Japanese Speech Understanding Using Grammar Specialization

Manny Rayner, Nikos Chatzichrisafis, Pierrette Bouillon
University of Geneva, TIM/ISSCO
40 bvd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland
mrayner@riacs.edu
{Pierrette.Bouillon,Nikolaos.Chatzichrisafis}@issco.unige.ch

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

Beth Ann Hockey
UCSC/NASA Ames Research Center
Moffet Field, CA 94035
bahockey@riacs.edu

Marianne Santaholma, Marianne Starlander
University of Geneva, TIM/ISSCO
40 bvd du Pont-d'Arve
CH-1211 Geneva 4, Switzerland
Marianne.Santaholma@eti.unige.ch
Marianne.Starlander@eti.unige.ch

The most common speech understanding architecture for spoken dialogue systems is a combination of speech recognition based on a class N-gram language model, and robust parsing. For many types of applications, however, grammar-based recognition can offer concrete advantages. Training a good class N-gram language model requires substantial quantities of corpus data, which is generally not available at the start of a new project. Head-to-head comparisons of class N-gram/robust and grammar-based systems also suggest that users who are familiar with system coverage get better results from grammar-based architectures (Knight et al., 2001). As a consequence, deployed spoken dialogue systems for real-world applications frequently use grammar-based methods. This is particularly the case for speech translation systems. Although leading research systems like Verbmobil and NESPOLE! (Wahlster, 2000; Lavie et al., 2001) usually employ complex architectures combining statistical and rule-based methods, successful practical examples like Phraselator and S-MINDS (Phraselator, 2005; Sehda, 2005) are typically phrasal translators with grammar-based recognizers.

Voice recognition platforms like the Nuance Toolkit provide CFG-based languages for writing grammar-based language models (GLMs), but it is challenging to develop and maintain grammars consisting of large sets of ad hoc phrase-structure rules. For this reason, there has been considerable interest in developing systems that permit language models be specified in higher-level formalisms, normally some kind of unification grammar (UG), and then compile these grammars down to the low-level platform formalisms. A prominent early example of this approach is the Gemini system (Moore, 1998).

Gemini raises the level of abstraction significantly, but still assumes that the grammars will be domain-dependent. In the Open Source REGULUS project (Regulus, 2005; Rayner et al., 2003), we have taken a further step in the direction of increased abstraction, and derive all recognizers from a single linguistically motivated UG. This derivation procedure starts with a large, application-independent UG for a language. An application-specific UG is then derived using an Explanation Based Learning (EBL) specialization technique. This corpus-based specialization process is parameterized by the training corpus and operationality criteria. The training corpus, which can be relatively small, consists of examples of utterances that should be recognized by the target application. The sentences of the corpus are parsed using the general grammar, then those parses are partitioned into phrases based on the operationality criteria. Each phrase defined by the operationality criteria is flattened, producing rules of a phrasal grammar for the application domain. This application-specific UG is then compiled into

Proceedings of HLT/EMNLP 2005 Demonstration Abstracts, pages 26–27,
Vancouver, October 2005.

a CFG, formatted to be compatible with the Nuance recognition platform. The CFG is compiled into the runtime recognizer using Nuance tools.

Previously, the REGULUS grammar specialization programme has only been implemented for English. In this demo, we will show how we can apply the same methodology to Japanese. Japanese is structurally a very different language from English, so it is by no means obvious that methods which work for English will be applicable in this new context: in fact, they appear to work very well. We will demo the grammars and resulting recognizers in the context of Japanese → English and Japanese → French versions of the Open Source MedSLT medical speech translation system (Bouillon et al., 2005; MedSLT, 2005).

The generic problem to be solved when building any sort of recognition grammar is that syntax alone is insufficiently constraining; many of the real constraints in a given domain and use situation tend to be semantic and pragmatic in nature. The challenge is thus to include enough non-syntactic constraints in the grammar to create a language model that can support reliable domain-specific speech recognition: we sketch our solution for Japanese.

The basic structure of our current general Japanese grammar is as follows. There are four main groups of rules, covering NP, PP, VP and CLAUSE structure respectively. The NP and PP rules each assign a sortal type to the head constituent, based on the domain-specific sortal constraints defined in the lexicon. VP rules define the complement structure of each syntactic class of verb, again making use of the sortal features. There are also rules that allow a VP to combine with optional adjuncts, and rules which allow null constituents, in particular null subjects and objects. Finally, clause-level rules form a clause out of a VP, an optional subject and optional adjuncts. The sortal features constrain the subject and the complements combining with a verb, but the lack of constraints on null constituents and optional adjuncts still means that the grammar is very loose. The grammar specialization mechanism flattens the grammar into a set of much simpler structures, eliminating the VP level and only permitting specific patterns of null constituents and adjuncts licenced by the training corpus.

We will demo several different versions of the Japanese-input medical speech translation system, differing with respect to the target language and the recognition architecture used. In particular, we will show a) that versions based on the specialized Japanese grammar offer fast and accurate recognition on utterances within the intended coverage of the system (Word Error Rate around 5%, speed under $0.1\times$RT), b) that versions based on the original general Japanese grammar are much less accurate and more than an order of magnitude slower.

References

P. Bouillon, M. Rayner, N. Chatzichrisafi s, 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.

S. Knight, G. Gorrell, M. Rayner, D. Milward, R. Koeling, and I. Lewin. 2001. Comparing grammar-based and robust approaches to speech understanding: a case study. In *Proceedings of Eurospeech 2001*, pages 1779–1782, Aalborg, Denmark.

A. Lavie, C. Langley, A. Waibel, F. Pianesi, G. Lazzari, P. Coletti, L. Taddei, and F. Balducci. 2001. Architecture and design considerations in NESPOLE!: a speech translation system for e-commerce applications. In *Proceedings of HLT: Human Language Technology Conference*, San Diego, California.

MedSLT, 2005. http://sourceforge.net/projects/medslt/. As of 9 June 2005.

R. Moore. 1998. Using natural language knowledge sources in speech recognition. In *Proceedings of the NATO Advanced Studies Institute*.

Phraselator, 2005. http://www.phraselator.com/. As of 9 June 2005.

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

Regulus, 2005. http://sourceforge.net/projects/regulus/. As of 9 June 2005.

Sehda, 2005. http://www.sehda.com/. As of 9 June 2005.

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