A Limited-Domain English to Japanese Medical Speech Translator Built Using REGULUS 2

Manny Rayner
Research Institute for Advanced Computer Science (RIACS),
NASA Ames Research Center,
Moffet Field, CA 94035
mrayner@riacs.edu

Pierrette Bouillon
University of Geneva
TIM/ISSCO,
40, bvd du Pont-d'Arve,
CH-1211 Geneva 4,
Switzerland
pierrette.bouillon@issco.unige.ch

Vol Van Dalsem III
El Camino Hospital
2500 Grant Road
Mountain View, CA 94040
vvandal3@aol.com

Hitoshi Isahara, Kyoko Kanzaki
Communications Research Laboratory
3-5 Hikaridai
Seika-cho, Soraku-gun
Kyoto, Japan 619-0289
{isahara,kanzaki}@crl.go.jp

Beth Ann Hockey
Research Institute for Advanced Computer Science (RIACS),
NASA Ames Research Center,
Moffet Field, CA 94035
bahockey@riacs.edu

Abstract

We argue that verbal patient diagnosis is a promising application for limited-domain speech translation, and describe an architecture designed for this type of task which represents a compromise between principled linguistics-based processing on the one hand and efficient phrasal translation on the other. We propose to demonstrate a prototype system instantiating this architecture, which has been built on top of the Open Source REGULUS 2 platform. The prototype translates spoken yes-no questions about headache symptoms from English to Japanese, using a vocabulary of about 200 words.

1 Introduction and motivation

Language is crucial to medical diagnosis. During the initial evaluation of a patient in an emergency department, obtaining an accurate history of the chief complaint is of equal importance to the physical examination. In many parts of the world there are large recent immigrant populations that require medical care but are unable to communicate fluently in the local language. In the US these immigrants are especially likely to use emergency facilities because of insurance issues. In an emergency setting there is acute need for quick accurate physician-patient communication but this communication is made substantially more difficult in cases where there is a language barrier. Our system is designed to address this problem using spoken machine translation.

Designing a spoken translation system to obtain a detailed medical history would be difficult if not impossible using the current state of the art. The reason that the use of spoken translation technology is feasible is because what is actually needed in the emergency setting is more limited. Since medical histories traditionally are obtained through two-way physician-patient conversations that are mostly physician initiative, there is a preestablished limiting structure that we can follow in designing the translation system. This structure allows a physician to sucessfully use one way translation to elicit and restrict the range of patient responses while still obtaining the necessary information.

Another helpful constraint on the conversational requirements is that the majority of medical conditions can be initiatlly characterized by a relatively small number of key questions about quality, quantity and duration of symptoms. For example, key questions about chest pain include intensity, location, duration, quality of pain, and factors that increase or decrease the pain. These answers to these questions can be sucessfully communicated by a limited number of one or two word responses (e.g. yes/no, left/right, numbers) or even gestures (e.g.

pointing to an area of the body). This is clearly a domain in which the constraints of the task are sufficient for a limited domain, one way spoken translation system to be a useful tool.

2 An architecture for limited-domain speech translation

The basic philosophy behind the architecture of the system is to attempt an intelligent compromise between fixed-phrase translation on one hand (e.g. (IntegratedWaveTechnologies, 2002)) and linguistically motivated grammar-based processing on the other (e.g. VERBMOBIL (Wahlster, 2000) and Spoken Language Translator (Rayner et al., 2000a)). At run-time, the system behaves essentially like a phrasal translator which allows some variation in the input language. This is close in spirit to the approach used in most normal phrase-books, which typically allow "slots" in at least some phrases ("How much does — cost?"; "How do I get to — ?"). However, in order to minimize the overhead associated with defining and maintaining large sets of phrasal patterns, these patterns are derived from a single large linguistically motivated unification grammar; thus the *compile-time* architecture is that of a linguistically motivated system. Phrasal translation at runtime gives us speed and reliability; the linguistically motivated compile-time architecture makes the system easy to extend and modify.

The runtime system comprises three main modules. These are respectively responsible for source language speech recognition, including parsing and production of semantic representation; transfer and generation; and synthesis of target language speech. The speech processing modules (recognition and synthesis) are implemented on top of the standard Nuance Toolkit platform (Nuance, 2003). Recognition is constrained by a CFG language model written in Nuance Grammar Specification Language (GSL), which also specifies the semantic representations produced. This language model is compiled from a linguistically motivated unification grammar using the Open Source REGULUS 2 platform (Rayner et al., 2003; Regulus, 2003); the compilation process is driven by a small corpus of examples. The language processing modules (transfer and generation) are a suite of simple routines written in SICStus Prolog. The speech and language processing modules communicate with each other through a minimal file-based protocol.

The semantic representations on both the source and target sides are expressed as attribute-value structures. In accordance with the generally minimalistic design philosophy of the project, semantic representations have been kept as simple as possible. The basic principle is that the representation of a clause is a flat list of attribute-value pairs: thus for example the representation of "Did your headache start suddenly?" is the attribute-value list

```
[[utterance_type,ynq],[tense,past],
 [symptom,headache],[state,start],
 [manner,suddenly]]
```

In a broad domain, it is of course trivial to construct examples where this kind of representation runs into serious problems. In the very narrow domain of a phrasebook translator, it has many desirable properties. In particular, operations on semantic representations typically manipulate lists rather than trees. In a broad domain, we would pay a heavy price: the lack of structure in the semantic representations would often make them ambiguous. The very simple ontology of the phrasebook domain however means that ambiguity is not a problem; the components of a flat list representation can never be derived from more than one functional structure, so this structure does not need to be explicitly present.

Transfer rules define mappings of sets of attribute-value pairs to sets of attribute-value pairs; the majority of the rules map single attribute-value pairs to single attribute-value pairs. Generation is handled by a small Definite Clause Grammar (DCG), which converts attribute-value structures into surface strings; its output is passed through a minimal post-transfer component, which applies a set of rules which map fixed strings to fixed strings. Speech synthesis is performed either by the Nuance Vocalizer TTS engine or by concatenation of recorded wavefiles, depending on the output language.

One of the most important questions for a medical translation system is that of reliability; we address this issue using the methods of (Rayner and Bouillon, 2002). The GSL form of the recognition grammar is run in generation mode using the Nuance `generate` utility to generate large numbers of random utterances, all of which are by construction within system coverage. These utterances are then processed through the system in batch mode using all-solutions versions of the relevant processing algorithms. The results are checked automatically to find examples where rules are either deficient or ambiguous. With domains of the complexity under consideration here, we have found that it is feasible to refine the rule-sets in this way so that holes and ambiguities are effectively eliminated.

3 A medical speech translation system

We have built a prototype medical speech translation system instantiating the functionality outlined in Section 1 and the architecture of Section 2. The system permits spoken English input of constrained yes/no questions about the symptoms of headaches, using a vocabulary of about 200 words. This is enough to support most of the standard examination questions for this subdomain. There are two versions of the system, producing spoken output in French and Japanese respectively. Since English → Japanese is distinctly the more interesting and challenging language pair, we will focus on this version.

Speech recognition and source language analysis are performed using REGULUS 2. The grammar is specialised from the large domain-independent grammar using the methods sketched in Section 2. The training corpus has been constructed by hand from an initial corpus supplied by a medical professional; the content of the questions was kept unchanged, but where necessary the form was revised to make it more appropriate to a spoken dialogue. When we felt that it would be difficult to remember what the canonical form of a question would be, we added two or three variant forms. For example, we permit “Does bright light make the headache worse?” as a variant for “Is the headache aggravated by bright light?”, and “Do you usually have headaches in the morning?” as a variant for “Does the headache usually occur in the morning?”. The current training corpus contains about 200 examples.

The granularity of the phrasal rules learned by grammar specialisation has been set so that the constituents in the acquired rules are VBARs, postmodifier groups, NPs and lexical items. VBARs may include both inverted subject NPs and adverbs[1]. Thus for example the training example “Are the headaches usually caused by emotional upset?” induces a top-level rule whose context-free skeleton is

```
UTT --> VBAR, VBAR, POSTMODS
```

For the training example, the first VBAR in the induced rule spans the phrase “are the headaches usually”, the second VBAR spans the phrase “caused”, and the POSTMODS span the phrase “by emotional upset”. The same rule could potentially be used to cover utterances like “Is the pain sometimes preceded by nausea?” and “Is your headache ever associated with blurred vision?”. The same training example will also induce several lower-level rules, the least trivial of which are rules for VBAR and POSTMODS with context-free skeletons

```
VBAR --> are, NP, ADV
POSTMODS --> P, NP
```

The grammar specialisation method is described in full detail in (Rayner et al., 2000b).

With regard to the transfer component, we have had two main problems to solve. Firstly, it is well-known that translation from English to Japanese requires major reorganisation of the syntactic form. Word-order is nearly always completely different, and category mismatches are very common. It is mainly for this reason that we chose to use a flat semantic representation. As long as the domain is simple enough that the flat representations are unambiguous, transfer can be carried out by mapping lists of elements into lists of elements. For example, we translate “are your headaches caused by fatigue” as “tsukare de zutsu ga okorimasu ka” (lit. “fatigue-CAUSAL headache-SUBJ occur-PRESENT QUESTION”). Here, the source-language representation is

```
[[utterance_type,ynq],
 [tense,present],
 [symptom,headache],
 [event,cause],
 [cause,fatigue]]
```

and the target-language one is

```
[[utterance_type,sentence],
 [tense,present],
 [symptom,zutsu],
```

[1]This non-standard definition of VBAR has technical advantages discussed in (Rayner et al., 2000c)

do your headaches often appear at night → yoku yoru ni zutsu ga arimasu ka (often night-AT headache-SUBJ is-PRES-Q)
is the pain in the front of the head → itami wa atama no mae no hou desu ka (pain-TOPIC head-OF front side is-PRES-Q)
did your headache start suddenly → zutsu wa totsuzen hajimari mashita ka (headache-TOPIC sudden start-PRES-Q)
have you had headaches for weeks → sushukan zutsu ga tsuzuite imasu ka (weeks headache-SUBJ have-CONT-PRES-Q)
is the pain usually superficial → itsumo itami wa hyomenteki desu ka (usually pain-SUBJ superficial is-PRES-Q)
is the severity of the headaches increasing → zutsu wa hidoku natte imasu ka (headache-TOPIC severe becoming is-PRES-Q)

Table 1: Examples of utterances covered by the prototype

```
[event,okoru],[postpos,causal],
[cause,tsukare]]
```

Each line in the source representation maps into the corresponding one in the target in the obvious way. The target-language grammar is constrained enough that there is only one Japanese sentence which can be generated from the given representation.

The second major problem for transfer relates to elliptical utterances. These are very important due to the one-way character of the interaction: instead of being able to ask a WH-question ("What does the pain feel like?"), the doctor needs to ask a series of Y-N questions ("Is the pain dull?", "Is the pain burning?", "Is the pain aching?", etc). We rapidly found that it was much more natural for questions after the first one to be phrased elliptically ("Is the pain dull?", "Burning?", "Aching?"). English and Japanese have however different conventions as to what types of phrase can be used elliptically. Here, for example, it is only possible to allow some types of Japanese adjectives to stand alone. Thus we can grammatically and semantically say "hageshii desu ka" (lit. "burning is-QUESTION") but not "*uzukuyona desu ka" (lit. "*aching is-QUESTION"). The problem is that adjectives like "uzukuyona" must combine adnominally with a noun in this context: thus we in fact have to generate "uzukuyona itami desu ka" ("aching-ADNOMINAL-USAGE pain is-QUESTION"). Once again, however, the very limited domain makes it practical to solve the problem robustly. There are only a handful of transformations to be implemented, and the extra information that needs to be added is always clear from the sortal types of the semantic elements in the target representation.

Table 1 gives examples of utterances covered by the system, and the translations produced.

References

IntegratedWaveTechnologies, 2002. http://www.i-w-t.com/investor.html. As of 15 Mar 2002.

Nuance, 2003. http://www.nuance.com. As of 25 February 2003.

M. Rayner and P. Bouillon. 2002. A phrasebook style medical speech translator. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (demo track)*, Philadelphia, PA.

M. Rayner, D. Carter, P. Bouillon, V. Digalakis, and M. Wirén, editors. 2000a. *The Spoken Language Translator*. Cambridge University Press.

M. Rayner, D. Carter, and C. Samuelsson. 2000b. Grammar specialisation. In Rayner et al. (Rayner et al., 2000a).

M. Rayner, B.A. Hockey, and F. James. 2000c. Compiling language models from a linguistically motivated unification grammar. In *Proceedings of the Eighteenth International Conference on Computational Linguistics*, Saarbrucken, Germany.

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.

Regulus, 2003. http://sourceforge.net/projects/regulus/. As of 24 April 2003.

W. Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer.