Automated Interpretation of Clinical Encounters with Cultural Cues and Electronic Health Record Generation

Daniel T. Heinze, PhD
A-Life Medical, Inc.
6055 Lusk Blvd. – Suite 200
San Diego, CA 92130
dheinze@alifemedical.com

Alexander Turchin, MD, MS
Brigham and Women's Hospital
221 Longwood Ave.
Boston, MA 02115
aturchin@partners.org

V. Jagannathan, PhD
West Virginia University and MedQuist, Inc.
235 High Street, Suite 2I3
Morgantown, WV 26505
juggy@medquist.com

Abstract

A review of publications by and about medical interpreters reveals a number of operational similarities and shared attitudes and beliefs with the medical coding and abstracting community as it existed ten years ago in the mid-1990's. At that time, the first of what have now become several successful commercial products using Natural Language Processing (NLP) for automated coding and abstracting appeared. The initial reaction was that machines could never do what human coders and abstractors do, and anecdotal accounts illustrating the difficulty of the task proliferated. The claims of superior human capabilities and the accuracy of the anecdotal accounts were and are substantially true, but the fact is that the machines are more capable than what they were initially given credit for, and the percentage of cases that can be handled with automation fairly well approximates the 80/20 rule.

In this paper, we present an early stage prototype medical interpreter system that is based on lessons learned in developing successful automated coding and abstracting systems and on the core infrastructure and techniques used in these systems. Specific techniques include leveraging standards based multi-lingual medical nomenclatures and clinical ontology systems, machine awareness of difficult situations, explanatory meta-knowledge, and an interactive environment that emphasizes the strengths of both the human and machine participants and mitigates the weaknesses of each.

1 Introduction

The task of medical interpretation is demanding and difficult, and although U.S. hospitals that receive federal funds are required to provide interpreter services, the demands on the system are generally beyond the availability of qualified interpreters. Less than one fourth of U.S. hospitals have professionally trained interpreters and among these, many have no training in medical terminology. [Loviglio, 2004] After noting that well-to-do, educated patients have a relatively similar grasp of the process and content of medical care regardless of national or cultural origin, Haffner [1992] describes a variety of scenarios in which communication regarding medical treatment is far more difficult with the poor and under-educated. Karliner, Perez-Stable and Gildengorin [2004] formalize the study of medical interpreting, expand on the ad hoc observations of interpreters, and detail many of the challenges and pitfalls that befall the medical interpreter as well as the errors in medical care that may arise from inadequate cross-language and cross-cultural communications. These include hesitancy of patients to communicate fully and openly with physicians due to embarrassment or cultural norms, misunderstandings regarding offered treatments based on differing medical prac-

tices in the patient's native environment, and, in some cases, a lack of terminology by which western medical concepts can be easily translated to the patient.

As in other areas of medicine such as medical coding and abstracting, the problem of an interpreter shortage is not likely to be self-limiting. For this reason, machine translation has undeniable interest. The demands of medicine, however, require that the matter be approached in a manner different or more comprehensive than those employed in translating web pages or interpreting tourism related queries and responses. Specific needs of both physicians and patients motivate the quest for medically accurate and culturally attuned communication. Experience in building successful systems that use NLP to automate medical coding and abstracting tasks teaches that success is achieved not in trying to create a machine that replaces the human but rather is achieved by creating a machine that assists and augments the human practitioner. Specifically, the machine should offload the portions of the job that are mundane, repetitive and that can be successfully automated. In the coding and abstracting area where A-Life Medical processes the free-text transcriptions for over two million clinical encounters per month, this equates to about 70% of the total volume. [Morris, et al., 2000] A second aspect of successful human-machine collaboration is that the machine needs the ability to accurately differentiate between language-based content that it can process independently and that which requires human review and/or intervention. We call this semi-knowledge in that it corresponds to the human capability to recognize that an utterance is of importance to the task at hand even though the full intent is not comprehended. In such cases, the machine must, like a human, seek expert guidance. In this regard, the machine will address the issue to a human expert, but the strengths of the machine can be used to provide on-line help and meta-data as an aid to the human expert. This is particularly helpful in the medical field where the sheer volume of knowledge is frequently beyond the ability of humans to keep in ready memory.

In regard to the volume of knowledge, the problem in medicine is in part mitigated by the ongoing developments in the area of medical ontologies that provide for the unambiguous representation of the majority of clinical concepts. In particular, this project relies on the Systematic Nomenclature of Medicine – Clinical Terminology (SNOMED-CT®)[1] for the core, multilingual nomenclature of clinical concepts and the Clinical Document Architecture, Release 2.0 (CDA2) [Dolin, et al., 2005] for the framework by which complex clinical events and communications can be represented using the core nomenclature.

2 Motivation

More than 21 million residents of the United States speak English poorly or not at all and for more than 46 million, English is not the first language. [Karliner, Perez-Stable and Gildengorin, 2004] In many urban settings, sizable minorities of patients speak a language other than English. [Loviglio, 2004] Unless the communications needs of both the physician and patient are met, the possibility for serious medical errors is exacerbated.

2.1 What Physicians Need

All areas of physician-patient communication are important to the quality of care, but the quality of communication that is required can be divided according to those communications that are only of immediate importance during the course of the encounter and those that have durable importance beyond the temporal scope of the encounter. For example, physician directives for the patient to stand, bend, take a deep breath, etc. are in the immediate class and the accuracy of a translation (often augmented by signing, example, and physical manipulation) can be easily judged by the patient's actions. Conversely, acquiring the patient history, the review of systems, explaining diagnoses and prescribing medications and a course of treatment

[1] This material includes SNOMED Clinical Terms® (SNOMED CT®), which is used by permission of the College of American Pathologists. ©2002-2006 College of American Pathologists. All rights reserved. SNOMED CT has been created by combining SNOMED RT® and a computer based nomenclature and classification known as Clinical Terms Version 3, formerly known as the Read Codes Version 3, which was created on behalf of the U.K. Department of Health and is a Crown Copyright. SNOMED and SNOMED CT are registered trademarks of the College of American Pathologists.

have import that continues beyond the time scope of the encounter in that they become part of the permanent record, are a basis for both current and future medical decision making, and are critical to accurate completion of the course of care. Further, it is the physician's responsibility, as the care provider, to ensure that the patient has been understood and that the patient understands the nature of their condition and the planned course of treatment. Without a means to validate the communications that represent what we are calling the durable aspects, the physician can neither be sure nor give assurance that the communications have been accurate, and that the course of treatment is appropriate.

Another important area where medical interpretation services are needed is physician-to-physician communication. Telemedicine is on the rise in the United States [Bauer, 2002] and worldwide. [Sood, Bhatia, 2005] Many applications of telemedicine involve communication between physicians located in different countries. [Wachter, 2006] Effective physician-to-physician communication usually requires proficiency in nuances medical terminology and can be challenging even for physicians who are fluent in lay language. [Bruzzi, 2006]

2.2 What patients need

Although communication that is primarily physician directed with yes/no or multiple-choice patient responses can cover a lot of territory, there are several areas for which it is necessary that the patient be able to have a more comprehensive input. These include the expression of concerns about the severity and prospective outcome relative to their medical condition, and communication of issues relative to their life situation that contributed to their condition or that may affect their ability to follow medical instructions. Karliner, et al. [2004] found that even when using interpreters, physician satisfaction levels with regard to their ability to elicit exact symptoms, explain treatments, elicit treatment preferences, and empower patients with regard to their own care was far lower than for the physician's satisfaction with their ability to diagnose and treat the medical condition. During medical encounters in which the physician and patient speak the same language, the physician may likely initiate dialogue on these topics with open-ended questions such as "How did this happen?", "Do you have any other questions?", "Does this concern you?", and the like. Because the answers to these questions may be complex and may be influenced by cultural sensitivities [Hudelson, 2005], cultural context must be accounted for in the design of an automated medical interpreter system.

2.3 Statistical translation systems

Statistical translations systems have been developed for many language pairs, but due to the nature of the available parallel training corpora one language in most pairs is English. [Waibel, et al., May 2004] Further, statistical translation systems rely on large parallel corpora for training and these may not be available for applications in clinical medicine. Advances in the general state of machine translation can be tracked in the results of the NIST Machine Translation Evaluations. [NIST, 2005] [Papineni, 2002] [Zhang, Vogel and Waibel, 2004] A more complete analysis that includes measures of adequacy, fluency, and meaning maintenance can be found in Eck and Hori [2005], who provide the following medical example that is illustrative of the unevenness in these three areas of measurement and demonstrates the need for other methods beyond straight statistical translation.

Reference	**i would like to have an allergy test please**
Translation 1	i would like to have an allergy test please
Translation 2	could you check I am allergic
Translation 3	i would like to make a
Translation 4	allergic to order room service please

Statistical systems can also be time and resource intensive [Fung, et al., 2004] such that quality must be sacrificed for speed in applications that require near real-time response. [Peterson, 2006]

2.4 Interlingua translation systems

Interlingua approaches for clinical applications seem to have a preferred status, in part due to the constrained nature of clinical speech and in part due to the ability to provide a structured back-translation from the interlingua to the physician's language for confirmation of adequacy, and in part

due to the fact that the interlingua provides a formally represented, deep analysis of meaning. [Schultz, et al., 2004] Two approaches to interlingua are common:

1. a formal representation of meaning [Bouillon, 2005]
2. a natural language, usually English, as interlingua. [Waibel, et al., May 2004]

With regard to using a formal interlingua, mapping speech to an unambiguous formal representation that can be validated by the speaker provides the requisite accuracy and a basis for accurate translation, but the time and expense required to build such a system for all patient languages is prohibitive. Secondly, a formal interlingua will not easily capture many nuances of natural language.

The use of a natural language as the interlingua provides the ability to represent a greater range of nuance, although all languages have subtleties that cannot easily be translated. Natural languages, however, introduce the problem of double translation errors. Further, the interlingua may be a language or format inaccessible to either speaker in a conversation, and so there is no way for either speaker to validate.

3 Approach

Our approach, currently designated as Accultran/Med or just Accultran (for Accurate, Acculturated Translator) is based, both philosophically and in terms of implementation, on LifeCode® NLP system that has been developed at A-Life Medical for coding and abstracting clinical documents. A complete description is beyond the scope of this paper but is available in [Heinze, et al. 2001]. Automated Speech Recognition (ASR) is performed using the SpeechMagic™ system from Philips, which is currently available for twenty-three languages. Non-CDA2/SNOMED-CT translations are via AltaVista Babelfish.

Many of the techniques in our approach are established in the practice. Particularly we note the use of physician directed communication with yes/no patient responses, back-translation on the physician side, and the use of multiple choice answer selections for patient responses. [Kazunori, et al., 2006] Beyond this, we are exploring the use of CDA2 and SNOMED-CT as the interlingua for use in those portions of the encounter where clinical accuracy is essential, the use of semi-knowledge for recognizing when an encounter is potentially moving in directions where cross-cultural communications problems may arise, and the use of patient waiting time for patient directed acculturation based on patient complaint information collected upon presentation. The primary emphasis here will be on the use of CDA2 and SNOMED-CT.

Based on the previous observations regarding physician and patient communication needs during a clinical encounter, we divide the clinical encounter into the following aspects: 1) establishing rapport; 2) chief complaint; 3) history; 4) review of systems; 5) physical examination; 6) diagnoses; 7) procedures; 8) medications; 9) instructions. Except for item 1, these all correspond to sections of the traditional clinical note or report and as such have extensive representations in CDA2 and SNOMED-CT. The Continuity of Care Document (CCD), a current effort to harmonize the ASTM and CDA2 standards in this realm, attempts to focus just on these elements using CDA2 representation capabilities. CDA2 is primarily declarative with some capabilities to represent contingencies. This is essentially what is needed for presenting information, but much of the encounter requires query and response.

The core of Accultran resides in the capability of the NLP engine to determine the appropriate context for each physician utterance and to appropriately process and route the content of the utterance. The overall communications flow for the system is illustrated in Figure 1, showing that upon receiving and processing an utterance from the physician, the NLP engine can choose one of several courses of action:

(1) Utterances that contain clinical questions or clinical statements for the patient to affirm or deny or instructions are: (a) converted to CDA2 and are (b) processed by a style sheet that produces the question/statement (c) first for physician validation and then (d) mapped to the patient language with, as needed, a request to affirm or deny.

(2) Utterances with content that cannot be converted to CDA2 are (a) routed to a general machine translation system, (b) optionally with back-translation and physician approval before (c) presentation to the patient.

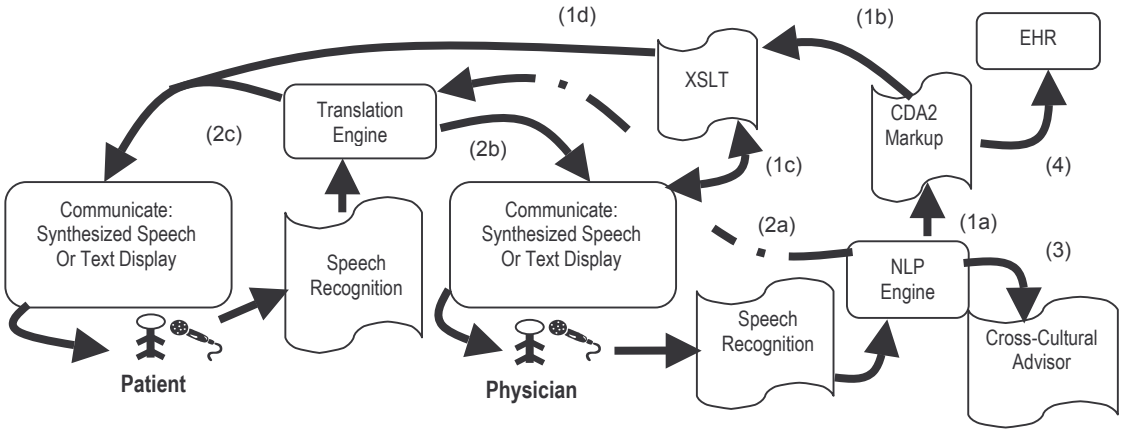

Figure 1: Framework for automated medical interpretation with cultural queues and EHR construction.

(3) Utterances that contain references to subject matter that is deemed culturally sensitive or subject to misunderstanding will trigger the Cross-Cultural Advisor.

(4) As the encounter progresses, the NLP engine appropriately directs information to the EHR via CDA2 for later physician review, and, as needed, revision.

The Cross-Cultural Advisor (CCA) module is a key feature. Technically it is based on the NLP engine's capability for recognizing and flagging clinical content that requires special attention beyond what the NLP system can independently provide. In this case, the flags are associated with warnings related to subject matter that is known to have either cultural sensitivities for patients in the target language group or that is difficult to translate into the target language. Options that the CCA could present to the physician for any particular flag would include warnings with explanation of the sensitivity, pre-formulated queries or informational presentations that are designed to mitigate any misunderstandings, or advise that a human interpreter be involved. In cases where the services of a human interpreter are called for, the CCA identified topic can be used to select, when available, an interpreter with training or skills appropriate to the case at hand. This can be particularly useful when Video Medical Interpretation (VMI) capabilities are used and there is a pool of remote interpreters from which to select.

Actual extracts (labels and annotation added) of a simplified encounter follow.

```
Physician: "Good morning."
Accultran: "Good day."
Physician: "Okay."
Accultran: "Buonas díaz."
Patient: "Buonas díaz."
Accultran: "Good day."
Physician: "What seems to be
the problem?"
Accultran: "What seems to be
the problem?"
Physician: "Okay."
Accultran: "¿Qué se parece
ser el problema?"
Patient: "Caí en la ducha y
lastimé mi muñeca." (patient
points to left wrist)
Accultran: "I fell in the
shower and hurt my wrist."
Physician: "You fell in the
shower and you have pain in
your left wrist?"
Accultran: "Please confirm:
fall in shower?"
Physician: "Okay."
Accultran: "Confirme por fa-
vor: ¿caída en ducha?"
Patient: "Si."
Accultran: "Yes."
Accultran: "Please confirm:
left wrist pain?"
```

Physician: “Okay.”
Accultran: “Confirme por favor: ¿dolor izquierdo estructura de la región de la muñeca?”
Patient: “Si.”
Accultran: “Yes.”
Physician: “Do you have any other pain?”
Accultran: “Do you have other pain?”
Physician: “Okay.”
Accultran: “¿Usted tiene otro dolor?”
Patient: “No.”
Accultran: “No.”

Examination follows a similar course of dialogue.

Physician: “You have a bruise on the left forearm and a mild sprain of the left wrist.”
Accultran: “You have: left forearm contusion; mild left wrist sprain. Do you understand?”
Physician: “Okay.”
Accultran: “Usted tiene: contusion izquierdo del antebrazo. ¿Usted entiende?”
Patient: “Si.”
Accultran: “Yes.”
Accultran: “Usted tiene: esguince izquierdo leve estructura de la región de la muñeca. ¿Usted entiende?”
Patient: “Si.”
Accultran: “Yes.”
Physician: “Instructions: no lifting for one week; take Tylenol 650 mg qid for two days.”
Accultran: “Instructions: no lifting; duration: 1 week. Medication: Tylenol, quantity: 650 millegrams, frequency: every 6 hours, duration: 2 days.”
Physician: “Okay.”

```
<section>
 <title>Question</title>
 <entry>
  <observation>
   <code
     codeSystemName="SNOMED CT"
     displayName="Chief Complaint" />
    <value xsi:type="CD"
     code="44188002"
     codeSystemName="SNOMED CT"
     displayName="fall in shower">
    </value>
  </observation>
  <observation>
     codeSystemName="SNOMED CT"
     displayName="Chief Complaint" />
    <value xsi:type="CD"
     code="22253000"
     codeSystemName="SNOMED CT"
     displayName="pain">
      <qualifier>
        <name
           displayName="finding site" />
        <value
           codeSystemName="SNOMED CT"
           displayName="wrist" />
      </qualifier>
      <qualifier>
        <name
           displayName="laterality" />
        <value
           codeSystemName="SNOMED CT"
           displayName="left" />
      </qualifier>
    </value>
  </observation>
 </entry>
</section>
```

Figure 2: Abbreviated SNOMED CT for Queries

Accultran: “Instrucciones: ninguna elevación; duración: 1 semana. ¿Usted entiende?”
Patient: “Si.”
Accultran: “Yes.”
Accultran: “Medicación: Tylenol, cantidad: 650 millegrams, frecuencia: cada 6 horas, duración: 2 días. ¿Usted entiende?”
Patient: “Si.”
Accultran: “Yes.”
Physician: “Do you have any other concerns?”
Accultran: “Do you have any other concerns?”
Physician: “Okay.”
Accultran: “¿Usted tiene otras preocupaciones?”
Patient: “Estoy asustado para mi bebé.”(Patient points to abdomen) “Mi marido está eno-

jado que pude haber lastimado al bebé."
Accultran: "I'm afraid for may baby. My husband is angry that I may have hurt the baby."
Physician: "Are you pregnant?"
Accultran: "Are you pregnant?"
Physician: "Okay."
Accultran: "¿Es usted embarazado?"
Patient: "Si."
Accultran: "Yes."
Physician: "Is your husband angry with you?"
Accultran: "Warning: …" (Accultran produces a cultural warning relative to the importance of bearing children in Hispanic cultures, marital relations…. The decision is made to involve an interpreter with skills in pregnancy and domestic issues.)

4 Discussion

As stated in the Introduction, Accultran is still an early prototype. The NLP engine has been evaluated at Partners Healthcare for mapping clinical free-text to CDA2. Publication of these results is expected in the near future. Development of the translation aspects is not yet sufficiently mature for field testing. Per work at A-Life, we currently see a number of strengths and several particular shortcomings with regard to CDA2 and SNOMED-CT as a framework for documenting and communicating clinical encounters. The strengths are in the coverage of medical concepts, the ability to formally assemble concepts in a coherent representation of an encounter, and the ability to easily map that formal representation to a variety of applications via XSLT (XML Style Sheets) and alternate language representation. However, although nomenclatures such as SNOMED-CT provide coverage for concepts such as embarrassment, inappropriate behavior, identification of cultural and value components related to pain management etc., they do not provide information or insights into the actual cultural components that affect these concepts. The cultural components must be developed separately and added to the system as metadata attached to specific semi-knowledge entries with attached flags and helps. This notion can be further expanded so as to use the considerable waiting time that patients typically experience in medical settings. During this time the patient would interact with the system, which would provide language and culture specific materials to educate and acculturate the patient.

Finally, and of no small import, SNOMED-CT is currently only available in two versions of English (US and UK), German and Spanish. In the US, at least, Spanish would be one of the primary languages in need. Although there are no current plans for complete and official versions of SNOMED-CT in languages other than those just noted (personal communication with author's SNOMED account manager), our requirements are for only a limited subset of the terms that could be independently translated in a commercial setting.

5 Conclusion

The difficulty of medical interpreting and the potential medical consequences should not be underestimated. Aside from the difficulties in sheer vocabulary size and multi-lingual representation, there are the added complications of diverse cultures. We have presented an architecture that addresses the issues of medical accuracy and cultural sensitivity. Although the use of such a system requires some patience and acclimation on the part of both medical practitioners and patients, the cost is small as compared to that of any morbidity or mortality that could result from inaccurate communication.

References

Bauer JC 2002. Insights on Telemedicine: How Big Is the Market? *Journal of Healthcare Information Management* 16(2): 10-11.

Bouillon P, Rayner M, Chatzichrisafis N, Hockey BA, Santaholma M, Starlander M, Isahara H, Kanzaki K, Nakao Y. May 2005. A generic Multi-Lingual Open Source Platform for Limited-Domain Medical Speech Translation. *Proceedings of the tenth Conference on European Association of Machine Translation.* Budapest, Hungary. 5-58.

Bruzzi JF 2006. The Words Count — Radiology and Medical Linguistics. *The New England Journal of Medicine* 354(7): 665-667.

Dolin RH, Alshhuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo A, editors. 2005. HL7 Clinical Document Architecture, Release 2.0. ANSI-approved HL7 Standard, May 2005. Ann Arbor, MI: Health Level Seven, Inc.

Eck M, Hori C. 2005. Overview of the IWSLT 2005 Evaluation Campaign. *International Workshop on Spoken Language Translation.* Pittsburgh, PA.

Flores G. 2005. The Impact of Medical Interpreter Services on the Quality of HealthCare: A Systematic Review. *Medical Care Research and Review,* 62(3): 255-299.

Fung P, Yi L, Yongsheng Y, Shen Y, Wu D. October 2004. A Grammar-Based Chinese to English Speech System for Portable Devices. *Interspeech-ICSLP*. Jeju, Korea.

Haffner L. September 1992. Translation is Not Enough: Interpreting in a Medical Setting. *Western Journal of Medicine,* 157(3).

Heinze DT, Morsch ML, Sheffer RE, Jimmink M, Jennings MA, Morris WC, Morsch AE. 2001. LifeCode: A Deployed Application for Automated Medical Coding. *AI Magazine,* 22(2): 76-88.

Hudelson P. 2005. Improving patient-provider communication: Insights from interpreters. *Family Practice*, 22(3): 311-316.

Imoto K, Sasajima M, Shimomori T, Yamanaka N, Yajima M, Masai Y. 2006. A multi modal supporting tool for multi lingual communication by inducing partner's reply. *Proceedings of the 11th International Conference on Intelligent User Interfaces.* Sydney, Australia.

Karliner LS, Perez-Stable EJ, Gildengorin G. February 2004. The Language Divide: The Importance of Training in the Use of Interpreters for Outpatient Practice. *Journal of General Internal Medicine,* 19(2): 175ff.

Loviglio J. November 22, 2004. Interpreters Lower Risks in Hospitals. Associated Press story available at http://ap.lancasteronline.com/4/hospital_babel Accessed Feb 28, 2006.

Morris WC, Heinze DT, Warner Jr. HR, Primack A, Morsch AE, Sheffer RE, Jennings MA, Morsch ML, Jimmink M. 2000. Assessing the accuracy of an automated coding system in emergency medicine. *Proceedings of the AMIA 2000 Annual Symposium.* Los Angeles, CA. 595-599.

NIST 2005 Machine Translation Evaluation Official Results. Version 3. August 1, 2005. http://www.nist.gov/speech/tests/mt/mt05eval_official_results_release_20050801_v3.html Accessed Feb 28, 2006.

Papineni K, Roukos S, Ward T, Zhu WJ. July 2002. Bleu: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA.

Peterson R. Jan 24, 2006. IBM Strives for Superhuman Speech Tech. *Pcmag.com* http://www.pcmag.com/article2/0,1895,1915071,00.asp Accessed Feb 28, 2006.

Schultz T, Alexander D, Black AW, Peterson K, Suebvisai S, Waibel A. 2004. A Thai Translation Stystem for Medical Dialogs. *Proceedings: Human Language Technologies*. Boston, MA.

Sood SP, Bhatia JS. 2005. Development of telemedicine technology in India: "Sanjeevani"-An integrated telemedicine application. *Journal of Postgraduate Medicine,* 51(4): 308-311.

Starlander M, Bouillon P, Rayner M, Chatzichrisafis N, Hockey BA, Isahara H, Kanzaki K, Nakao Y, Santaholma M. 2005. Breaking the Language Barrier: Machine Assisted Diagnosis using the Medical Speech Translator. *Proceedings of the XIX Internation Congress of the European Federation for Medical Informatics MIE.* Geneva, Switzerland.

Wachter RM 2006. The "Dis-location" of U.S. Medicine — The Implications of Medical Outsourcing. *The New England Journal of Medicine,* 354(7): 661-665

Waibel A, Schultz T, Vogel S, Fügen C, Honal M, Kolss M, Reichert J, Stüker S. May 2004. Towards Language Portability in Statistical Speech Translation. *Special Session on Multilinguality in Speech Processing, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.* Montreal, Canada.

Zhang Y, Vogel S, Waibel A. May 2004. Interpreting BLEU/NIST Scores: How Much Improvement Do We Need to Have a Better System? *Proceedings of LREC 2004*. Lisbon, Portugal.