]). As a result, most language vendors now make use of some form of MT, and postediting has become part of many translators' standard skill repertoire. Companies such as Intel, Adobe, or Continental Airlines regularly use MT for document and website localization. Several government and non-profit organizations, including the Pan-American Health Organization[23] and the Canadian–UN Global Public Health Intelligence Network project,[24] already make use of machine translation in their workflow processes; however, their MT engines are proprietary systems developed and tuned for in-house needs rather than generic, freely available translation engines. Studies of the usefulness of freely available SMT are beginning to emerge in certain domains (cf Zuo[25] for UN documents). Ramos recently described a language vendor's experience with using Google Translate as a first step in localizing documents for the Word Health Organization's Department of Reproductive Health and Research.[26] We are not aware of any study investigating freely available SMT technology for document translation in public-health settings in the USA.

METHODS

Our objective is to study the feasibility of using a generic, state-of-the-art SMT system followed by human postediting to replace the step of human-only translation in the standard workflow of producing public health materials for LEP audiences. This approach will only be useful if the performance of the MT system is not too poor, that is, postediting should not take an inordinate amount of time and effort. The final documents should be of equivalent quality to those produced by standard human translation and review. Finally, the proposed process should offer time and cost savings vis-à-vis the traditional process. In this pilot study, a set of English public health documents were collected and translated into Spanish by an SMT engine. The output was manually rated and analyzed for errors. The translations were then postedited and reviewed by human evaluators. Time measurements were obtained and compared to the traditional workflow process.

Document collection and translation

We collected 25 English health-promotion documents from various public-health agencies' websites, including Public Health Seattle & King County, the Centers for Disease Control and Prevention, and the DHHS. The consumer-oriented information covered a variety of health and safety topics (eg, HIV, maternal health, floodwater emergencies, rat infestation, etc). The documents were passed through Google Translate (Mexican Spanish option) (http://translate.google.com/toolkit). Since Google Translate uses existing parallel web data for training translation models, we took care to include documents that do not have a corresponding website in Spanish in addition to those that do.

Translation quality evaluation and error analysis

Each translated document was analyzed by two native speakers of Spanish with fluent knowledge of English. The analysis was restricted to the main text of each page, excluding side bars, figures, etc. Translation errors were identified and classified into one of the six categories listed in table 1, and the percentage of errors was computed for each category.

Note that a simple overall error percentage (the total number of errors divided by the total number of words) is not an adequate measure of translation accuracy: a given source language word can correspond to several words in the target language. Additionally, one word may combine several translation errors—for example, it can have the wrong morphologic form and can also be in the wrong position in the sentence, yielding two error counts. In order to obtain quantitative measurements of translation quality, we conducted human evaluations, adopting the commonly used criteria of fluency and adequacy (see section 'Machine translation'). In line with previous evaluation studies,[27] both fluency and adequacy were rated on the five-point scale shown in table 2.

A randomly selected subset of 385 sentences was subsequently rated by two native speakers of Spanish fluent in English. Evaluators were presented with each sentence within a context window consisting of three sentences immediately preceding or following the current sentence, and with the automatic translation. An initial calibration exercise was carried out on a separate document. During this exercise, evaluators first rated each translated sentence individually and then compared their ratings to ensure that their interpretations of the scales shown in table 2 did not diverge too strongly. They then evaluated the set of 385 sentences independently. The means and standard deviations were computed from the two sets of scores, and the interannotator agreement was computed in the form of a weighted kappa coefficient:

$$K = \frac{P_o - P_e}{1 - P_e}$$

where

$$P_o = \sum_{ij} \left(1 - w_{ij}\right) \frac{n_{ij}}{N}$$

and

$$P_e = \sum_{ij} \left(1 - w_{ij}\right) \frac{n_i n_j}{2N}$$

N is the number of scores (385), i and j range over the possible ratings (1 through 5), n_{ij} is the number of times the combination

Table 1 Error categories for English to Spanish translation

Category	Description
Missing term	A term essential to the meaning of a sentence has been left out
Untranslated word	A source-language word is left untranslated
Word sense error	A word sense inappropriate to the context was chosen
Syntactic error	Wrong word order
Morphologic error	Wrong inflectional word form (eg, singular subject but plural verb form)
Other grammatical error	Missing particle, preposition, etc
Pragmatic error	Culturally inappropriate translation (eg, unintentionally comical or offensive)

Table 2 Rating scales for fluency and adequacy

Level	Fluency (how grammatically correct the translation is)	Adequacy (how much of the original meaning is preserved in the translation)
5	Flawless Spanish	All
4	Good Spanish	Most
3	Non-native Spanish	Much
2	Disfluent Spanish	Little
1	Incomprehensible	None

of ratings i and j was seen, and w_{ij} is a combination-specific weight, that is, computed as the difference between i and j, divided by the maximum possible difference (4). In this way, each combination of scores is weighted by the relative distance between the ratings. An outcome where, for example, Annotator 1 assigns a 1, and Annotator 2 assigns a 4 is weighted more strongly than a divergence of one point on the five-point scale.

Postediting and final quality comparison

As explained above, the use of MT in our intended setting requires a postediting stage during which machine translation errors are corrected by a human editor. For this pilot study, 13 of the translated documents were postedited by a public-health professional with native knowledge of Spanish and fluency in English. One additional document was used to familiarize the posteditor with the postediting guidelines in an initial practice session.

The posteditor was instructed to apply all and only those edit operations (deleting, adding, or replacing words, changing the positions of words, changing punctuation) necessary to ensure that the output was (a) grammatical; (b) conveyed the same meaning as the original English document; (c) was culturally appropriate; and (d) preserved the linguistic style of the source document. Extensive rewriting was discouraged. Prior to postediting, the editor was asked to read the source document in order to identify potential comprehension problems. Postediting of each document was then performed in a single, uninterrupted time period, using the interface provided in the Google Translator Toolkit (http://translate.google.com/toolkit). Up to five documents were processed per session. Time measurements were taken for each documents, starting with the click of the 'Send' button to generate the translation and ending when the posteditor signaled that he had finished editing the document.

As a final evaluation step, we asked a medical professional (a native speaker of Spanish fluent in English) to perform a blind comparison of original human translations of six randomly selected English documents and the corresponding postedited machine translations (the evaluator did not know which translation was human- or machine-derived). The evaluator was asked to indicate whether the translations were equivalent and, if not, which translation was preferred and why.

RESULTS

Human evaluation of adequacy and fluency yielded a mean fluency score of 3.73 (SD 0.74) and a mean adequacy score of 4.19 (SD 0.71). The interevaluator agreement was 0.85.

The detailed error analysis is shown in table 3. The most significant error categories are morphologic errors, word sense errors, and other grammatical errors. Of these, annotators found the word-sense errors to be the most disruptive to human processing and understanding. An example of a word-sense error is the use of the Spanish term *junta* for *board* in the sentence *A glue board is sometimes used to catch a rat* (from a document on pest control). Here, *board* was translated in the sense of a *governing body* rather than *wooden plank*.

Postediting took between 15 and 53 min per document, with an average of 30 min.

The average throughput was 2.4 words/min. Of note, the typical translation and review process for health department information is reported to take between 2 and 7 days when translations are outsourced and subsequently go through a similar post-translation process to insure accuracy and clarity. The final quality comparison of human-only versus machine-assisted translations on six documents showed the following results:

- Clear preference for human-only translation: two documents
- Clear preference for postedited machine translation output: two documents
- Minimal preference for the human-only translation: one document
- Human-only and machine-assisted translations were equivalent: one document

Reasons for preferences included better fluency of the human-only translations and higher adequacy of the MT-supported translations (the translation was closer to the source text). No difference in translation quality was observed for those source documents that did have existing translations on the web versus those that did not.

DISCUSSION

There is a growing need for rapid and cost-effective generation of quality translations of public-health material, particularly in emergency situations. Our results show that adequate translations of public health materials can be generated within a short period of time (sometimes on the order of minutes) by substituting the traditional process of manual translation, which often takes days or weeks, with machine translation plus postediting. While the accuracy of MT certainly needs to be improved, especially in the realm of word sense errors, our results also indicate that our proposed process results in translations that are comparable in quality to human-only translations.

With respect to the final quality evaluation, it might seem surprising that postedited machine translation output is sometimes even preferred to translations produced by human translators. However, it is often the case that posteditors expect machine translation output to contain more errors than human-generated translations and therefore scrutinize the text more closely, resulting in a better final product. In addition, human translators also might take undue liberties in translating, reshaping the original text in a way deemed more appropriate for the target audience. When considering a final quality comparison, it is important to bear in mind that the

Table 3 Results of error analysis on Spanish machine translations

Error category	Percentage
Missing term	4.43
Untranslated word	3.75
Word sense error	21.97
Syntax error	18.43
Morphologic error	24.39
Other grammatical error	22.59
Pragmatic error	4.44

evaluation of machine-translated documents judges both the contributions of the machine translation component and the quality of the postediting; any disfluencies or errors left in the translations are an indication of poor postediting. It is noteworthy that we observed overall quality equivalence, even though our posteditor was not a language professional, suggesting that postediting could be performed by bilingual public-health staff.

The quality equivalence and time savings suggested by our pilot study represent a conservative estimate of the potential gains, for the following reasons: the translation system was a generic MT engine that was not tuned to our particular domain. In most applications of MT, the basic system can be fine-tuned by including specialized dictionaries, terminology lists, and translation memories that store translations and corrections produced previously. Furthermore, our posteditor was not a trained translator and did not have prior experience in postediting. Typically, the speed and accuracy of postediting increases with experience. For these reasons, we consider the use of MT for translating public-health materials a promising future direction.

Limitations

The scope of this feasibility study was restricted to a small sample of documents, determined by access to already-translated materials, availability of expert annotators, and time and financial resource constraints. We therefore do not claim the results are statistically significant. It was also beyond the scope of this study to conduct side-by-side time and cost comparisons of the existing versus proposed translation process on the same set of input documents. The primary purpose of this study was to determine whether current MT technology was sufficiently accurate to warrant an expanded study, and to design and test the evaluation methodology. A larger study is currently under way. However, in the light of reported translation practices at local and regional public-health departments, our results suggest that machine translation may offer substantial time and cost savings.

Another question is to what extent the proposed approach can be applied to languages other than Spanish. Spanish is a well-researched language with a grammatical structure similar to English, and a large body of parallel Spanish—English data is available to train SMT systems. Translation between languages with highly divergent linguistic structures or fewer data resources is more difficult. We have completed an initial error analysis on machine translations of English public-health documents into Vietnamese, a language with a different structure and comparatively few parallel training data. The dataset used for this purpose only included 10 documents and was deemed too small to be included in the above report. Initial results indicate that MT performance is lower than in the case of Spanish but still sufficient to generate usable results in combination with postediting. The nature of machine translation errors is similar to those obtained on the Spanish data; in particular, word-sense errors predominate.

Future work

Several questions resulting from this study will be addressed in future work. In addition to extending the present analysis to more languages and larger data sets, we will investigate how SMT can be best integrated into the workflow of a typical (regional) public-health agency. Second, we will examine which machine translation errors are the most disruptive to human processing and understanding, and thus deserve priority in addressing the shortcomings of current MT technology. Third, we plan to devise appropriate ways of improving MT technology for our purpose (including on-the-fly context disambiguation and terminology management).

CONCLUSIONS

Our pilot study indicates that machine translation technology holds great promise for assisting health agencies with the growing task of providing quality translations of health promotion materials to individuals with LEP. Although machine translation quality is imperfect, it is sufficiently accurate to be used as the initial step in a human—machine collaborative translation framework that involves error correction and fine tuning by a posteditor. Such a framework would greatly facilitate the process of producing translated materials by reducing the time and financial resources required to generate quality translations for those in need.

Acknowledgments We would like to thank I Hendrickson, for providing assistance with document collection, and D Capurro Nario, for providing quality judgments on translated materials. In addition, we would like to thank members of the PHSKC communication teams, in particular H Karasz and M Valenzuela, for providing us with information about their current translation processes.

Funding This work was supported by the Northwest Preparedness and Response Research Center (PERRC) grant number P01TP000297, CDC Center of Excellence in Public Health Informatics grant P01 HK 000027, the National Library of Medicine Medical Informatics Training Grant T15 LM007442-07, National Library of Medicine Grant 1R01LM010811-01 and a grant from the University of Washington's Provost's Office to KK.

Competing interests None.

Ethics approval The University of Washington Institutional Review Board approved this study.

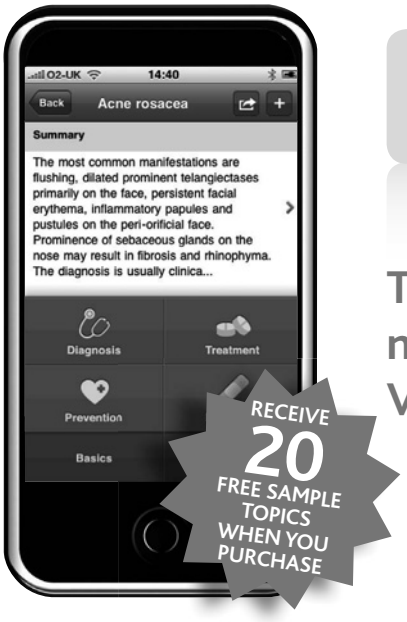
Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Cardelle AJF,** Rodriguez EG. The quality of Spanish health information websites. *J Prev Interv Community* 2006;**29**:85—102.
2. **Peña-Purcell N.** Hispanics' use of internet health information: an exploratory study. *J Med Libr Assoc* 2008;**96**:101—7.
3. **ACS.** *American Community Survey*, 2009. http://www.census.gov/acs/www/ (accessed 31 Jan 2011).
4. **Goel MS,** Wee CC, McCarthy EP, *et al*. Racial and ethnic disparities in cancer screening: the importance of foreign birth as a barrier to care. *J Gen Intern Med* 2003;**18**:1028—35.
5. **Jacobs EA,** Shepard DS, Suaya JA, *et al*. Overcoming language barriers in health care: costs and benefits of interpreter services. *Res Pract* 2004;**94**:866—9.
6. **Ponce NA,** Hays RD, Cunningham WE. Lingustic disparities in health care access and health status among older adults. *J Gen Intern Med* 2006;**21**:786—91.
7. **Berland GK,** Elliott MN, Morales LS. Health information on the internet: accessibility. *J Am Med Assoc* 2001;**285**:2612—21.
8. **Turner AM,** Stavri Z, Revere D, *et al*. From the ground up: determining the information needs and uses of local public health nurses in Oregon. *J Med Libr Assoc* 2008;**96**:335—42.
9. **DHHS.** Department of Health and Human Services-Guidance to federal financial assistance recipients regarding title VI prohibition against national origin discrimination affecting limited english proficient persons. *Fed Regist* 2003;**68**:47311—23.
10. **DoJ.** *Report to Congress. Assessment of the Total Benefits and Costs of Implementing Executive Order No. 13166: Improving Access to Services for Persons with Limited English Proficiency*. Washington DC: 2002. http://www.usdoj.gov/crt/cor/lep/omb-lepreport.pdf.
11. **IOM.** *Institute of Medicine. Health Literacy: A Prescription to End Confusion*. Washington, DC: National Academy Press, 2004.
12. **NLM.** *Charting a Course for the 21st Century—NLM's Long Range Plan 2006—2016*. 2007. http://www.nlm.nih.gov/pubs/plan/lrp06/report/executivesummary.html (accessed 23 Sep 2009).
13. **DHHS.** *Department of Health and Human Services Centers for Disease Control and Prevention*. Washington DC: Local Solutions to Racial & Ethnic Health Disparities, 2007.
14. **Valenzuela M,** Beebe A, Beleford J, *et al*. *Internal and External Translation Assessment Summary—October 2008. Research Summary*. Seattle, WA: Public Health-Seattle & King County, 2008.

15. **Cancedda N,** Dymetman M, Foster G, *et al*. A statistical machine translation primer. In: Goutte C, Cancedda N, Dymetman M, *et al*, eds. *Learning Machine Translation*. Cambridge, Mass: The MIT Press, 2009:1—37.
16. **Koehn P,** Och FJ, Marcu D. Statistical phrase-based translation. *Proceedings of the Meeting of the North American Association for Computational Linguistics (NAACL)*. Edmonton, Canada: Association for Computational Linguistics, 2003:48—54.
17. **Och FJ,** Zens R, Ney H. Phrase-Based Statistical Machine Translation. *KI (German Conference on Artificial Intelligence)*. Berlin: Springer, 2002:18—32.
18. **Papineni K,** Roukos S, Ward T, *et al*. A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Los Angeles, CA: Association for Computational Linguistics, 2002:311—18.
19. **Popovic M,** Ney H. Error Analysis of Verb Inflections in Spanish Translation Output *Error Analysis of Verb Inflections in Spanish Translation Output*. Barcelona, Spain, 2006:99—103.
20. **Doyon J,** Doran C, Means CD. Automated machine translation improvement through post-editing techniques: analyst and translator experiments. *Proceedings of the Conference of the American Machine Translation Association (AMTA)*. Waikiki, Hawai'i: AMTA, 2008.
21. **Allen J.** Case study: implementing machine translation for the translation of pre-sales marketing and post-sales software deployment documentation at mycom international. In: Richardson SD, ed. *Machine Translation: From Real Users to Research*. Berlin: Springer, 2004:1—6.
22. **Groves D.** Bringing humans into the loop: localization with machine translation at traslan. *Proceedings of the Conference of the American Machine Translation Association (AMTA)*. Waikiki, Hawai'i: AMTA, 2008.
23. **Aymerich J.** *Using Machine Translation for Fast, Inexpensive, and Accurate Health Information Assimilation and Dissemination: Experiences at the Pan American Health Organization*. Salvador - Bahia, Brazil: 9th World Congress on Health Information and Libraries, 2005.
24. **Blench M.** Global public health intelligence network (GPHIN). *Proceedings of the Conference of the American Machine Translation Association (AMTA)*. Waikiki, Hawai'i: AMTA, 2008.
25. **Zuo L.** Performance of Google translation tool on UN documents. *Proceedings of the American Machine Translation Association (AMTA)*. Denver, CO: AMTA, 2010.
26. **Ramos LC.** Post-editing free machine translation: from a language vendor's perspective. *Proceedings of the Conference of the American Machine Translation Association (AMTA)*. Denver, CO: AMTA, 2010.
27. **Callison-Burch C,** Fordyce C, Koehn P, *et al*. (Meta-)evaluation of machine translation. *Proceedings of the Second ACL Workshop on Statistical Machine Translation*. Prague, Czech Republic: Association for Computational Linguistics, 2007:136—58.

Have confidence in your decision making.



The best clinical decision support tool is now available as an app for your iPhone. Visit bestpractice.bmj.com/app

BestPractice
FROM THE BMJ EVIDENCE CENTRE

clinicians • medical students • nurses • healthcare practitioners