MedSLT: A Limited-Domain Unidirectional Grammar-Based Medical Speech Translator

Rayner, Emmanuel; Bouillon, Pierrette; Chatzichrisafis, Nikolaos; Santaholma, Marianne Elina; Starlander, Marianne; Hockey, Beth Ann; Nakao, Yukie; Isahara, Hitoshi; Kanzaki, Kyoko

How to cite

RAYNER, Emmanuel et al. MedSLT: A Limited-Domain Unidirectional Grammar-Based Medical Speech Translator. In: Proceedings of the First International Workshop on Medical Speech Translation (HLT-NAACL). ACL - Association for Computational Linguistics (Ed.). New York (USA). [s.l.] : [s.n.], 2006. p. 44–47.

This publication URL: https://archive-ouverte.unige.ch/unige:3440

MedSLT: A Limited-Domain Unidirectional Grammar-Based Medical Speech Translator

Manny Rayner, Pierrette Bouillon, Nikos Chatzichrisafis, Marianne Santaholma, Marianne Starlander
University of Geneva, TIM/ISSCO
40 bvd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland
mrayner@riacs.edu
Pierrette.Bouillon@issco.unige.ch, Nikos.Chatzichrisafis@vozZup.com
Marianne.Santaholma@eti.unige.ch, Marianne.Starlander@eti.unige.ch

Beth Ann Hockey
UCSC/NASA Ames Research Center
Moffet Field, CA 94035
bahockey@email.arc.nasa.gov

Yukie Nakao, Hitoshi Isahara, Kyoko Kanzaki
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan 619-0289
yukie-n@khn.nict.go.jp, {isahara, kanzaki}@nict.go.jp

Abstract

MedSLT is a unidirectional medical speech translation system intended for use in doctor-patient diagnosis dialogues, which provides coverage of several different language pairs and subdomains. Vocabulary ranges from about 350 to 1000 surface words, depending on the language and subdomain. We will demo both the system itself and the development environment, which uses a combination of rule-based and data-driven methods to construct efficient recognisers, generators and transfer rule sets from small corpora.

1 Overview

The mainstream in speech translation work is for the moment statistical, but rule-based systems are still a very respectable alternative. In particular, nearly all systems which have actually been deployed are rule-based. Prominent examples are (Phraselator, 2006; S-MINDS, 2006; MedBridge, 2006).

MedSLT (MedSLT, 2005; Bouillon et al., 2005) is a unidirectional medical speech translation system for use in doctor-patient diagnosis dialogues, which covers several different language pairs and subdomains. Recognition is performed using grammar-based language models, and translation uses a rule-based interlingual framework. The system, including the development environment, is built on top of Regulus (Regulus, 2006), an Open Source platform for developing grammar-based speech applications, which in turn sits on top of the Nuance Toolkit.

The demo will show how MedSLT can be used to carry out non-trivial diagnostic dialogues. In particular, we will demonstrate how an integrated intelligent help system counteracts the brittleness inherent in rule-based processing, and rapidly leads new users towards the supported system coverage. We will also demo the development environment, and show how grammars and sets of transfer rules can be efficiently constructed from small corpora of a few hundred to a thousand examples.

2 The MedSLT system

The MedSLT demonstrator has already been extensively described elsewhere (Bouillon et al., 2005; Rayner et al., 2005a), so this section will only present a brief summary. The main components are a set of speech recognisers for the source languages, a set of generators for the target languages, a translation engine, sets of rules for translating to and from interlingua, a simple discourse engine for dealing with context-dependent translation, and a top-level which manages the information flow between the other modules and the user.

MedSLT also includes an intelligent help module, which adds robustness to the system and guides the user towards the supported coverage. The help module uses a backup recogniser, equipped with a statistical language model, and matches the results from this second recogniser against a corpus of utterances which are within system coverage and translate correctly. In previous studies (Rayner et al., 2005a; Starlander et al., 2005), we showed that the grammar-based recogniser performs much better than the statistical one on in-coverage utterances, but worse on out-of-coverage ones. Having the help system available approximately doubled the speed at which subjects learned, measured as the average difference in semantic error rate between the results for their first quarter-session and their last quarter-session. It is also possible to recover from recognition errors by selecting a displayed help sentence; in the cited studies, this increased the number of acceptably processed utterances by about 10%.

We will demo several versions of the system, using different source languages, target languages and subdomains. Coverage is based on standard examination questions obtained from doctors, and consists mainly of yes/no questions, though there is also support for WH-questions and elliptical utterances. Table 1 gives examples of the coverage in the English-input headache version, and Table 2 summarises recognition performance in this domain for the three main input languages. Differences in the sizes of the recognition vocabularies are primarily due to differences in use of inflection. Japanese, with little inflectional morphology, has the smallest vocabulary; French, which inflects most parts of speech, has the largest.

Where?
"is the pain above your eye" "do you experience the pain in your jaw" "does the pain spread to the shoulder"
When?
"have you had the pain for more than a month" "do the headaches ever occur in the morning"
How long?
"does the pain typically last a few minutes" "does the pain ever last more than two hours"
How often?
"do you get headaches several times a week" "are the headaches occurring more often"
How?
"is it a stabbing pain" "is the pain usually severe"
Associated symptoms?
"do you vomit when you get the headaches" "is the pain accompanied by blurred vision"
Why?
"does bright light make the pain worse" "do you get headaches when you eat cheese"
What helps?
"does sleep make the pain better" "does massage help"
Background?
"do you have a history of sinus disease" "have you had an e c g"

Table 1: Examples of English MedSLT coverage

3 The development environment

Although the MedSLT system is rule-based, we would, for the usual reasons, prefer to acquire these rules from corpora using some well-defined method. There is, however, little or no material available for most medical speech translation domains, including ours. As noted in (Probst and Levin, 2002), scarcity of data generally implies use of some strategy to obtain a carefully structured training corpus. If the corpus is not organised in this way, conflicts between alternate learned rules occur, and it is hard to induce a stable set of rules. As Probst and Levin suggest, one obvious way to attack the problem is to implement a (formal or informal) elicitation strategy, which biases the informant towards translations which are consistent with the existing ones. This is the approach we have adopted in MedSLT.

The Regulus platform, on which MedSLT is based, supports rapid construction of complex grammar-based language models; it uses an example-based method driven by small corpora of disambiguated parsed examples (Rayner et al., 2003; Rayner et al., 2006), which extracts most of the structure of the model from a general linguistically motivated resource grammar. The result is a specialised version of the general grammar, tailored to the example corpus, which can then be com-

Language	Vocab	WER	SemER
English	441	6%	18%
French	1025	8%	10%
Japanese	347	4%	4%

Table 2: Recognition performance for English, French and Japanese headache-domain recognisers. “Vocab” = number of surface words in source language recogniser vocabulary; “WER” = Word Error Rate for source language recogniser, on in-coverage material; “SemER” = semantic error rate for source language recogniser, on in-coverage material.

piled into an efficient recogniser or into a generation module. Regulus-based recognisers and generators are easy to maintain, and grammar structure is shared automatically across different subdomains. Resource grammars are available for several languages, including English, Japanese, French and Spanish.

Nuance recognisers derived from the resource grammars produce both a recognition string and a semantic representation. This representation consists of a list of key/value pairs, optionally including one level of nesting; the format of interlingua and target language representations is similar. The formalism is sufficiently expressive that a reasonable range of temporal and causal constructions can be represented (Rayner et al., 2005b). A typical example is shown in Figure 1. A translation rule maps a list of key/value pairs to a list of key/value pairs, optionally specifying conditions requiring that other key/value pairs either be present or absent in the source representation.

When developing new coverage for a given language pair, the developer has two main tasks. First, they need to add new training examples to the corpora used to derive the specialised grammars used for the source and target languages; second, they must add translation rules to handle the new key/value pairs. The simple structure of the MedSLT representations makes it easy to support semi-automatic acquisition of both of these types of information. The basic principle is to attempt to find the minimal set of new rules that can be added to the existing set, in order to cover the new corpus example; this is done through a short elicitation dialogue with the developer. We illustrate this with a simple example.

Suppose we are developing coverage for the English → Spanish version of the system, and that the English corpus sentence “does the pain occur at night” fails to translate. The acquisition tool first notes that processing fails when converting from interlingua to Spanish. The interlingua representation is

```
[[utterance_type,ynq],
 [pronoun,you],
 [state,have_symptom],
 [symptom,pain],[tense,present],
 [prep,in_time],[time,night]]
```

Applying Interlingua → Spanish rules, the result is

```
[[utterance_type,ynq],
 [pronoun,usted],
 [state,tener],[symptom,dolor],
 [tense,present],
 [prep,por_temporal],
 failed:[time,night]]
```

where the tag `failed` indicates that the element `[time,night]` could not be processed. The tool matches the incomplete transferred representation against a set of correctly translated examples, and shows the developer the English and Spanish strings for the three most similar ones, here

```
does it appear in the morning
-> tiene el dolor por la mañana

does the pain appear in the morning
-> tiene el dolor por la mañana

does the pain come in the morning
-> tiene el dolor por la mañana
```

This suggests that a translation for “does the pain occur at night” consistent with the existing rules would be “tiene el dolor por la noche”. The developer gives this example to the system, which parses it using both the general Spanish resource grammar and the specialised grammar used for generation in the headache domain. The specialised grammar fails to produce an analysis, while the resource grammar produces two analyses,

```
[[utterance_type,ynq],
 [pronoun,usted],
```

```
[[utterance_type,ynq],[pronoun,you],[state,have_symptom],
 [tense,present],[symptom,headache],[sc,when],
 [[clause,[[utterance_type,dcl],[pronoun,you],
  [action,drink],[tense,present],[cause,coffee]]]]]
```

Figure 1: Representation of “do you get headaches when you drink coffee”

```
 [state,tener],[symptom,dolor],
 [tense,present],
 [prep,por_temporal],
 [temporal,noche]]
```

and

```
[[utterance_type,dcl],
 [pronoun,usted],
 [state,tener],[symptom,dolor],
 [tense,present],
 [prep,por_temporal],
 [temporal,noche]]
```

The first of these corresponds to the YN-question reading of the sentence (“do you have the pain at night”), while the second is the declarative reading (“you have the pain at night”). Since the first (YN-question) reading matches the Interlingua representation better, the acquisition tool assumes that it is the intended one. It can now suggest two pieces of information to extend the system’s coverage.

First, it adds the YN-question reading of “tiene el dolor por la noche” to the corpus used to train the specialised generation grammar. The piece of information acquired from this example is that `[temporal,noche]` should be realised in this domain as “la noche”. Second, it compares the correct Spanish representation with the incomplete one produced by the current set of rules, and induces a new Interlingua to Spanish translation rule. This will be of the form

```
[time,night] -> [temporal,noche]
```

In the demo, we will show how the development environment makes it possible to quickly add new coverage to the system, while also checking that old coverage is not broken.

References

P. Bouillon, M. Rayner, N. Chatzichrisafis, B.A. Hockey, M. Santaholma, M. Starlander, Y. Nakao, K. Kanzaki, and H. Isahara. 2005. A generic multi-lingual open source platform for limited-domain medical speech translation. In *In Proceedings of the 10th Conference of the European Association for Machine Translation (EAMT)*, Budapest, Hungary.

MedBridge, 2006. http://www.medtablet.com/index.html. As of 15 March 2006.

MedSLT, 2005. http://sourceforge.net/projects/medslt/. As of 15 March 2005.

Phraselator, 2006. http://www.phraselator.com. As of 15 March 2006.

K. Probst and L. Levin. 2002. Challenges in automatic elicitation of a controlled bilingual corpus. In *Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation*.

M. Rayner, B.A. Hockey, and J. Dowding. 2003. An open source environment for compiling typed unification grammars into speech recognisers. In *Proceedings of the 10th EACL (demo track)*, Budapest, Hungary.

M. Rayner, P. Bouillon, N. Chatzichrisafis, B.A. Hockey, M. Santaholma, M. Starlander, H. Isahara, K. Kankazi, and Y. Nakao. 2005a. A methodology for comparing grammar-based and robust approaches to speech understanding. In *Proceedings of the 9th International Conference on Spoken Language Processing (ICSLP)*, Lisboa, Portugal.

M. Rayner, P. Bouillon, M. Santaholma, and Y. Nakao. 2005b. Representational and architectural issues in a limited-domain medical speech translator. In *Proceedings of TALN/RECITAL*, Dourdan, France.

M. Rayner, B.A. Hockey, and P. Bouillon. 2006. *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler*. CSLI Press, Chicago.

Regulus, 2006. http://sourceforge.net/projects/regulus/. As of 15 March 2006.

S-MINDS, 2006. http://www.sehda.com. As of 15 March 2006.

M. Starlander, P. Bouillon, N. Chatzichrisafis, M. Santaholma, M. Rayner, B.A. Hockey, H. Isahara, K. Kanzaki, and Y. Nakao. 2005. Practicing controlled language through a help system integrated into the medical speech translation system (MedSLT). In *Proceedings of the MT Summit X*, Phuket, Thailand.