Language Engineering and the Pathway to Healthcare: A user-oriented view

Harold Somers
School of Informatics
University of Manchester
PO Box 88
Manchester M61 0QD, England
Harold.Somers@manchester.ac.uk

Abstract

This position paper looks critically at a number of aspects of current research into spoken language translation (SLT) in the medical domain. We first discuss the user profile for medical SLT, criticizing designs which assume that the doctor will necessarily need or want to control the technology. If patients are to be users on an equal standing, more attention must be paid to usability issues. We focus briefly on the issue of feedback in SLT systems, pointing out the difficulties of relying on text-based paraphrases. We consider the delicate issue of evaluating medical SLT systems, noting that some of the standard and much-used evaluation techniques for all aspects of the SLT chain might not be suitable for use with real users, even if they are role-playing. Finally, we discuss the idea that the "pathway to healthcare" involves much more than a face-to-face interview with a medical professional, and that different technologies including but not restricted to SLT will be appropriate along this pathway.

1 Introduction

The doctor–patient consultation is a central element of the "pathway to healthcare", and with language problems recognised as the single most significant barrier on this pathway, spoken-language translation (SLT) of doctor–patient dialogues is an obvious and timely and attractive application of language technology. As Bouillon et al. (2005) state, the task is both useful and manageable, particularly as interactions are highly constrained, and the domain can be divided into smaller domains based on symptom types. In this position paper, we wish to discuss a number of aspects of this research area, and suggest that we should broaden our horizons to look beyond the central doctor–patient consultation to consider the variety of interactions on the pathway to healthcare, and beyond the confines of SLT as an appropriate technology for patient–provider communication.

In particular we want to stress the importance of the users – both practitioners and patients – in the design, especially considering computer- and conventional literacy. We will argue that the pathway to healthcare involves a range of communicative activities requiring different language skills and implying different technologies, not restricted to SLT. We will comment on the different situations which have been targeted by research in this field so far, and the impact of different target languages on research, and how the differing avilability of resources and software influences research. We also need to consider more carefully the design of the feedback and verification elements of systems, and the need for realistic evaluations.

2 Who are the users?

We start by looking at the assumed profile of users of medical SLT systems. Systems that have been developed so far can be divided into those for use in the doctors office – notably, MedSLT (Rayner and

Bouillon, 2002), CCLINC (Lee et al., 2002), and (honourable mention) the early work done at CMU (Tomita et al., 1988)[1] – and those for use for first contact with medical professionals "in the field", developed under DARPA's CAST programme:[2] MASTOR (Zhou et al., 2004), Speechalator (Waibel et al., 2003), Transonics (Narayanan et al., 2004) and SRI's system (Precoda et al., 2004). This distinction mainly motivates differences in hardware, overall design, and coverage, but there may be other more subtle differences that result especially from the situation in which it was envisaged that the CAST systems would be used.

Some descriptions of the systems talk of "doctors" and "patients" though others do use more inclusive terms such as "medical professional". A significant common factor in the descriptions of the systems seems to be that it is the doctor who controls the device. This may be because it can only handle one-way translation, as is the case of MedSLT, ". . . the dialogue can be mostly initiated by the doctor, with the patient giving only non-verbal responses" (Bouillon et al., 2005), or may be an explicit design decision:

> There is, however, an assymmetry in the dialogue management in control, given the *desire* for the English-speaking doctor *to be in control* of the device and the primary "director" of the dialog. (Ettelaie et al., 2005, 89) [emphasis added]

It is understandable that as a regular user, the medical professional may *eventually* have more familiarity with the system, but this should be reflected in there being *different* user-interfaces (see Somers and Lovel 2003). We find regrettable however the assumption that "the English speaker [. . .] is expected to have greater technological familiarity" (Precoda et al., 2004, 9) or that

> the medical care-giver will maintain the initiative in the dialogue, will have sole access to the controls and display of the translation device, and will operate the push-to-talk controls for both him or herself and the [P]ersian patient. (Narayanan et al., 2004, 101)

In fact, although the early use of computers in doctor–patient consultations was seen as a threat, more recently the help of computers to increase communication and rapport has begun to be recognised (Mitchell and Sullivan, 2001). This may be at the expense of patient-initiated activities however, and many practitioners are suspicious of the negative impact of technology on relationships with patients, especially inasmuch as it increases the perceived power imbalance in the relationship.

Figure 1, a snapshot from Transonics demo,[3] leaves in no doubt who is in control.

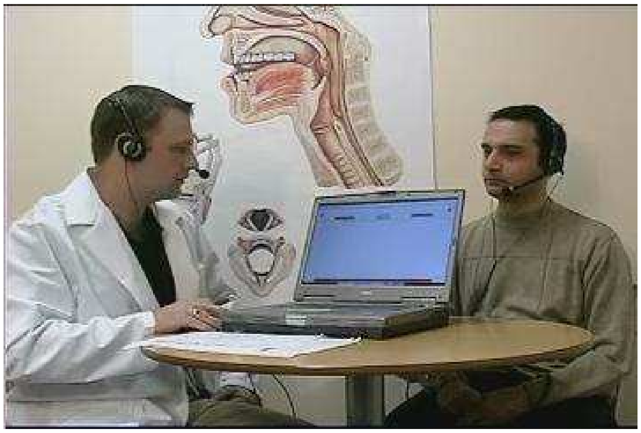


Figure 1: Snapshot from Transonics' demo movie. The patient is not even allowed to see the screen!

Equipment whose use and "ownership" can be equally shared between the participants goes some way to redressing the perceived power-balance in the consultation. We have evidence of this effect in ongoing experiments comparing (non-speech) communication aids on laptops and tablet PCs: with the laptop, controlled by a mouse or mouse-pad, the practitioner tends to take the initiative, while with the tablet, which comes with a stylus, the patient takes the lead. Bouillon et al. (2005) comment that "patients [. . .] will in general have had no previous exposure to speech recognition technology, and may be reluctant to try it." On the other hand, patients also have suffered from failed consultations

[1] We give here one indicative reference for each system.

[2] Formerly known as Babylon. See www.darpa.mil/ipto/ programs/cast/.

[3] http://sail.usc.edu/transonics/demo/transedit02lr.mov

which break down through inability to communicate, and in our experience are pleased to be involved in experiments to find alternatives. In our view, one should not underestimate patients' adaptability, or their potential as users of technology on an equal status with the practitioners.

This being the case, we feel that some effort needs to be devoted to usability issues. We will return to this below, but note that text-based interfaces are not appropriate for users with limited literacy (which may be due to low levels of education, visual impairment, or indeed the lack of a written standard for the language). Use of images and icons also needs to be evaluated for appropriateness, an issue not addressed in any of the reports on research in medical SLT that we have read. For example, Bouillon et al. (2005) show a screenshot which includes the graphic reproduced in Figure 2. The text suggests that the user (i.e. the doctor?) can click on the picture to set the topic domain. It is not clear why a graphic is more suitable for the doctor-user than a drop-down text menu; there is no mention of whether the patient is encouraged to use the diagram, but if so one wonders for what purpose, and if it is the best choice of graphic. Research (e.g. by Costantini et al. 2002) suggests that multimodal interfaces are superior to speech-only systems, so there is some scope for exploration here.

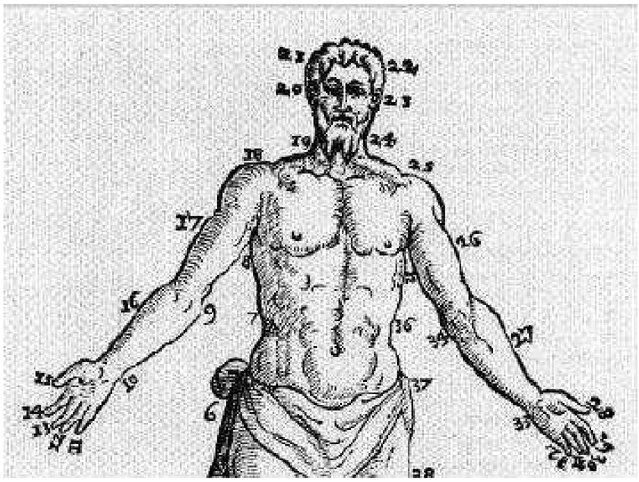
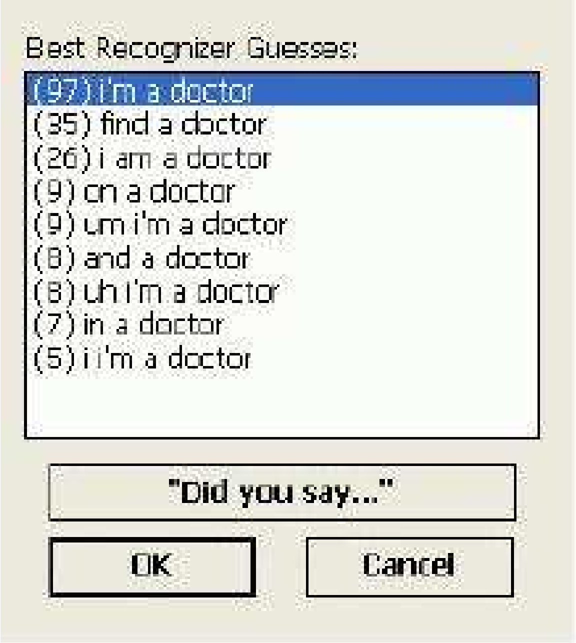


Figure 2: Graphic taken from screenshot in Bouillon et al. (2005)

Incorporating more symbolic graphics into an interface is an area of complexity, as Johnson et al. (2006) report. Iconic text-free symbols, for example to represent "please repeat", or "next question", or abstract concepts such as "very" are not always as instantly understandable as some designers think. Considering the use of symbols from AAC (augmentative and alternative communication) designed for speech-impaired disabled users by patients with limited English, we noticed that AAC symbol sets have a systematic iconicity that regular users learn, but which may be opaque to first-time (or one-time) untrained users (Johnson, 2004).

3 Feedback and verification

Translation accuracy is of course crucial in the medical domain, and sometimes problematic even with human interpreters, if not trained properly (Flores, 2005). Both speech recognition (SR) and translation are potential sources of error in the SLT chain, so it is normal and necessary to incorporate in SLT systems the provision of feedback and verification for users. The standard method for SR is textual representation, often in the form of a list of choices, for example as in Figure 3, from Precoda et al. (2004).

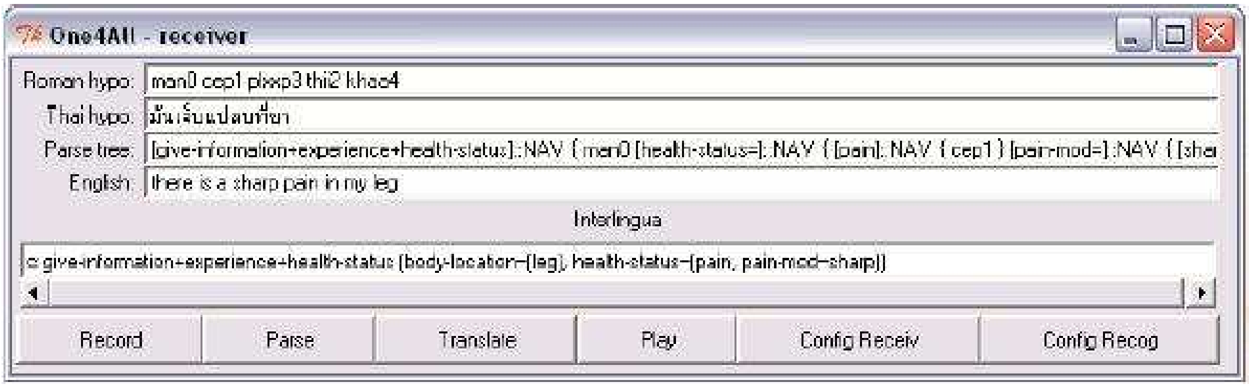


Best Recognizer Guesses:

(97) i'm a doctor
(35) find a doctor
(26) i am a doctor
(9) on a doctor
(9) um i'm a doctor
(8) and a doctor
(8) uh i'm a doctor
(7) in a doctor
(5) i i'm a doctor

Figure 3: Choice of recognizer outputs, from Precoda et al. (2004:10)

For translation output, some form of paraphrase or back-translation is offered, often facilitated by the



Roman hypo: man0 cep1 plxxp3 thii2 khaa4

Thai hypo: มันเจ็บแปลบที่ขา

Parse tree: [give-information+experience+health-status]::NAV (man0 [health-status=]::NAV ([pain]::NAV (cep1) [pain-mod=]::NAV ([sha

English: there is a sharp pain in my leg

Interlingua

c:give-information+experience+health-status (body-location=(leg), health-status=(pain, pain-mod=sharp))

Figure 4: Choice of recognizer outputs, from Precoda et al. (2004:10)

particular design of the machine translation (MT) component (e.g. use of an interlingua representation, as in MedSLT, Speechalator). In the Transonics system, the SR accuracy is automatically assessed by the MT component: SR output that conforms to the expectations of the MT systems grammar is preferred.

For the literate English-speaking user, this approach seems reasonable, although an interface such as the one shown in Figure 4, detailing the output of the parse must be of limited utility to a doctor with no linguistics training, and we must assume that the prototype is designed more for the developers' benefit than for the end-users.

For the patient with limited or no English, the issue of feedback and verification is much more difficult. As mentioned above, and reiterated by Precoda et al. (2004), the user may not be (wholly) literate, or indeed the language (or dialect) may not have an established writing system. For some languages, displaying text in the native orthography may be an added burden. Figure 5 shows Speechalator's Arabic input screen (Waibel et al., 2003). It is acknowledged that the users must "know something about the operation of the machine", and although it is stated that the display uses the writing system of the language to be recognised, in the illustration the Arabic is shown in transcription.

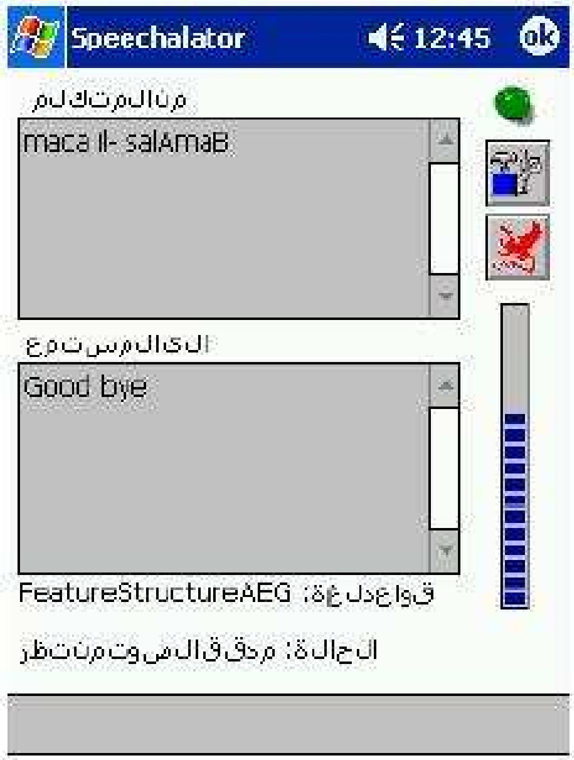
Another issue concerns the ease with which a lay user can make any sense of a task in which they are asked to judge a number of paraphrases, some ungrammatical. This is an intellectual task that is difficult for someone with limited education or no experience of linguistic "games". For example, for this reason we have rejected the use of semantically unpredictable sentences (SUS) (Benoît et al., 1996) in our attempts to evaluate Somali speech synthesis (Somers et al., 2006). This leads us to a consideration of how medical SLT can best be evaluated.

4 Evaluation

MT evaluation is notoriously difficult, and SLT evaluation even more so. Most researchers agree that measures of translation fidelity in comparison with a gold-standard translation, as seen in text MT evaluation, are largely irrelevant: a task-based evaluation is more appropriate. In the case of medical SLT this presumably means simulating the typical situation that the technology will be used in, which involves patients with medical problems seeking assistance.

Since SLT is a pipeline technology, the individual components could be evaluated separately, and indeed the effects of the contributing technologies assessed (cf. Somers and Sugita 2003). Once again, literacy issues will cloud any evaluation of speech recognition accuracy that relies on its speech-to-text function, and evaluation of speech synthesis must simulate a realistic task (cf. comments on SUS, above).

Evaluations that have been reported suggest using real medical professionals and actors playing the part of patients: this scenario is well established in the medical world, where "standardized patients" (SPs) – actors trained to behave like patients – have been used since the 1960s. One problem with SPs for systems handling "low density" languages like Persian, Pashto and so on, is the need for the vol-



من المتكلم

maca Il- salAmaB

الى المستمع

Good bye

قواعد لغة: FeatureStructureAEG

الحالة: مدقق الصوت منتظر

Figure 5: Speechalator's Arabic input screen (Waibel et al., 2003, 372)

unteers to understand English so that they can be trained as an SP, in conflict with the need for them to not understand English in order to give the system a realistic test. Ettelaie et al. (2005) for example report that their evaluation was somewhat compromised by the fact that two of their patient role-players did speak some English, while a third participant did not adequately understand what they were supposed to do.

Another problem is that there is no obvious baseline against which evaluations can be assessed. One could set up "with and without" trials, and measure how much and how accurately information was elicited in either mode. But this would be a waste of effort: it is widely, although anecdotally, reported that when patients with limited English arrive for a consultation where no provision for interpretation has been made, the consultations simply halt. It is also reported, as already mentioned, that human interpreters are not 100% reliable (Flores, 2005). Often, an untrained interpreter is used, whether a family member or friend that the patient has brought with them, or even another health-seeker who happens to be sitting in the waiting room. The potential for an unreliably interpreted consultation (or worse) is massive.

Ettelaie et al. (2005) mention a number of metrics that were used in their evaluation, but unfortunately do not have space for a full discussion. The principle metric is task completion, but they also mention an evaluation of a scripted dialogue, with translations evaluated against model translations using a modified version of BLEU, and SR evaluated with word-error rate. These do not seem to me to be extremely valuable evaluation techniques.

Starlander et al. (2005) report an evaluation in which the translations were judged for acceptability by native speakers. Given the goal-based nature of the task, rating for intelligibility rather than acceptability might have been more appropriate, though it is widely understood that the two features are closely related. On the positive side, Starlander et al. used only a three-point rating ("good", "ok" or "bad"): evaluations of other target languages might be subject to the problem, reported by Johnson et al. (in prep.) and by ADD REF that rating scales are highly culture-dependent, so that for example Somali participants in an evaluation of the suitability of symbols in doctor–patient communication mostly used only points 1 and 7 of a 7-point scale.

Another evaluation method[4] is to assess the number and type of translation or interpretation errors made, including whether there was any potential or actual error of clinical consequence.

As Starlander et al. (2005) say:

> In the long-term, the real question we would like to answer when evaluating the prototype is whether this system is practically useful for doctors

to which we can only add, reiterating our comments in Section 2, ". . . and for patients".

5 The Pathway to Healthcare

Let us move on finally to a more wide-ranging issue. "Medical SLT" is often assumed to focus on doctorpatient consultations or, as we have seen in

[4] Thanks to the anonymous reviewer for pointing this out.

the case of systems developed under the CAST programme, interactions between medical professionals and affected persons in the field. Away from that scenario, although it is natural to think of "going to the doctor" as involving chiefly an interview with a doctor, and while everything in medical practice arguably derives from this consultation, the pathway to healthcare in normal circumstances involves several other processes, all of which involve language-based encounters that present a barrier to patients with limited English. None of the medical SLT systems that have been reported in the literature address this variety of scenarios, although the website for the Phraselator (which is of course not an SLT system as such) does list a number of different scenes, such as the front desk, labour ward and so on.

In this section, we would like to survey the pathway to healthcare, and note the range of language technologies – not always speech or translation oriented – that might be appropriate at any point. The purpose of this is both to make a plea to widen our vision of what "medical SLT" covers, but also to note that SLT is not necessarily the most appropriate technology in every case.

The pathway might begin with a person suspecting that there may be something wrong with them. Many people nowadays would in this situation first try to find out something about their condition on their own, typically on the Web, though of course there is still a major "digital divide" for racial and ethnic minorities, and the poor, partly due to the langauge barriers this research is addressing. If you need this information in your own language, and you have limited literacy skills, technologies implied are multilingual information extraction. MT perhaps coupled with text simplification, with synthesized speech output. For specific conditions which may be treated at specialist clinics (our own experience is based on Somalis with respiratory difficulties) it may be possible to identify a series of frequently asked questions and set up a pre-consultation computer-mediated help-desk and interview (cf. Osman et al. 1994). See Somers and Lovel (2003) for more details.

Having decided that a visit to the doctor is indicated, the next step is to make an appointment. Appointment scheduling is the classical application of SLT, as seen in most of the early work in the field, and is a typical case of a task-oriented cooperative dialogue. Note that the "practitioner" – the receptionist in the clinic – does not necessarily have any medical expertise, nor possibly the high level of education and openness to new technology that is often assumed in the literature on medical SLT which talks of the "doctor" controlling the device.

If this is the patient's first encounter with this particular healthcare institution, there may be a process of gathering details of the patient's medivcal history and other details, done separately from the main doctor–patient consultation, to save the doctor's time. This might be a suitable application for computer-based interviewing (cf. Bachman 2003).

The next step might be the doctor–patient consultation, which has been the focus of much attention. For no doubt practical purposes, some medical SLT developers have assumed that the patients role in this can be reduced to simple responses involving yes/no responses, gestures and perhaps a limited vocabulary of simple answers at the limit. This view unfortunately ignores current clinical theory. *Patient*-centred medicine (cf. Stewart et al. 2003) is widely promoted nowadays. The session will see the doctor eliciting information in order to make a diagnosis as foreseen, but also explaining the condition and the treatment, and exploring the patients feelings about the situation. While it may be unrealistic at present to envisage fully effective support for all these aspects of the doctorpatient consultation, we feel that its purpose should be explicitly appreciated, and the limitations of current technology in this respect acknowledged.

After the initial consultation, the next step may involve a trip to the pharmacist to get some drugs or equipment. Apart from the human interaction, the drugs (or whatever) will include written instructions and information: frequency and amount of use, contraindications, warnings and so on. This is an obvious application for controlled language MT: drug dose instructions are of the same order of complexity as weather bulletins. For non-literate patients, "talking pill boxes" are already available:[5] why can't they talk in a variety of languages?

Another outcome might involve another practitioner – a nurse or a therapist – and a series of meet-

[5] Marketed by MedivoxRx. See Orlovsky (2005).

ings where the condition may be treated or managed. Apart from more scheduling, this will almost certainly involve explanations and demonstrations by the practitioner, and typically also elicitation of further information from the patient. Hospital treatment would involve interaction with a wide range of staff, again not all medical experts. If a communication device is to be used, it makes more sense for it to be under the control and "ownership" of the person who is going to be using it regularly: the patient.

6 Conclusion

Some of the comments made in this position paper may seem critical, but it has not been my intention to be negative about the field.[6] It has been my intention in this paper to draw attention to the following aspects of medical SLT which I believe so far have been somewhat neglected:

- What is the ideal user profile for medical SLT? Should the doctor control the system, or could it be seen as a shared resource?

- If the patient is also a user, devices need to be more user-friendly, taking into account cultural differences, and problems of low literacy.

- This particularly applies to feedback and verification modules in the system.

- Evaluation should focus on the ability of the technology to aid the completion of the task, from the perspective of both the practitioner and the patient.

- Evaluation methods should not involve participants in meaningless or incomprehensible tasks (such as rating nonsensical output), nor rely on skills (such as literacy) that they may lack.

- The pathway to healthcare involves more than the one-way doctor–patient dialogues covered by most systems. A wide range of technologies can be brought to bear on the problem.

Acknowledgments

I would like to acknowledge the contribution to this paper of my colleagues on the CANES project, Ann Caress, Gareth Evans, Marianne Johnson, Hermione Lovel, Zeinab Mohamed. Some of the work reported here is funded by the Economic and Social Research Council (ESRC), project reference RES-000-23-0610. Thanks also to the anonymous referees for their very useful comments.

References

Bachman, J.W. 2003. 'The patient-computer interview: a neglected tool that can aid the clinician.' *Mayo Clinic Proceedings*, 78:67–78.

Benoît, Christian, Martine Grice and Valérie Hazan. 1996. 'The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences'. *Speech Communication*, 18:381–392.

Bouillon, Pierrette, Manny Rayner, Nikos Chatzichrisafis, Beth Ann Hockey, Marianne Santaholma, Marianne Starlander, Yukie Nakao, Kyoko Kanzaki and Hitoshi Isahara. 2005. 'A generic multi-lingual open source platform for limited-domain medical speech translation'. In *Proceedings of the Tenth Conference on European Association of Machine Translation*, Budapest, Hungary, pp. CHECK

Costantini, Erica, Fabio Pianesi and Susanne Burger. 2002. 'The added value of multimodality in the NESPOLE! speech-to-speech translation system: an experimental study'. In *Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, Pittsburgh, PA, pp. 235–240.

Ettelaie, Emil, Sudeep Gandhe, Panayiotis Georgiou, Robert Belvin, Kevin Knight, Daniel Marcu, Shrikanth Narayanan and D. Traum. 2005. 'Transonics: A practical speech-to-speech translator for English-Farsi medical dialogues'. In *43rd Annual Meeting of the Association for Computational Linguistics: ACL-05 Interactive Poster and Demonstration Sessions*, Ann Arbor, MI, pp. 89–92.

Flores, Glenn. 2005. 'The impact of medical interpreter services on the quality of health care: a systematic review'. *Medical Care Research and Review*, 62:255–299.

Johnson, M.J. 2004. 'What can we learn from drawing parallels between people who use AAC and people whose first language is not English?' *Communication Matters*, 18(2):15–17.

[6] In particular, it should perhaps be acknowledged that in terms of practical accomplishment we have yet to match others in the field.

Johnson, M.J., D.G. Evans and Z. Mohamed. 2006. ‘A pilot study to investigate alternative communication strategies in provider-patient interaction with Somali refugees’. In *Current Perspectives in Healthcare Computing Conference*, Harrogate, England, pp. 97–106.

Johnson, M.J., G. Evans, Z. Mohamed and H. Somers (in prep.) An investigation into the perception of symbols by UK-based Somalis and English-speaking nursing students using a variety of symbol assessment techniques.

Lee, Young-Suk, Daniel J. Sinder and Clifford J. Weinstein. 2002. ’Interlingua-based English–Korean two-way speech translation of doctor–patient dialogues with CCLINC’. *Machine Translation*, 17:213–243.

Mitchell, E. and F. Sullivan. 2001. ‘A descriptive feast but an evaluative famine: systematic review of published articles on primary care computing during 1980-97’. *British Medical Journal*, 322:279–282.

Narayanan, S., S. Ananthakrishnan, R. Belvin, E. Ettelaie, S. Gandhe, S. Ganjavi, P. G. Georgiou, C. M. Hein, S. Kadambe, K. Knight, D. Marcu, H. E. Neely, N. Srinivasamurthy, D. Traum, and D. Wang. 2004. ‘The Transonics spoken dialogue translator: an aid for English-Persian doctor-patient interviews’. In Timothy Bickmore (ed.) *Dialogue Systems for Health Communication: Papers from the 2004 Fall Symposium*, American Association for Artificial Intelligence, Menlo Park, California, pp. 97–103.

Orlovsky, Christina. 2005. ‘Talking pill bottles let medications speak for themselves’. *NurseZone.com* (online magazine), www.nursezone.com/Job/DevicesandTechnology.asp ?articleID=14396. Accessed 15 March 2006.

Osman, L., M. Abdalla, J. Beattie, S. Ross, I. Russell, J. friend, J. Legge and J. Douglas. 1994. ‘Reducing hospital admissions through computer supported education for asthma patients’. *British Medical Journal*, 308:568–571.

Precoda, Kristin, Horacio Franco, Ascander Dost, Michael Frandsen, John Fry, Andreas Kathol, Colleen Richey, Susanne Riehemann, Dimitra Vergyri, Jing Zheng and Christopher Culy. 2004. ‘Limited-domain speech-to-speech translation between English and Pashto’. In *HLT-NAACL 2004, Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 9–12.

Rayner, Manny and Pierrette Bouillon. 2002. ’A flexible speech to speech phrasebook translator’. In *Proceedings of the ACL-02 Workshop on Speech-to-Speech Translation: Algorithms and Systems*, Philadelphia, PA, pp. 69–76.

Somers, Harold, Gareth Evans and Zeinab Mohamed. 2006. ‘Developing speech synthesis for under-resourced languages by “faking it”: an experiment with Somali’. In *Proceedings of LREC: 5th Conference on Language Resources and Evaluation*, Genoa.

Somers, Harold and Hermione Lovel. 2003. ‘Computer-based support for patients with limited English’. In *Association for Computational Linguistics EACL 2003, 10th Conference of The European Chapter, Proceedings of the 7th International EAMT Workshop on MT and other language technology tools*, Budapest, pp. 41–49.

Somers, Harold and Yuriko Sugita. 2003. ‘Evaluating commercial spoken language translation software.’ In *MT Summit IX: Proceedings of the Ninth Machine Translation Summit*, New Orleans, pp. 370–377.

Starlander, Marianne, Pierrette Bouillon, Manny Rayner, Nikos Chatzichrisafis, Beth Ann Hockey, Hitoshi Isahara, Kyoko Kanzaki, Yukie Nakao and Marianne Santaholma. 2005. ‘Breaking the language barrier: machine assisted diagnosis using the medical speech translator’. In *Proceedings of the XIX International Congress of the European Federation for Medical Informatics*, Geneva, Switzerland.

Stewart, Moira, Judith Belle Brown, W. Wayne Weston, Ian R. McWhinney, Carol L. McWilliam and Thomas R. Freeman. 2003. *Patient-Centered Medicine: Transforming the Clinical Method* (2nd ed.). Radcliffe, Abingdon, Oxon.

Tomita, Masaru, Marion Kee, Hiroaki Saito, Teruko Mitamura and Hideto Tomabechi. 1988. ’The universal parser compiler and its application to a speech translation system’. In *Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Pittsburgh, Pennsylvania, pages not numbered.

Waibel, Alex, Ahmed Badran, Alan W. Black, Robert Frederking, Donna Gates, Alon Lavie, Lori Levin, Kevin Lenzo, Laura Mayfield Tomokiyo, Jürgen Reichert, Tanja Schultz, Dorcas Wallace, Monika Woszczyna, and Jing Zhang. 2003. ‘Speechalator: two-way speech-to-speech translation on a consumer PDA’. In *Proceedings of EUROSPEECH 2003, 8th European Conference on Speech Communication and Technology*, Geneva, pp. 369–372.

Zhou, Bowen, Daniel Déchelotte and Yuqing Gao. 2004. ‘Two-way speech-to-speech translation on handheld devices’. In *INTERSPEECH 2004 – ICSLP, 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, pp. 1637–1640.