Making Texts in Electronic Health Records Comprehensible to Consumers: A Prototype Translator

Qing Zeng-Treitler PHD[1], Sergey Goryachev MS[1], Hyeoneui Kim RN PHD[1], Alla Keselman PHD[2], Douglas Rosendale MD[3]
[1]Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts [2]National Library of Medicine, National Institute of Health, Bethesda, Maryland [3]Veteran Affairs Hospital, Washington DC

Abstract

Narrative reports from electronic health records are a major source of content for personal health records. We designed and implemented a prototype text translator to make these reports more comprehensible to consumers. The translator identifies difficult terms, replaces them with easier synonyms, and generates and inserts explanatory texts for them. In feasibility testing, the application was used to translate 9 clinical reports. Majority (68.8%) of text replacements and insertions were deemed correct and helpful by expert review. User evaluation demonstrated a non-statistically significant trend toward better comprehension when translation is provided (p=0.15).

Introduction

Personal health records (PHRs) have the potential to empower health care consumers and improve self-care [1]. In order to fulfill this promise, PHRs need to be easily understood by lay readers. Much of today's PHR content comes from electronic health records (EHR), made up largely of physician progress notes, discharge summaries, and procedure reports (e.g., from radiology, pathology, or surgery). Understanding EHR information can be difficult for an average consumer. Even a highly educated consumer is likely to find many radiology reports and discharge summary statements incomprehensible, such as "Calcified opacity is noted in the left femoral head and left ilium" and "H/O abnormal pap with laser therapy for ablation." Our previous work suggests that professional medical terminology and abbreviations significantly impede patients' comprehension of their medical records [2].

Efforts to improve consumer-friendliness of PHR and EHR information focused on user interface design and/or the links to references or educational materials (i.e., infobuttons) [3-5]. Less attention has been given to the underlying logic and linguistic features of PHRs. This study employs text translation/simplification as a method for improving the readability of EHR texts for consumers.

Few monolingual automatic text translation tools presently exist in or outside of the medical domain. Our literature search produced three pilot natural language processing (NLP) systems for syntactic simplification [6-8]. All of them were used to simplify general newspaper articles. It is not obvious, however, how to apply their techniques to clinical reports, as there are significant syntactic differences between medical and non-medical corpora [8]. None of the systems attempted lexical simplification, i.e. simplifying the vocabulary. Since vocabulary is a key factor in health text readability [2, 9], we focused on term replacement and explanation generation as a first step toward making texts in EHR comprehensible to consumers.

Background

Two sources of vocabulary knowledge were used by the study: the Unified Medical Langrage System (UMLS)[1], and the open-access collaborative (OAC) consumer health vocabulary (CHV)[2]. UMLS is a comprehensive source of medical terms and concepts as well as concept semantic types and relations. The OAC CHV, on the other hand, provides consumer health specific information, which complements the UMLS.

Most of the OAC CHV concepts have one-to-one match with UMLS concepts. OAC is much smaller than the UMLS, containing only 58,000 concepts. Each OAC concept has a consumer-friendly display name, which is often different from the UMLS preferred name of the same concept. These names were identified through a combination of automated analysis and manual review [10].

Each term in the OAC CHV also has several familiarity scores. Familiarity scores estimate the likelihood that a term or concept will be recognized by an average consumer [11]. The score ranges from 0 to 1, with 1 being the most familiar and least difficult. OAC offers a frequency-based, a context-based, and a combination familiarity score for terms.

[1] http://www.nlm.nih.gov/research/umls/
[2] www.consumerhealthvocab.org

The frequency-based score was calculated by a support-vector machine model based on term-occurrence frequency in several health text corpora. The context-based score was calculated based on term co-occurrence patterns in a health-specific query log data set. The combination score is derived from the frequency-based and context-based scores; in most cases, it is the average of the two. The correlation between frequency-based scores and actual consumer comprehension has been validated statistically in two small-scale user studies [12]. In this study, the combination score was used.

We chose not to use UMLS or other dictionary definitions as explanatory information for direct insertion into the text. Although they provide excellent external resources to link to, these definitions tend to be long and sometimes even more difficult than the terms they define. For instance, one UMLS source defines "heart valve" as "Flaps of tissue that prevent regurgitation of blood from the heart ventricles to the heart atria or from the pulmonary arteries or aorta to the ventricles."

Design

We employ two strategies to mitigate the vocabulary difficulty of clinical reports: (a) synonym replacement, and (b) explanation insertion.

Synonym replacement. Medical concepts often have multiple names, one of which may be easier for consumers to understand than others. Replacing a term with its more comprehensible synonym may improve readability without loss of semantic information. For example, "Pharyngitis" can be safely replaced with "sore throat".

Explanation insertion. While not all terms have easy synonyms, many are connected to other simpler terms through hierarchical or non-hierarchical relations. According to ontological theory [13], all concepts can be eventually defined using a set of basic concepts and relations.

Hierarchical relations are particularly useful: A term can be explained as a specific incidence of its ancestors. A term's decedents could also be provided as examples of the concept. For instance, "Pulmonary emboli" is a type of "lung disease"; While "Chicken" is an example of "Poultry".

For terms of some semantic types, certain non-hierarchical relations are especially useful for explanatory purposes. A body part, for instance, may be explained as part of another body part (e.g. "right atrium" is part of a "heart").

Implementation

Our implemented prototype text translator (Figure 1), has three main components: (a) concept extraction, (b) synonym identification and replacement, and (c) explanation generation and insertion.

Concept extraction. An existing NLP system called HITEx [14] is used to parse EHR reports and map report terms to UMLS concepts. This step is necessary because we identify synonyms and related terms for a term based on the concept it represents. To avoid errors that might be introduced by disambiguation, this prototype does not attempt to replace and explain ambiguous terms (i.e. terms that map to multiple concepts).

Synonym identification. As described in the Background, many UMLS concepts have one-to-one match with OAC CHV concepts, and all OAC concepts have pre-defined consumer friendly display names. This enables the translator to look up the OAC consumer friendly display name for a UMLS concept and use it to replace the original term.

Some abbreviations (e.g., "WBC") will be replaced with their more understandable full names (e.g., "white blood cell count"), but the consumer-friendly ones such as "AIDS" will not be replaced by their less friendly full names such as "acquired immuno deficiency syndrome", because "AIDS" and "white blood cell count" are the consumer friendly display names for their respective underlying concepts.

Explanation generation. If a term or its replacement has a familiarity score below a threshold (which can be adjusted by users), the application generates explanatory phrases based on the semantic relations in UMLS. The application searches two levels of hierarchical relationships (i.e. parents, grandparents, children and grandchildren) for ancestors or descendents with more comprehensible names.

The only non-hierarchical relations used in this prototype are "has site of" for concepts with the UMLS semantic type of "Disease or Syndrome", and "is part of" for concepts with the UMLS semantic type of "Body part, Organ or Organ Component".

If a more comprehensible related term is found using hierarchical or non-hierarchical relations, an explanatory phrase will be generated describing the relations between the original and the related term. Otherwise, a term will be explained as an instance of its UMLS semantic type.

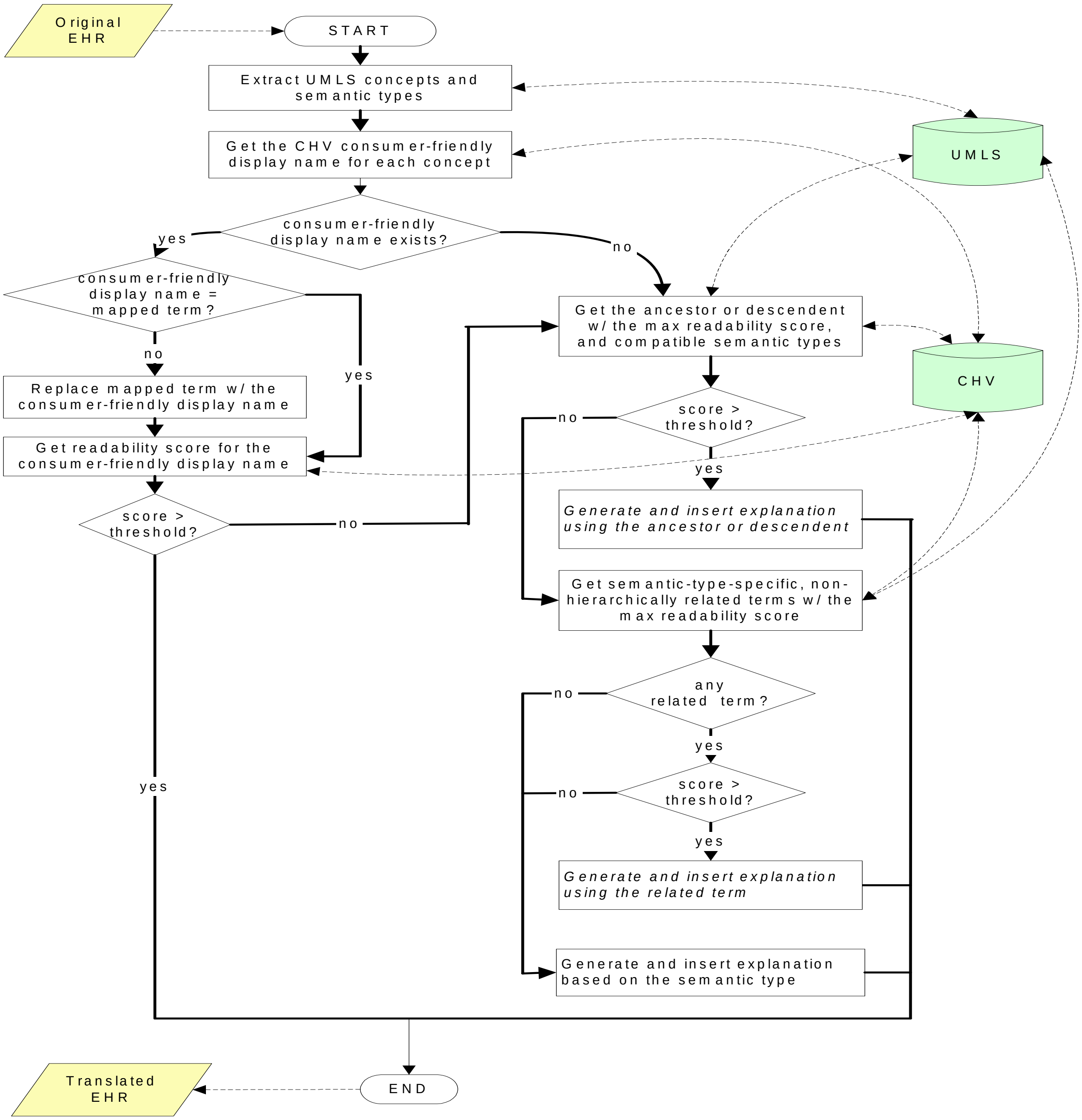

Figure 1 Schematic diagram of the translation process.

Feasibility testing

We collected nine anonymized sample EHR reports including a admission note, five discharge summaries, a radiology and two surgery reports from the Web, and translated them using the prototype translator. The threshold of minimal readability score was set to 0.6. Table 1 presents three pairs of original and translated sentences from the reports.

The first 250 words of each report were extracted for human review and cloze testing (a standard comprehension test procedure) [15]. A clinician reviewed the translations for correctness and helpfulness.

For cloze testing, incorrect translations identified by the clinician were removed. Following standard cloze procedure, every 5th word in the first 250 words of each document was replaced with a blank space.

A convenience sample of nine subjects was recruited from within the Decision Systems Group and the National Library of Medicine. The subjects are not clinicians but highly educated (1 at college and 8 at graduate school level). Each subject was given 2 different reports: one original and one translated. They were asked to fill in the blank spaces.

A cloze score for each document was calculated as the percentage of answers that matched with the deleted words exactly. We then compared the average cloze scores of the original and translated reports.

Table 1 Examples of sentence translation. The replaced or inserted texts are highlighted.

He denies radiation of pain, nausea, vomiting, diaphoresis, palpitations, syncope or near syncope.
He denies radiating pain, nausea, vomiting, excessive sweating, palpitation, fainting, or near fainting.
Laboratory data on 1/10 showed glucose level test 94.
Laboratory data on 1/10 showed glucose level test (e.g. blood sugar level) 94.
The patient is to be discharged home to continue Rifampin orally.
The patient is to be discharged home to continue Rifampin (a type of antibiotic) orally.

Results

On average, 14.7 terms were translated in the first 250 words of the reports. Majority (68.8%) of the translations were deemed correct and helpful by the human reviewer, 23.0% was deemed unhelpful, and 8.2% incorrect. Table 2 provides examples of translations which were considered correct and helpful, correct but unhelpful, and incorrect.

Table 2 Examples of translations which were considered helpful, unhelpful and incorrect.

	Original	Translation
Helpful	tachypnea	rapid breathing
Unhelpful	nulligravida	nulligravida (a type of finding)
Incorrect	cyst	cyst (a type of tumor)

The incorrect translations were mainly caused by problems in two areas:

- Term to concept mapping. For example, the word "ascend" was mapped to a veterinarian medication called "Ascend".
- Hierarchical relations. For example, "Tobacco abuse" is a child of "Psychiatric problem".

The term mappings and semantic relations were obtained from the UMLS, and some of them were not applicable to the context of the reports or consumer-oriented translation. To explain "cyst" as "a type of tumor" or "tobacco abuse" as "a type of psychiatric problem", for example, may falsely alarm or unnecessarily alienate the reader.

The average cloze score was 36.6% for the original reports and 43.0% for the translated. Since a higher cloze score signifies better comprehension, the data suggest a trend toward improved comprehension when reading the translated reports. This trend, however, is not statistically significant (p=0.15).

For a document to be considered readable for an audience, the cloze score of the document should be in the 50-60% range. Cloze scores of the original and translated EHR reports ranged from 23.8% to 54.4% (Figure 2), indicating that the reports are fairly difficult for the subjects to comprehend.

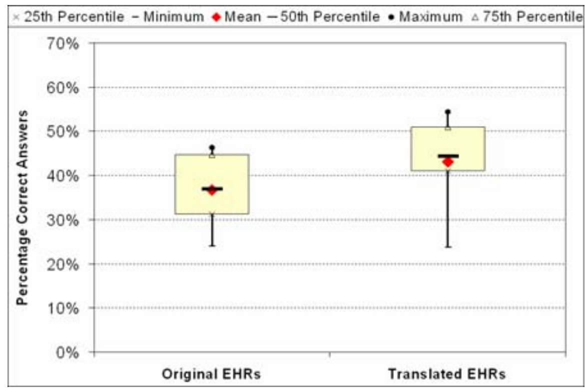

Figure 2 Box-blocker graph of cloze score distributions for original and translated EHR reports.

Discussion

The need to improve the readability of EHRs for lay readers is growing, as more patients gain access to their records through PHRs. We have designed and implemented a prototype translator to improve the readability of EHR reports for consumers. While there had been a few prior studies on automated text simplification, they were conducted with general newspaper articles which are more readable than EHRs. These studies also focused on syntactic transformation, and did not address vocabulary-related issues.

This study focused on reducing vocabulary difficulty. A prior study reported that vocabulary and "main point" are the two key text features used by health communication experts to assess readability [9]. While Infobuttons [5] have been used to provide vocabulary/knowledge support, we explored a different approach – text translation through synonym replacement and explanation generation. Infobuttons are very helpful, however, clicking the infobuttons and following the links to external resources does interrupt the flow of reading. The optimal approach may be to combine text translation with infobutton.

Machine translation is inherently challenging, be it multilingual or monolingual. In feasibility testing, 68% of the translations were deemed to be correct and helpful, which is very encouraging. The 8.2%

incorrect rate, however, is troubling. Most of the errors could be corrected through a more selective use of the UMLS and a careful examination of the mapping of some relatively common English words. Some of the problematic term mappings were already modified in the latest (2007AA) UMLS, compared with the 2005AA version we used. Nevertheless, EHR translation should not introduce errors and it would require considerable amount of effort for us to eliminate all the errors.

As a first prototype, our EHR text translator needs much improvement. Besides reducing translation errors, we intend to explore syntax translation, and conduct more extensive user testing.

Another observation we made is that EHR records are indeed very difficult for lay people to comprehend, and there is a non-statistically significant trend toward improved comprehension when translation is provided. Materials that score 50-60% in cloze tests are considered to be fairly readable. While most of the participants of our cloze test are much more educated than the average consumers, the average cloze score of the original reports was only 36.6%. The average cloze score of the translated reports was higher (43.0%), though no statistical significance was found partially due to the small sample size and high variability in the report sample.

Conclusion

We have developed and implemented a prototype EHR text translator and obtained promising results from feasibility testing. While such a translator has the potential to improve consumer comprehension, we also recognize the importance and challenge of eliminating translation errors.

Acknowledgments

This work is supported by the NIH grant R01 LM07222 and by the Intramural Research Program of the NIH, NLM/Lister Hill National Center for Biomedical Communications.

References

1. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-6.
2. Keselman. A, Slaughter L, Arnott-Smith C, Kim H, Divita G, Browne A, Zeng-Treitler Q. Towards Consumer-Friendly PHRs: Patients' Experience with Reviewing Their Health Records. 2007 (Unpublished manuscript).
3. De Clercq PA, Hasman A, Wolffenbuttel BH. A consumer health record for supporting the patient-centered management of chronic diseases. Med Inform Internet Med 2003;28(2):117-27.
4. Wald JS, Middleton B, Bloom A, Walmsley D, Gleason M, Nelson E, Li Q, Epstein M, Volk L, Bates DW. A patient-controlled journal for an electronic medical record: issues and challenges. Medinfo 2004;11(Pt 2):1166-70.
5. Baorto DM, Cimino JJ. An "infobutton" for enabling patients to interpret on-line Pap smear reports. Proc AMIA Symp 2000:47-50.
6. Carroll J, Minnen G, Pearce D, Canning Y, Devlin S, Tait J. Simplifying text for language-impaired readers. In: EACL 99: European Chapter of the Association for Computational Linguistics; 1999; 1999.
7. Chandrasekar R, Bangalore S. Automatic induction of rules for text simplification. Knowledge-Based Systems 1997;10(3):183-190.
8. Campbell DA, Johnson SB. Comparing syntactic complexity in medical and non-medical corpora. Proc AMIA Symp 2001:90-4.
9. Rosemblat G, Logan R, Tse T, Graham L. How Do Text Features Affect Readability? Expert Evaluations on Consumer Health Web Site Text. In: MEDNET; 2006; Toronto, CA; 2006. p. In press.
10. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying Consumer-Friendly Display (CFD) Names for Health Concepts. Proc AMIA Symp 2005:859-63.
11. Zeng Q, Kim E, Crowell J, Tse T. A Text Corpora-Based Estimation of the Familiarity of Health Terminology. Lecture Notes in Computer Science 2005;3745 / 2005:184-192.
12. Keselman A, Massengalea L, Ngo L, Browne A, Zeng Q. The Effect of User Factors on Consumer Familiarity with Health Terms: Using Gender as a Proxy for Background Knowledge about Gender-Specific Illnesses. In: International Symposium on Biological and Medical Data Analysis; 2006; Thessaloniki, Greece; 2006. p. In press.
13. Sowa J. Knowledge Representation:Logical, Philosophical, and Computational Foundations: Brookes/Cole; 2000.
14. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006;6:30.
15. Taylor WL. Cloze procedure: A new tool for measuring readability. Journalism Quarterly 1953(30):415-433.