

Translating Electronic Health Record Notes from English to Spanish: A Preliminary Study

Weisong Liu, PhD

University of Massachusetts
Medical School,
Worcester, MA

Weisong.Liu@umassmed.edu

Shu Cai

Information Sciences Institute
Marina del Rey, CA

shuca@isi.edu

Balaji P Ramesh, PhD

University of Massachusetts
Medical School,
Worcester, MA

balaji288@gmail.com

Germán Chiriboga, MPH

University of Massachusetts
Medical School,
Worcester, MA

German.Chiriboga@umassmed.edu

Kevin Knight, PhD

Information Sciences Institute
Marina del Rey, CA

knight@isi.edu

Hong Yu, PhD

University of Massachusetts
Medical School,
Worcester, MA

Hong.Yu@umassmed.edu

Abstract

The Centers for Medicare & Medicaid Services Incentive Programs promote meaningful use of electronic health records (EHRs), which, among many benefits, allow patients to receive electronic copies of their EHRs and thereby empower them to take a more active role in their health. In the United States, however, 17% population is Hispanic, of which 50% has limited English language skills. To help this population take advantage of their EHRs, we are developing English-Spanish machine translation (MT) systems for EHRs. In this study, we first built an English-Spanish parallel corpus and trained *NoteAid_{Spanish}*, a statistical MT (SMT) system. Google Translator and Microsoft Bing Translator are two baseline MT systems. In addition, we evaluated hybrid MT systems that first replace medical jargon in EHR notes with lay terms and then translate the notes with SMT systems. Evaluation on a small set of EHR notes, our results show that Google Translator outperformed *NoteAid_{Spanish}*. The hybrid SMT systems first map medical jargon to lay language. This step improved the translation. A fully implemented hybrid MT system is available at <http://www.clinicalnotesaid.org>. The English-Spanish parallel-aligned MedlinePlus corpus is available upon request.

1 Introduction

The Centers for Medicare & Medicaid Services Incentive Programs promote meaningful use of electronic health records (EHRs), which, among many benefits, allow patients to receive electronic copies of their health records and

thereby empower them to take a more active role in their health. EHRs present a new and *personalized* communication channel that has the potential to increase patient involvement in care and improve communication between physicians and patients and their caregivers. In particular, allowing patients access to their physicians' notes has the potential to enhance patients' understanding of their conditions and disease and improve medication adherence and self-managed care.

However, most EHRs are written in English. In the United States, 17% population is Hispanic, of which 50% has limited English language skills. Many general-purpose MT systems are available. For example, Google Translate is a free service that has been used by health professions. Like most general-purpose MT systems, it is based on SMT, looking for patterns in hundreds of millions of WWW documents. In contrast, EHRs contain medical terms, shortened forms, complex disease and medication names, and other domain-specific jargon that do not typically appear in WWW documents, and therefore Google Translate may not perform well for EHRs, as was found in a prior study that evaluated general-purpose MT systems (Zeng-Treitler et al., 2010). Furthermore, the Health Insurance Portability and Accountability Act of 1996 protects the privacy and security of individually identifiable health information, so a secure MT system may be needed for US hospitals.

Therefore we are developing an EHR domain-specific English-Spanish MT system called *NoteAid_{Spanish}*, which may help over 37 million Spanish speaking US residents to meaningfully use their EHRs.

2 Background

MT has been an active research field for the past 60–70 years. Early systems mainly applied bilingual dictionaries and manually crafted rules. However, since the 1990s, research has turned to SMT (Brown et al., 1990). The best SMT systems are built from translation patterns that are learned automatically from parallel, human-translated text corpora (Koehn, 2010). Translation patterns include phrase translations that translate input text by translating sequences of words at a time (Koehn et al., 2003; Och, 2002), re-ordering tendencies allowing swapping of words or phrases (Tillmann, 2004), hierarchical phrase translations with variables (Chiang, 2007), and syntax-based transformations (Galley et al., 2004). Automatic learning enables systems to imitate human translation behavior and adapt to particular domains. The bulk of current MT research is tested on domains such as news and politics. The BLEU (Papineni et al., 2002) score is a standard evaluation metric for MT. It measures n-gram overlap with human translations and has shown correlation with human judgment.

Comparatively few MT systems have been developed in the medical domain. Early work focused on knowledge-based approaches for phrase translation (Eck et al., 2004; Humphrey et al., 1998; Liu et al., 2006; Merabti et al., 2011). Several research groups built parallel corpora, then trained SMT systems (Wu et al., 2011; Yepes et al., 2013). The Shared Task of Medical Translation provided both parallel aligned and monolingual corpora (Bojar et al., 2014). Eight teams participated the shared task and most of the systems were based on the Moses phrase-based toolkit with in-domain and out-of-domain language models.

Zeng-Treitler et al (Zeng-Treitler et al., 2010) evaluated a general-purpose MT tool called Babel Fish to translate 213 EHR note sentences from English into Spanish, Chinese, Russian, and Korean and then evaluated the comprehensibility and accuracy of the translation. They found, however, the majority of the translations were incomprehensible and/or incorrect.

3 Methods

We first built a domain-specific English-Spanish parallel aligned corpus and then developed and evaluated SMT and hybrid machine translation (HMT) systems for translating EHR notes from English to Spanish. This study was approved by the Institutional Review Board of University of Massachusetts Medical School. All EHR notes have been deidentified.

3.1 English-Spanish Parallel Aligned Bio-medical Corpora

The MedlinePlus ($ESPAC_{MedlinePlus}$)

Source: The NIH’s MedlinePlus ((U.S.),) web site hosts web pages of medical articles of different health topics. Most of the articles in English have a corresponding Spanish version translated by human. 2,999 articles have Spanish translations, which we crawled to build the parallel aligned corpus. We conducted data cleaning and sentence alignment. We split $ESPAC_{MedlinePlus}$ into a training set (60%), a tuning set (20%) and a testing set (20%) by interleaving sentence by sentence. Table 1 shows the statistics of the data. Unknown words or word types on the English side are 4,580 and 3,308 for tuning and 4,558 and 3,309 for testing.

	Sentence Pairs	Word tokens (English)	Sent. Length (English)	Word tokens (Spanish)	Sent. Length (Spanish)
Training	85,540	1,005,342	11.7	1,135,080	13.27
Tuning	29,299	341,821	11.7	386,754	13.20
Testing	29,258	338,431	11.6	382,239	13.06

Table 1. Statistics of $ESPAC_{MedlinePlus}$

The EHR Corpus ($ESPAC_{EHR}$)

The UMass Amherst Translation Center translated three de-identified EHR notes (108 sentences, 13.4 word tokens per sentence, and a total of 1,445 words) from English to Spanish.

3.2 MT Systems

Phrase-Based SMT

Using $ESPAC_{MedlinePlus}$, we trained an initial phrase-based Moses (Koehn et al., 2007) system. The training aligns the words in sentence pairs and extracts phrase pairs consistent with those alignments. We set the maximum phrase pair length to 7 words. We trained a 3-gram language model on the Spanish side using SRILM (Stolcke, 2002; Stolcke et al., 2011). We first used the default feature weights in Moses, then adjusted these feature weights using MERT (Och, 2003).

HMT Systems

EHR notes contain medical jargon that differs significantly from the consumer-oriented medical corpora most MT systems are trained on. We therefore speculate that if we replace medical jargon with lay terms and then feed the transformed EHR note to a SMT system, we may improve the MT performance. In our HMT system, we first applied the Metemap (Aronson, 2001) to map free text to UMLS concepts. For those mapped concepts, we replace the medical jargon with lay terms. A concept is clinically relevant if it belongs to one of the 18 UMLS semantic types, as described in the *NoteAid* system (Ramesh et al., 2013). A term is a lay term if it appears in the Consumer Health Vocabulary of the UMLS. A term is also a lay term if it appears in MedlinePlus. We also identify abbreviations and replace them with their expanded full terms. The second component of the HMT systems is an SMT system. We explored two state-of-the-art SMT systems, Google Translate and Microsoft Bing Translator, resulting in two HMT systems, *NoteAid-Google_{Spanish}* and *NoteAid-Bing_{Spanish}*.

Baseline MT Systems

The baseline systems are the state-of-the-art general purpose Google and Bing MT systems in which EHR notes are directly fed into the systems without any medical jargon replacement.

3.3 Evaluation Metrics and Procedure

All the MT systems were evaluated by single-reference, case-insensitive BLEU score using the Moses package. We also asked a bilingual domain expert to manually evaluate the five MT system outputs of the three EHR notes.

4 Results

4.1 Automatic Evaluation

The BLEU score of *NoteAid-Moses_{Spanish}* on the tuning and testing medical parallel data are 41.8 (1.097) and 41.2 (1.104) before MERT and 50.4 (0.99) and 49.8 (0.99) after MERT. The BLEU score of Google Translate, which was 49.9 (0.99). Table 2 shows the performance (macro-average) of the translation systems on the three de-identified EHR notes. We found that 17.9% of all terms in EHR notes do not appear in the MedlinePlus corpus,

	BLEU score (ave. \pm SD)	Sentence length ratio (ave. \pm SD)
Bing	21.33 \pm 7.38	1.02 \pm 0.07
<i>NoteAid-Bing_{Spanish}</i>	18.17 \pm 7.38	1.03 \pm 0.07
Google	14.05 \pm 6.30	1.24 \pm 0.04
<i>NoteAid-Google_{Spanish}</i>	11.05 \pm 5.63	1.23 \pm 0.04
<i>NoteAid-Moses_{Spanish}</i>	5.82 \pm 1.95	1.10 \pm 0.02

Table 2. MT systems on ESPAC_{EHR}

4.2 Evaluation by a Domain Expert

A bilingual human expert performed a blind review of the outputs of all five MT systems on the three EHR notes (a total of 15 Spanish outputs). He ranked all five MT systems. In addition, he marked up the errors by each MT system.

The expert judged that each MT system had a few translation omissions. For example, “symptomatically,” was omitted by all the MT systems. Of the three EHR notes, *Google Translate* performed the best for two. *NoteAid-Google_{Spanish}* and *NoteAid-Bing_{Spanish}* were second on three. Bing Translator was the best for one. *NoteAid-Moses_{Spanish}* was the last.

The expert also performed a blind comparison of *Google Translate* versus *NoteAid-Google_{Spanish}*. He found that the hybrid system simplified the medical jargon and translated well. However, it introduced inconsistencies a few times. Therefore, the rating for *Google translation* is slightly better on two out of the three EHR notes.

5 Discussion

There are a number of challenges for translating EHR notes from English to Spanish. Spanish translation frequently increases token length. In addition, rhetoric styles differ, which can considerably affect text length in cases where the medical note is more of a narrative than a sequence of facts and isolated sentences (Valero-Garces, 1996). Finally, it is expensive to create English-Spanish parallel aligned EHR corpora.

Both *NoteAid-Moses_{Spanish}* and Google Translate achieved a competitive performance for *ESPAC_{MedlinePlus}*. Several factors could have contributed to the excellent MT performance. Since 25% of our data is redundant, during the training process the decoder memorized those sentences. This combined with the fact that the total percentage of unknown words and sentences were small (~16%) may have contributed to the good results. In addition, we

found that 37% of the sentences in the tuning and testing sets had less than seven words, and about half of those sentences overlapped with the training set. These sentences were memorized as phrases during training, although their contribution to the overall performance was less significant than longer sentences. Finally, translating sentences with one word is easier than translating sentences with multiple words because one-word sentences do not have a re-ordering problem, which is one of the challenges in MT.

The evaluation of MT systems on EHR notes (Table 2) showed much reduced performance. The results are not surprising since 17.9% terms in EHR notes do not appear in the MedlinePlus.

In addition, all HMT systems performed worse than their SMT counterparts. The lower performance of HMT systems can be attributed to the lack of gold standards that exactly match the source text of hybrid systems. The gold standard consists of original English notes translated to Spanish by human translators. But, the HMT systems modify the original notes by replacing the medical jargon with lay terms and then translate the notes to Spanish. Since, the BLEU score calculates what percentage of the n-grams or phrases from the translations also appear in the gold standard and the HMT systems modify the original text before translation, it is expected to yield a lower performance.

We also found that sentences in EHR notes were not always grammatically well formed. Whereas, when humans translated the text, they inferred the context from the note and formed coherent and logical sentences by inserting the missing verb or conjunction. The translation systems translated the original ill-formed sentences into Spanish word for word. This resulted in a lower BLEU score performance for MT systems.

Our manual analyses show that the baseline and the HMT systems perform well and make very few mistakes on EHR notes. The mistakes include:

- Translation omission when they encounter typos in the source language. For example, the MT systems failed to translate typos like “possily” and “phychological.”
- Failure to take context into consideration when translating the text. For example, in

“we are redrawing blood cultures,” the MT systems failed to recognize that “redraw” refers to removing blood cultures, and translated it as “redibujando” or “rediseñando,” meaning redrawing or redesigning something.

- Incorrect grammatical gender assignment although the translation is correct. For example, “Skin: Warm and dry” is translated as “Piel: Cálido y seco” ignoring the fact that the grammatical gender context of “Piel”/Skin is feminine.
- Errors in verb conjugation. For example, “to drain” is translated as “para drenar” instead of “á drenar.”

We select and describe three examples of errors by MT systems, as shown below.

In the example below, all the five MT systems fail to accurately translate the sentence and change the meaning when translated back to English. We also observed that human translators often translate the text using different words while maintaining the semantic sense of the sentence.

Source: Acute renal failure with neutropenia likely medication induced

Human Translation: fallo renal grave con neutropenia probablemente debido a medicamento.

Human Back Translation: severe renal failure with neutropenia probably due to medication

Google Translate: La insuficiencia renal aguda con neutropenia probable medicación inducida

Human Back Translation: acute renal failure with neutropenia probably induced medication

Bing Translate: Insuficiencia renal aguda con medicación probable neutropenia inducida

Human Back Translation: acute renal failure with medication, probably induced neutropenia

NoteAid-Moses_{Spanish}: insuficiencia renal aguda con la neutropenia probable Medicines induced

Human Back Translation: Probable medication induced acute renal failure with neutropenia

NoteAid-Google_{Spanish}: insuficiencia renal aguda con neutropenia probables Medicamentos inducidos

Human Back Translation: acute renal failure with neutropenia, probable induced medications

NoteAid-Bing_{Spanish}: la insuficiencia renal aguda con neutropenia indujeron probables medicamentos

Human Back Translation: acute renal failure with neutropenia induced probable medications

In this example, *NoteAid-Moses_{Spanish}* conserves only some of the source text’s context and format but omits translation of several words, including medical jargon. The *NoteAid-Bing_{Spanish}* omits only one word but the remaining MT systems do not omit any word. The Google translate and both the hybrid systems make a grammatical mistake by assigning incorrect gender to the patient in Spanish.

Source: ASSESSMENT AND PLAN: The patient was scheduled for a kidney biopsy today, but she was informed by the Renal Transplant Service that they were going to delay this since there was some improvement in her creatinine (today's creatinine is not yet available).

Human Translation: EVALUACION Y PLAN: Se proyectaba que la paciente tuviese una biopsia del riñón hoy, pero el Servicio de Trasplante Renal le informó que iban a retrasarla pues ha habido una mejora en su creatinina (la creatinina de hoy todavía no está disponible).

Google Translate: EVALUACIÓN Y PLAN: **El** paciente fue programado para una biopsia de riñón hoy, pero fue informado por el Servicio de Trasplante Renal de que iban a demorar esto, ya hubo alguna mejora en su creatinina (creatinina de hoy todavía no está disponible).

Bing Translate: EVALUACIÓN Y PLAN: La paciente estaba programada para una biopsia de riñón hoy, pero fue informada por el servicio de Trasplante Renal que iban a retrasar esto ya que hubo cierta mejora en la creatinina sérica (creatinina de hoy aún no está disponible).

NoteAid-Moses^{Spanish}: ASSESSMENT AND PLAN: The paciente se programado para una biopsia del riñón today, pero que ella estaba informado por la Renal Transplant Service que fueron de irse a retrasar este dado que no hubo alguna mejora en su creatinina (today's creatinina aún no se available).

NoteAid-Google^{Spanish}: EVALUACIÓN Y PLAN: **El** paciente fue programado para una biopsia de riñón hoy, pero fue informado por el Servicio de Trasplante Renal de que iban a demorar esto, ya hubo alguna mejora en sus creatininas (hoy creatininas aún no está disponible).

NoteAid-Bing^{Spanish}: EVALUACIÓN Y PLAN: **El** paciente fue programado para una biopsia de riñón hoy, pero fue informada por el servicio de trasplante de riñón que iban a retrasar esto ya que hubo cierta mejora en sus creatinines (hoy creatinines aún no está disponible).

Unlike the narrative in the previous examples, the non-narrative context consists of a list of medical jargon, the format of which is quite common in clinical notes. While the baseline and hybrid MT translation systems perform well, *NoteAid-Moses^{Spanish}* omits many of the terms.

Source: Lungs: bilateral: reduced breath sounds, expiratory wheezes present (in the upper lobe, in the lower lobe), at or near baseline? , no rhonchi present, . Skin: lesion, . b/l venous stasis changes on distal anterior tibias trace tibial edema

Human Translation: Pulmones: bilateral: sonidos respiratorios reducidos, presencia de resoplido espiratorio (en el lóbulo superior, en el lóbulo inferior), ¿en o cerca de la línea de base? , no hay presencia de estertor roncus, . Piel: lesión, . cambios bilaterales de estasis venosa en tibias anteriores distales indicio edema tibial

Google Translate: Pulmones: bilateral : ruidos respiratorios reducida , sibilancias espiratorias presentes (en el lóbulo superior , en el lóbulo inferior) , en o cerca de la línea de base ? , No hay roncus presentes , . Piel: lesión , . b / l estasis venosa cambios en la tibia distal tibial anterior traza edema

Bing Translate: Pulmones: bilateral: sonidos respiratorios reducida, resuellos espiratorios presentes (en el lóbulo superior, en el lóbulo inferior), en o cerca de línea de base. , presentes, no roncus. Piel: lesión. b/l estasis venosa cambios en tibias anteriores distales rastrear el edema tibial

NoteAid-Moses^{Spanish}: Lungs: bilateral: reducción de aliento sounds, wheezes presente (in dispositivos de la parte superior del lobe, en la parte inferior de lobe), , en o cerca de baseline?

, no present, estertores . Skin: lesion, . b/l de estasis venosa cambios en el edema tibial anterior distal tibias trace

NoteAid-Google^{Spanish}: Pulmones: bilateral: reducción de sonidos pulmonares espiratorio sibilancias presentes (en el lóbulo superior , en el lóbulo inferior) , en o cerca de la línea de base la visión ? , No hay rhonchis presentes , . Piel: lesión . cambios b / l venostasis en distal rastro tibias anterior tibial Edema

NoteAid-Bing^{Spanish}: Pulmones: bilateral: reducido sibilancias espiratorio de sonidos pulmonares presentes (en el lóbulo superior, en el lóbulo inferior),, en o cerca de base de la visión? , no rhonchis presente,. Piel: lesión. cambios b/l lindo tibias anteriores distales rastrear el Edema tibial

NoteAid-Moses^{Spanish} performed poorly on the EHR notes, suggesting that the system needs to be trained on bigger data sets, or be trained directly on the EHR notes. We found that some errors by *NoteAid-Google^{Spanish}* were due to engineering errors, which can be fixed.

6 Limitations, Conclusion and Future Work

This pilot study has limitations. The SMT system was built on the limited MedlinePlus data. We plan to incorporate other biomedical corpora (e.g., Medline and ClinicalTrial.gov). The corpus size of EHR notes for evaluation is small and we plan to build such a corpus.

The BLEU score does not provide a measurement in terms of whether the semantic content is correctly translated. In the future work we may explore other domain-specific evaluation metrics (Castilla et al., 2005).

In this application, we have experimented with simple MT approaches. In the future we may explore other MT approaches, including incorporating biomedical knowledge resources (e.g., the UMLS), domain adaptation, semantic role labelling and abstract meaning representation.

Acknowledgement: The authors thank the anonymous reviewers for invaluable comments.

References

- A. R. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp*:17–21.
- M Baldry, C Cheal, B Fisher, M Gillett, and V Huet. 1986. Giving patients their own records in general practice: experience of patients and staff. *British Medical Journal (Clinical research ed.)*, 292(6520):596–598, March. PMID: 3081187PMCID: PMC1339574.
- Ondrej Bojar, Christian Buck, Christian Federmann, Barry Haddow, Philipp Koehn, Johannes Leveling, Christof Monz, Pavel Pecina, Matt Post, Herve Saint-Amand, and others. 2014. Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 12–58. Association for Computational Linguistics Baltimore, MD, USA.
- Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational linguistics*, 16(2):79–85.
- A. C. Castilla, A. S. Bacic, and S. S. Furuie. 2005. Machine Translation on the Medical Domain: The Role of BLEU/NIST and METEOR in a Controlled Vocabulary Setting. *Proceedings of the Tenth Machine Translation Summit. Phuket, Thailand*:47.
- David Chiang. 2007. Hierarchical phrase-based translation. *computational linguistics*, 33(2):201–228.
- James J Cimino, Vimla L Patel, and Andre W Kushniruk. 2002. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. *International journal of medical informatics*, 68(1-3):113–127, December. PMID: 12467796.
- Tom Delbanco, Jan Walker, Sigall K. Bell, Jonathan D. Darer, Joann G. Elmore, Nadine Farag, Henry J. Feldman, Roanne Mejilla, Long Ngo, James D. Ralston, Stephen E. Ross, Neha Trivedi, Elisabeth Vodicka, and Suzanne G. Leveille. 2012. Inviting Patients to Read Their Doctors’ Notes: A Quasi-experimental Study and a Look Ahead. *Annals of Internal Medicine*, 157(7):461–470, October.
- Darren A DeWalt, Robert M Malone, Mary E Bryant, Margaret C Kosnar, Kelly E Corr, Russell L Rothman, Carla A Sueta, and Michael P Pignone. 2006. A heart failure self-management program for patients of all literacy levels: a randomized, controlled trial [ISRCTN11535170]. *BMC health services research*, 6:30. PMID: 16533388.
- Matthias Eck, Stephan Vogel, and Alex Waibel. 2004. Improving statistical machine translation in the medical domain using the Unified Medical Language System. In *Proceedings of the 20th international conference on Computational Linguistics*, page 792.
- Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What’s in a Translation Rule? In *HLT-NAACL*, pages 273–280.
- B. Humphrey, D. A. B. Lindberg, H. M. Schoolman, and G. O. Barnett. 1998. The Unified Medical Language System: An Informatics Research Collaboration. *Journal of the American Medical Association*, 5:1–11.
- Philipp Koehn. 2010. *Statistical Machine Translation*. Cambridge University Press, 1 edition edition, January.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54.
- P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, and others. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180.
- Fang Liu, Michael Ackerman, and Paul Fontelo. 2006. BabelMeSH: development of a cross-language tool for MEDLINE/PubMed. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*:1012. PMID: 17238631 PMCID: PMC1839504.
- Bradley M Mathers, Louisa Degenhardt, Hammad Ali, Lucas Wiessing, Matthew Hickman, Richard P Mattick, Bronwyn Myers, Atul Ambekar, and Stefanie A Strathdee. 20. HIV prevention, treatment, and care services for people who inject drugs: a systematic review of global, regional, and national coverage. *The Lancet*, 375(9719):1014–1028.
- Michael Meltsner. 2012. A Patient’s View of Open-Notes. *Annals of Internal Medicine*, 157(7):523–524, October.
- Tayeb Merabti, Lina F. Soualmia, Julien Grosjean, Olivier Palombi, Jean-Michel Müller, and Stéfan J. Darmoni. 2011. Translating the Foundational Model of Anatomy into French using knowledge-based and lexical methods. *BMC medical informatics and decision making*, 11(1):65.

- F.J. Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 160–167.
- Franz Josef Och. 2002. *Statistical machine translation: from single-word models to alignment templates*. Ph.D. thesis, Bibliothek der RWTH Aachen.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Balaji Ramesh, Thomas Houston, Cynthia Brandt, Julia Fang, and Hong Yu. 2013. Improving Patients' Electronic Health Record Comprehension with NoteAid. *The 14th World Congress on Medical and Health Informatics. Best Student Paper*.
- Dean Schillinger, Margaret Handley, Frances Wang, and Hali Hammer. 2009. Effects of self-management support on structure, process, and outcomes among vulnerable patients with diabetes: a three-arm practical clinical trial. *Diabetes care*, 32(4):559–566, April. PMID: 19131469.
- J F Seitz, A Ward, and W H Dobbs. 1978. Granting patients access to records: the impact of the Privacy Act at a federal hospital. *Hospital & community psychiatry*, 29(5):288–289, May. PMID: 640590.
- Andreas Stolcke. 2002. SRILM—An extensible language modeling toolkit. In *In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*, pages 901–904.
- Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at sixteen: Update and outlook. In
- Christoph Tillmann. 2004. A unigram orientation model for statistical machine translation. In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 101–104.
- National Library of Medicine (U.S.). Fact SheetMedlinePlus®.
- C. Valero-Garces. 1996. Contrastive ESP Rhetoric: Metatext in Spanish-English Economics Texts. *English for Specific Purposes*, 15(4):279–294, November.
- Warren J. Winkelman, Kevin J. Leonard, and Peter G. Rossos. 2005. Patient-Perceived Usefulness of Online Electronic Medical Records: Employing Grounded Theory in the Development of Information and Communication Technologies for Use by Patients Living with Chronic Illness. *Journal of the American Medical Informatics Association*, 12(3):306–314, May.
- Cuijun Wu, Fei Xia, Louise Deleger, and Imre Solti. 2011. Statistical Machine Translation for Biomedical Text: Are We There Yet? *AMIA Annual Symposium Proceedings*, 2011:1290–1299. PMID: 22195190PMCID: PMC3243244.
- Antonio Jimeno Yepes, Élise Prieur-Gaston, and Aurélie Névéol. 2013. Combining MEDLINE and publisher data to create parallel corpora for the automatic translation of biomedical text. *BMC Bioinformatics*, 14(1):146, April. PMID: 23631733.
- Qing Zeng-Treitler, Hyeoneui Kim, Graciela Rosemblat, and Alla Keselman. 2010. Can multilingual machine translation help make medical record content more comprehensible to patients? *Studies in health technology and informatics*, 160(Pt 1):73–77. PMID: 20841653.
2014. ACL 2014 Ninth Workshop on Statistical Machine Translation Shared Task: Medical Translation. Technical report.