A Bidirectional Grammar-Based Medical Speech Translator

Pierrette Bouillon[1], Glenn Flores[2], Marianne Starlander[1], Nikos Chatzichrisafis[1]
Marianne Santaholma[1], Nikos Tsourakis[1], Manny Rayner[1,3], Beth Ann Hockey[4]

[1] University of Geneva, TIM/ISSCO, 40 bvd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland
Pierrette.Bouillon@issco.unige.ch
Marianne.Starlander@eti.unige.ch, Nikos.Chatzichrisafis@vozZup.com
Marianne.Santaholma@eti.unige.ch, Nikolaos.Tsourakis@issco.unige.ch

[2] Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226
gflores@mcw.edu

[3] Powerset, Inc., 475 Brannan Street, San Francisco, CA 94107
manny@powerset.com

[4] Mail Stop 19-26, UCSC UARC, NASA Ames Research Center, Moffett Field, CA 94035–1000
bahockey@ucsc.edu

Abstract

We describe a bidirectional version of the grammar-based MedSLT medical speech system. The system supports simple medical examination dialogues about throat pain between an English-speaking physician and a Spanish-speaking patient. The physician's side of the dialogue is assumed to consist mostly of WH-questions, and the patient's of elliptical answers. The paper focusses on the grammar-based speech processing architecture, the ellipsis resolution mechanism, and the online help system.

1 Background

There is an urgent need for medical speech translation systems. The world's current population of 6.6 billion speaks more than 6,000 languages (Graddol, 2004). Language barriers are associated with a wide variety of deleterious consequences in healthcare, including impaired health status, a lower likelihood of having a regular physician, lower rates of mammograms, pap smears, and other preventive services, non-adherence with medications, a greater likelihood of a diagnosis of more severe psychopathology and leaving the hospital against medical advice among psychiatric patients, a lower likelihood of being given a follow-up appointment after an emergency department visit, an increased risk of intubation among children with asthma, a greater risk of hospital admissions among adults, an increased risk of drug complications, longer medical visits, higher resource utilization for diagnostic testing, lower patient satisfaction, impaired patient understanding of diagnoses, medications, and follow-up, and medical errors and injuries (Flores, 2005; Flores, 2006). Nevertheless, many patients who need medical interpreters do not get them. For example, in the United States, where 52 million people speak a language other than English at home and 23 million people have limited English proficiency (LEP) (Census, 2007), one study found that about half of LEP patients presenting to an emergency department were not provided with a medical interpreter (Baker et al., 1996). There is thus a substantial gap between the need for and availability of language services in health care, a gap that could be bridged through effective medical speech translation systems.

An ideal system would be able to interpret accurately and flexibly between patients and health care professionals, using unrestricted language and a large vocabulary. A system of this kind is, unfortunately, beyond the current state of the art. It is, however, possible, using today's technology, to build speech translation systems for specific scenarios and language-pairs, which can achieve acceptable levels of reliability within the bounds

Proceedings of SPEECHGRAM 2007, pages 41–48,
Prague, Czech Republic, June 2007.

of a well-defined controlled language. MedSLT (Bouillon et al., 2005) is an Open Source system of this type, which has been under construction at Geneva University since 2003. The system is built on top of Regulus (Rayner et al., 2006), an Open Source platform which supports development of grammar-based speech-enabled applications. Regulus has also been used to build several other systems, including NASA's Clarissa (Rayner et al., 2005b).

The most common architecture for speech translation today uses statistical methods to perform both speech recognition and translation, so it is worth clarifying why we have chosen to use grammar-based methods. Even though statistical architectures exhibit many desirable properties (purely data-driven, domain independent), this is not necessarily the best alternative in safety-critical medical applications. Anecdotally, many physicians express reluctance to trust a translation device whose output is not readily predictable, and most of the speech translation systems which have reached the stage of field testing rely on various types of grammar-based recognition and rule-based translation (Phraselator, 2007; Fluential, 2007).

Statistical speech recognisers can achieve impressive levels of accuracy when trained on enough data, but it is a daunting task to collect training material in the requisite quantities (usually, tens of thousands of high-quality utterances) when trying to build practical systems. Considering that the medical speech translation applications we are interested in constructing here need to work for multiple languages and subdomains, the problem becomes even more challenging. Our experience is that grammar-based systems which also incorporate probabilistic context-free grammar tuning deliver better results than purely statistical ones when training data are sparse (Rayner et al., 2005a).

Another common criticism of grammar-based systems is that out-of-coverage utterances will neither be recognized nor translated, an objection that critics have sometimes painted as decisive. It is by no means obvious, however, that restricted coverage is such a serious problem. In text processing, work on several generations of controlled language systems has developed a range of techniques for keeping users within the bounds of system coverage (Kittredge, 2003; Mitamura, 1999), and variants of these methods can also be adapted for spoken language applications. Our experiments with MedSLT show that even a quite simple help system is enough to guide users quickly towards the intended coverage of a medium-vocabulary grammar-based speech translation application, with most users appearing confident after just an hour or two of exposure (Starlander et al., 2005; Chatzichrisafis et al., 2006).

Until recently, the MedSLT system only supported unidirectional processing in the physician to patient direction. The assumption was that the physician would mostly ask yes/no questions, to which the patient would respond non-verbally, for example by nodding or shaking their head. A unidirectional architecture is easier to make habitable than a bidirectional one. It is reasonable to assume that the physician will use the system regularly enough to learn the coverage, but most patients will not have used the system before, and it is less clear that they will be able to acclimatize within the narrow window at their disposal. These considerations must however be balanced against the fact that a unidirectional system does not allow for a patient-centered interaction characterized by meaningful patient-clinician communication or shared decision making. Multiple studies in the medical literature document that patient-centeredness, effective patient-clinician communication, and shared decision making are associated with significant improvements in patient health outcomes, including reduced anxiety levels, improved functional status, reduced pain, better control of diabetes mellitus, blood pressure reduction among hypertensives, improved adherence, increased patient satisfaction, and symptom reduction for a variety of conditions (Stewart, 1995; Michie et al., 2003). A bidirectional system is considered close to essential from a healthcare perspective, since it appropriately addresses the key issues of patient centeredness and shared decision making. For these reasons, we have over the last few months developed a bidirectional version of MedSLT, initially focussing on a throat pain scenario with an English-speaking physician and a Spanish-speaking patient. The physician uses full sentences, while the patient answers with short responses.

One of the strengths of the Regulus approach is that it is very easy to construct parallel versions of a grammar; generally, all that is required is to vary the training corpus. (We will have more to say about this soon). We have exploited these properties of the platform to create two different configurations of the bidirectional system, so that we can compare competing approaches to the problem of accommodating patients unfamiliar with speech technology. In Version 1 (less restricted), the patient is allowed to answer using both elliptical utterances and short sentences, while in Version 2 (more restricted) they are only permitted to use elliptical utterances. Thus, for example, if the physician asks the question "How long have you had a sore throat?", Version 1 allows the patient to respond both "Desde algunos días" ("For several days") and "Me ha dolido la garganta desde algunos días" ("I have had a sore throat for several days"), while Version 2 only allows the first of these. Both the short and the long versions are translated uniformly, with the short version resolved using the context from the preceding question.

In both versions, if the patient finds it too challenging to use the system to answer WH-questions directly, it is possible to back off to the earlier dialogue architecture in which the physician uses Y-N questions and the patient responds with simple yes/no answers, or nonverbally. Continuing the example, if the patient is unable to find an appropriate way to answer the physician's question, the physician could ask "Have you had a sore throat for more than three days?"; if the patient responds negatively, they could continue with the follow-on question "More than a week?", and so on.

In the rest of the paper, we first describe the system top-level (Section 2), the way in which grammar-based processing is used (Section 3), the ellipsis processing mechanism (Section 4), and the help system (Section 5). Section 6 presents an initial evaluation, and the final section concludes.

2 Top-level architecture

The system is operated through the graphical user interface (GUI) shown in Figures 1 and 2. In accordance with the basic principles of patient-centeredness and shared decision-making outlined in Section 1, the patient and the physician each have their own headset, use their own mouse, and share the same view of the screen. This is in sharp contrast to the majority of the medical speech translation systems described in the literature (Somers, 2006).

As shown in the screenshots, the main GUI window is separated into two tabbed panes, marked "Doctor" and "Patient". Initially, the "Doctor" view (the one shown in Figure 1) is active. The physician presses the "Push to talk" button, and speaks into the headset microphone. If recognition is successful, the GUI displays four separate results, listed on the right side of the screen. At the top, immediately under the heading "Question", we can see the actual words returned by speech recognition. Here, these words are "Have you had rapid strep test". Below, we have the help pane: this displays similar questions taken from the help corpus, which are known to be within system coverage. The pane marked "System understood" shows a back-translation, produced by first translating the recognition result into interlingua, and then translating it back into English. In the present example, this corrects the minor mistake the recogniser has made, missing the indefinite article "a", and confirms that the system has obtained a correct grammatical analysis and interpretation at the level of interlingua. At the bottom, we see the target language translation. The left-hand side of the screen logs the history of the conversation to date, so that both sides can refer back to it.

If the physician decides that the system has correctly understood what they said, they can now press the "Play" button. This results in the system producing a spoken output, using the Vocalizer TTS engine. Simultaneously with speaking, the GUI shifts to the "Patient" configuration shown in Figure 2. This differs from the "Doctor" configuration in two respects: all text is in the patient language, and the help pane presents its suggestions immediately, based on the preceding physician question. The various processing components used to support these functionalities are described in the following sections.

3 Grammar-based processing

Grammar-based processing is used for source-language speech recognition and target-side generation. (Source-language analysis is part of the recognition process, since grammar-based recognition includes creating a parse). All of these functionalities

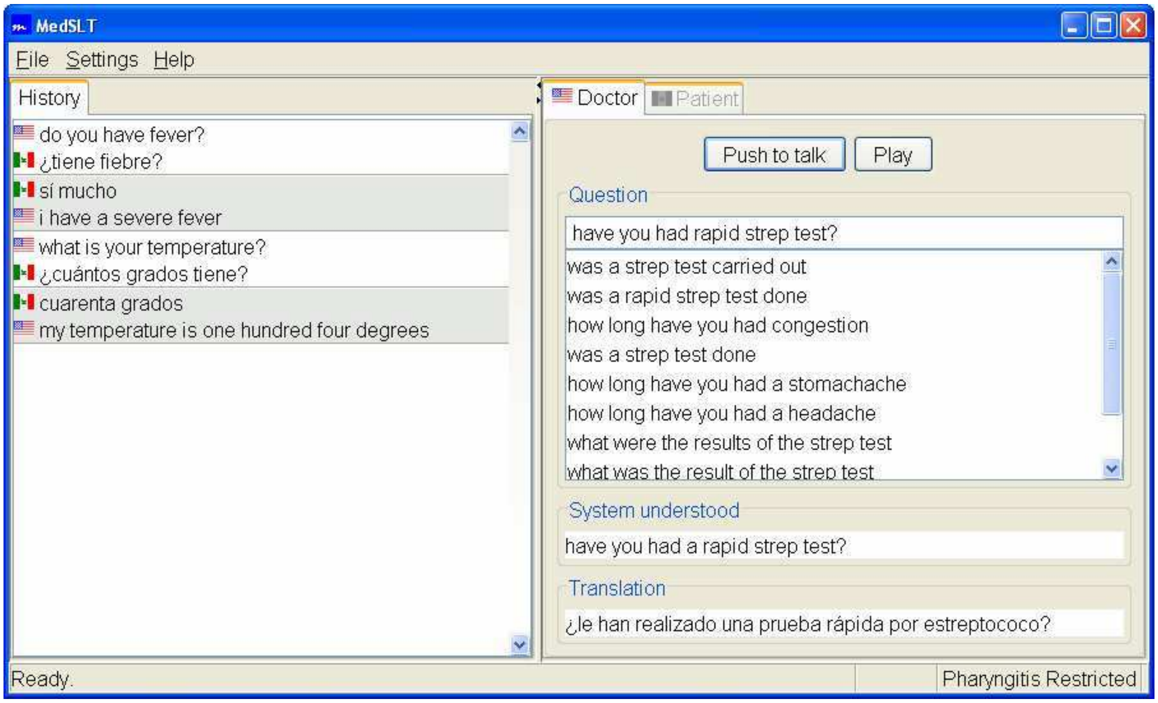

Figure 1: Screenshot showing the state of the GUI after the physician has spoken, but before he has pressed the "Play" button. The help pane shows similar queries known to be within coverage.

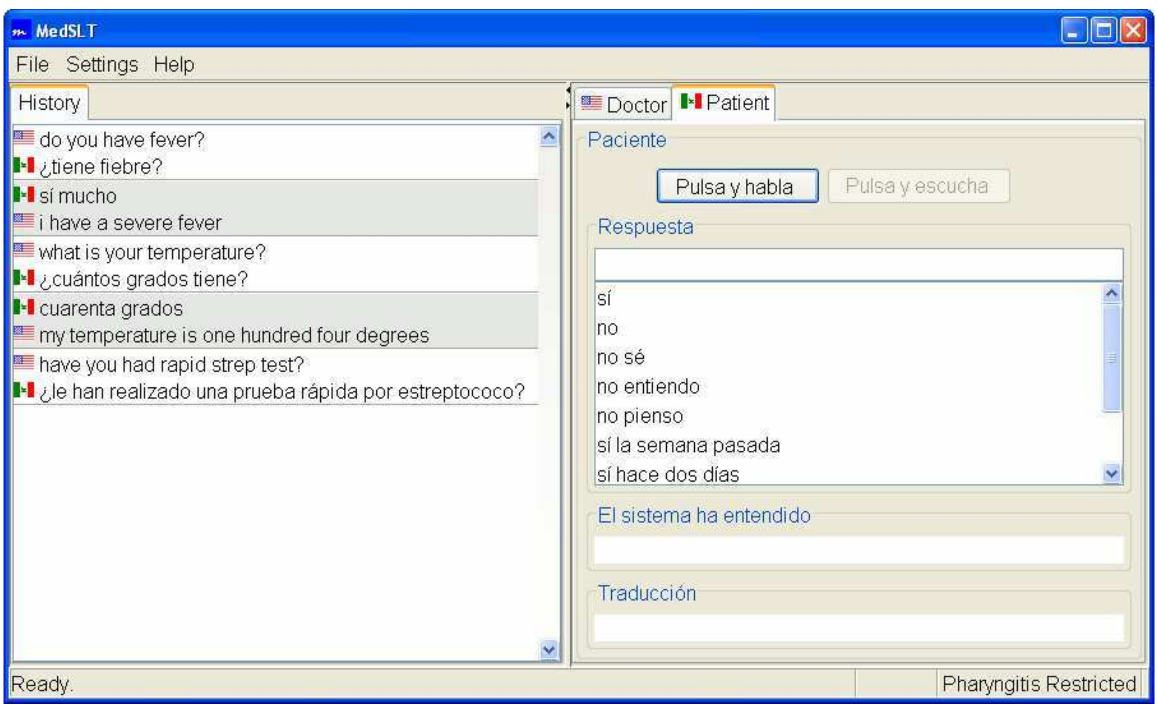

Figure 2: Screenshot showing the state of the GUI after the physician has pressed the "Play" button. The help pane shows known valid responses to similar questions.

are implemented using the Regulus platform, with the task-specific grammars compiled out of general feature grammar resources by the Regulus tools. For both recognition and generation, the first step is to extract a domain-specific feature grammar from the general one, using a version of the Explanation Based Learning (EBL) algorithm.

The extraction process is driven by a corpus of examples and a set of "operationality criteria", which define how the rules in the original resource grammar are recombined into domain-specific ones. It is important to realise that the domain-specific grammar is *not* merely a subset of the resource grammar; a typical domain-specific grammar rule is created by merging two to five resource grammar rules into a single "flatter" rule. The result is a feature grammar which is less general than the original one, but more efficient. For recognition, the grammar is then processed further into a CFG language model, using an algorithm which alternates expansion of feature values and filtering of the partially expanded grammar to remove irrelevant rules. Detailed descriptions of the EBL learning and feature grammar $\rightarrow$ CFG compilation algorithms can be found in Chapters 8 and 10 of (Rayner et al., 2006). Regulus feature grammars can also be compiled into generators using a version of the Semantic Head Driven algorithm (Shieber et al., 1990).

The English (physician) side recogniser is compiled from the large English resource grammar described in Chapter 9 of (Rayner et al., 2006), and was constructed in the same way as the one described in (Rayner et al., 2005a), which was used for a headache examination task. The operationality criteria are the same, and the only changes are a different training corpus and the addition of new entries to the lexicon. The same resources, with a different training corpus, were used to build the English language generator. It is worth pointing out that, although a uniform method was used to build these various grammars, the results were all very different. For example, the recognition grammar from (Rayner et al., 2005a) is specialised to cover only second-person questions ("Do you get headaches in the mornings?"), while the generator grammar used in the present application covers only first-person declarative statements ("I visited the doctor last Monday."). In terms of structure, each grammar contains several important constructions that the other lacks. For example, subordinate clauses are central in the headache domain ("Do the headaches occur when you are stressed?") but are not present in the sore throat domain; this is because the standard headache examination questions mostly focus on generic conditions, while the sore throat examination questions only relate to concrete ones. Conversely, relative clauses are important in the sore throat domain ("I have recently been in contact with someone who has strep throat"), but are not sufficiently important in the headache domain to be covered there.

On the Spanish (patient) side, there are four grammars involved. For recognition, we have two different grammars, corresponding to the two versions of the system; the grammar for Version 2 is essentially a subset of that for Version 1. For generation, there are two separate and quite different grammars: one is used for translating the physician's questions, while the other produces back-translations of the patient's questions. All of these grammars are extracted from a general shared resource grammar for Romance languages, which currently combines rules for French, Spanish and Catalan (Bouillon et al., 2006; Bouillon et al., to appear 2007b).

One interesting consequence of our methodology is related to the fact that Spanish is a prodrop language, which implies that many sentences are systematically ambiguous between declarative and Y-N question readings. For example, "He consultado un médico" could in principle mean either "I visited a doctor" or "Did I visit a doctor?". When training the specialised Spanish grammars, it is thus necessary to specify which readings of the training sentences are to be used. Continuing the example, if the sentence occurred in training material for the answer grammar, we would specify that the declarative reading was the intended one[1].

4 Ellipsis processing and contextual interpretation

In Version 1 of the system, the patient is permitted to answer using elliptical phrases; in Ver-

[1]The specification can be formulated as a preference that applies uniformly to all the training examples in a given group.

sion 2, she is obliged to do so. Ability to process elliptical responses makes it easier to guide the patient towards the intended coverage of the system, without degrading the quality of recognition (Bouillon et al., to appear 2007a). The downside is that ellipses are also harder to translate than full sentences. Even in a limited domain like ours, and in a closely related language-pair, ellipsis can generally not be translated word for word, and it is necessary to look at the preceding context if the rules are to be applied correctly. In examples 1 and 2 below, the locative phrase "In your stomach" in the English source becomes the subject in the Spanish translation. This implies that the translation of the ellipsis in the second physician utterance needs to change syntactic category: "In your head" (PP) becomes "La cabeza" (NP).

(1)	Doctor:	Do you have a pain in your stomach?
	(Trans):	Le duele el estomago?
(2)	Doctor:	In your head?
	(Trans):	*En la cabeza?

Since examples like this are frequent, our system implements a solution in which the patient's replies are translated in the context of the preceding utterance. If the patient-side recogniser's output is classified as an ellipsis (this can done fairly reliably thanks to use of suitably specialised grammars; cf. Section 3), we expand the incomplete phrase into a full sentence structure by adding appropriate structural elements from the preceding physician-side question; the expanded semantic structure is the one which is then translated into interlingual form, and thence back to the physician-side language.

Since all linguistic representations, including those of elliptical phrases and their contexts, are represented as flat attribute-value lists, we are able to implement the resolution algorithm very simply in terms of list manipulation. In YN-questions, where the elliptical answer intuitively adds information to the question ("Did you visit the doctor?"; "El lunes" → "I visited the doctor on Monday"), the representations are organised so that resolution mainly amounts to concatenation of the two lists[2]. In WH-questions, where the answer intuitively substitutes the elliptical answer for the WH-phrase ("What is your temperature?"; "Cuarenta grados" → "My temperature is forty degrees"), resolution substitutes the representation of the elliptical phrase for that of a semantically similar element in the question.

[2]It is also necessary to replace second-person pronouns with first-person counterparts.

The least trivial aspect of this process is providing a suitable definition of "semantically similar". This is done using a simple example-based method, in which the grammar developer writes a set of declarations, each of which lists a set of semantically similar NPs. At compile-time, the grammar is used to parse each NP, and extract a generalised skeleton, in which specific lexical information is stripped away; at run-time, two NPs are held to be semantically similar if they can each be unified with skeletons in the same equivalence class. This ensures that the definition of the semantic similarity relation is stable across most changes to the grammar and lexicon. The issues are described in greater detail in (Bouillon et al., to appear 2007a).

5 Help system

Since the performance of grammar-based speech understanding is only reliable on in-coverage material, systems based on this type of architecture must necessarily use a controlled language approach, in which it is assumed that the user is able to learn the relevant coverage. As previously noted, the MedSLT system addresses this problem by incorporating an online help system (Starlander et al., 2005; Chatzichrisafis et al., 2006).

On the physician side, the help system offers, after each recognition event, a list of related questions; similarly, on the patient side, it provides examples of known valid answers to the current question. In both cases, the help examples are extracted from a precompiled corpus of question-answer pairs, which have been judged for correctness by system developers. The process of selecting the examples is slightly different on the two sides. For questions (physician side), the system performs a second parallel recognition of the input speech, using a statistical recogniser. It then compares the recognition result, using an N-gram based metric, against the set of known correct in-coverage questions from the question-answer corpus, to extract the most similar ones. For answers (patient side), the help system searches the question-answer corpus to find the questions most similar to the current one, and shows the list of corresponding valid answers, using the whole list in the case of Version 1 of the system, and only the subset consisting of elliptical phrases in the case of Version 2.

6 Evaluation

In previous studies, we have evaluated speech recognition and speech understanding performance for physician-side questions in English (Bouillon et al., 2005) and Spanish (Bouillon et al., to appear 2007b), and investigated the impact on performance of the help system (Rayner et al., 2005a; Starlander et al., 2005). We have also carried out recent evaluations designed to contrast recognition performance on elliptical and full versions of the same utterance; here, our results suggest that elliptical forms of (French-language) MedSLT utterances are slightly easier to recognise in terms of semantic error rate than full sentential forms (Bouillon et al., to appear 2007a). Our initial evaluation studies on the bidirectional system have focussed on a specific question which has particular relevance to this new version of MedSLT. Since we are assuming that the patient will respond using elliptical utterances, and that these utterances will be translated in the context of the preceding physician-side question, how confident can we be that this context-dependent translation will be correct?

In order to investigate these issues, we performed a small data-collection using Version 2 of the system, whose results we summarise here. One of the authors of the paper played the role of an English-speaking physician, in a simulated medical examination scenario where the goal was to determine whether or not the "patient" was suffering from a viral throat infection. The six subjects playing the role of the patient were all native speakers of Spanish, and had had no previous exposure to the system, or indeed any kind of speech technology. They were given cards describing the symptoms they were supposed to be displaying, on which they were asked to based their answers. From a total of 92 correctly recognised patient responses, we obtained 50 yes/no answers and 42 examples of real elliptical utterances. Out of these, 36 were judged to have been translated completely correctly, and a further 3 were judged correct in terms of meaning, but less than fluent. Only 3 examples were badly translated: of these two were caused by problems in a translation rule, and one by incorrect treatment of ellipsis resolution. We show representative exchanges below; the last of these is the one in which ellipsis processing failed to work correctly.

(3)	Doctor:	For how long have you had your sore throat?
	Patient:	Desde hace más de una semana
	(Trans):	I have had a sore throat for more than one week
(4)	Doctor:	What were the results?
	Patient:	Negativo
	(Trans):	The results were negative
(5)	Doctor:	Have you seen a doctor for your sore throat?
	Patient:	Sí el lunes
	(Trans):	I visited the doctor for my sore throat monday
(6)	Doctor:	Have you been with anyone recently who has a strep throat?
	Patient:	Si más de dos semanas
	(Trans):	I was in contact with someone more than two weeks recently who had strep throat

7 Conclusions

We have presented a bidirectional grammar-based English $\leftrightarrow$ Spanish medical speech translation system built using a linguistically motivated architecture, where all linguistic information is ultimately derived from two resource grammars, one for each language. We have shown how this enables us to derive the multiple grammars needed, which differ both with respect to function (recognition/generation) and to domain (physician questions/patient answers). The system is currently undergoing initial lab testing; we hope to advance to initial trials on real patients some time towards the end of the year.

References

[Baker et al.1996] D.W. Baker, R.M. Parker, M.V. Williams, W.C. Coates, and Kathryn Pitkin. 1996.

Use and effectiveness of interpreters in an emergency department. *Journal of the American Medical Association*, 275:783–8.

[Bouillon et al.2005] 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 *Proceedings of the 10th Conference of the European Association for Machine Translation (EAMT)*, pages 50–58, Budapest, Hungary.

[Bouillon et al.2006] P. Bouillon, M. Rayner, B. Novellas Vall, Y. Nakao, M. Santaholma, M. Starlander, and N. Chatzichrisafis. 2006. Une grammaire multilingue partagée pour la traduction automatique de la parole. In *Proceedings of TALN 2006*, Leuwen, Belgium.

[Bouillon et al.to appear 2007a] P. Bouillon, M. Rayner, M. Santaholma, and M. Starlander. to appear 2007a. Les ellipses dans un système de traduction automatique de la parole. In *Proceedings of TALN 2006*, Toulouse, France.

[Bouillon et al.to appear 2007b] P. Bouillon, M. Rayner, B. Novellas Vall, Y. Nakao, M. Santaholma, M. Starlander, and N. Chatzichrisafis. to appear 2007b. Une grammaire partagée multi-tâche pour le traitement de la parole : application aux langues romanes. *Traitement Automatique des Langues*.

[Census2007] U.S. Census, 2007. *Selected Social Characteristics in the United States: 2005. Data Set: 2005 American Community Survey*. Available here.

[Chatzichrisafis et al.2006] N. Chatzichrisafis, P. Bouillon, M. Rayner, M. Santaholma, M. Starlander, and B.A. Hockey. 2006. Evaluating task performance for a unidirectional controlled language medical speech translation system. In *Proceedings of the HLT-NAACL International Workshop on Medical Speech Translation*, pages 9–16, New York.

[Flores2005] G. Flores. 2005. The impact of medical interpreter services on the quality of health care: A systematic review. *Medical Care Research and Review*, 62:255–299.

[Flores2006] G. Flores. 2006. Language barriers to health care in the united states. *New England Journal of Medicine*, 355:229–231.

[Fluential2007] Fluential, 2007. http://www.fluentialinc.com. As of 24 March 2007.

[Graddol2004] D. Graddol. 2004. The future of language. *Science*, 303:1329–1331.

[Kittredge2003] R. I. Kittredge. 2003. Sublanguages and comtrolled languages. In R. Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, pages 430–447. Oxford University Press.

[Michie et al.2003] S. Michie, J. Miles, and J. Weinman. 2003. Patient-centeredness in chronic illness: what is it and does it matter? *Patient Education and Counseling*, 51:197–206.

[Mitamura1999] T. Mitamura. 1999. Controlled language for multilingual machine translation. In *Proceedings of Machine Translation Summit VII*, Singapore.

[Phraselator2007] Phraselator, 2007. http://www.voxtec.com/. As of 24 March 2007.

[Rayner et al.2005a] M. Rayner, P. Bouillon, N. Chatzichrisafis, B.A. Hockey, M. Santaholma, M. Starlander, H. Isahara, K. Kanzaki, 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)*, pages 1103–1107, Lisboa, Portugal.

[Rayner et al.2005b] M. Rayner, B.A. Hockey, J.M. Renders, N. Chatzichrisafis, and K. Farrell. 2005b. A voice enabled procedure browser for the International Space Station. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (interactive poster and demo track)*, Ann Arbor, MI.

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

[Shieber et al.1990] S. Shieber, G. van Noord, F.C.N. Pereira, and R.C. Moore. 1990. Semantic-head-driven generation. *Computational Linguistics*, 16(1).

[Somers2006] H. Somers. 2006. Language engineering and the path to healthcare: a user-oriented view. In *Proceedings of the HLT-NAACL International Workshop on Medical Speech Translation*, pages 32–39, New York.

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

[Stewart1995] M.A. Stewart. 1995. Effective physician-patient communication and health outcomes: a review. *Canadian Medical Association Journal*, 152:1423–1433.