Computer Assistance in Bilingual Task-Oriented Human-Human Dialogues

Sven Schmeier, Matthias Rebel, Renlong Ai

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
Alt-Moabit 91c, 10559 Berlin, Germany
{schmeier, matthias.rebel, renlong.ai}@dfki.de

Abstract. . In 2008, the percentage of people with a migration background in Germany had already reached more than 15% (12 Million people). Among that 15%, the ratio of seniors aged 50 years or older was 30% [1]. In most cases, their competence of the German language is adequate for dealing with everyday situations. However sometimes in emergency or medical situations, their knowledge of German is not sufficient to communicate with medical professionals and vice versa. These seniors are part of the main target group within the German Ministry of Research and Education (BMBF) research project SmartSenior [2] and we have developed a software system that assists multilingual doctor-patient conversations to overcome language and cultural barriers. The main requirements of such a system are robustness, accurate translations in respect to context and mobility, adaptability to new languages and topics and of course an appropriate user interface. Furthermore, we have equipped the system with additional information to convey cultural facts about different countries. In this paper, we present the architecture and ideas behind the system as a whole as well as related work in the area of computer aided translation and a first evaluation of the system.

Keywords: language barriers, human-human dialogue system, health care

1 Introduction

The research project SmartSenior aims to develop technological and comprehensive solutions to aid senior citizens in leading independent lives. The aim is to support elderly people in their day-to-day lives and social interactions, and, in terms of health, support them in a way that helps them continue living within their familiar environment. The project targets both senior citizens leading overall independent lives, as well as acutely or chronically ill elderly people in need of assistance and care.

Within the context of “support”, a subproject in SmartSenior is dedicated to helping seniors with migration backgrounds carry out conversations, especially in emergency situations.

High quality patient-centered care depends heavily on communication. Understanding the patient's needs and abilities to follow their doctor’s advice are the keys for

successful treatment [3]. In multicultural, multilingual contexts, language and cultural barriers present a key challenge to effective communication between patient and service provider. Unfortunately, access to translation services or human translators, is not always available, especially in emergency situations where time is a very limiting factor. Numerous studies show the effect that the absence of language interpreters have in such situations ([4], [5], and [6]). Furthermore, those translators need to be trained to handle specific situations, especially when dealing with patients that come from different social cultures. It is also known that interpreter errors may occur if untrained ad-hoc interpreters are used [7].
Starting from these observations, we developed a software system that makes conversations across different languages possible.

The main requirements on such a system are listed in the following points:

1. Robustness: the system should work robustly and not quit services under any circumstances
2. Accuracy: the system should provide accurate translation
3. Mobility: the system must be accessible anywhere and anytime
4. Adaptivity: the system should be able to handle new languages and topics
5. Usability: the user interface must be designed with respect to the target group

To meet these requirements the following design, implementation and platform decisions have been made: with respect to mobility and usability, we chose touch-sensitive Tablet PC as our target platform. The software system is divided into UI and database content sections. This facilitates the easy incorporation of new languages, situations and information. The system itself is designed as a cross lingual, mixed-initiative dialogue and information system. The content of the dialogues have been created with the help of doctors and medical staff personnel with respect to the given task. The translations have been done by human expert interpreters. The system itself runs off-line, no internet connection is needed.
In the remainder of this paper, we first present related work that has been done in the area of automated translation (services). In the next sections, we describe the system specifications, the dialogue design and the evaluation of our first in house test series. We finally end with the conclusion and outlook.

2 Related Work

Overcoming language barriers has been the focus of research for more than six decades. Since then much progress has been made, starting from bilingual dictionaries and hand coded rules for the final translation, to linguistic parsing of the input in one language generating the output in a second language. However, due to the high ambiguity of natural language the results are still poor compared to the costs. Restrictions of the domain improve translation quality, but the process of adapting new or different domains turned out to be very costly.

With the fast developing and growing internet and search engine companies, huge amounts of corpora have been collected in the last decade. Hence research has focused more and more on pure statistical machine translation. Machine translation services like Babelfish[1] or Google Translate[2] are now available for the public. Unfortunately, the accuracy is still quite poor. For example, the German sentence: "Haben Sie heute morgen schon Wasser gelassen?" (Engl.: "Did you urinate this morning?") produces the translation: "Did you this morning has left the water?"
Another kind of translation service are phrasebooks which contain frequently used sentences and their translation. In the last two years, more and more electronic phrasebooks have become available for mobile devices. Phrasebooks allow users to express a limited set of utterances in a language they do not know. Unfortunately, the partner is unable to reply, i.e. they do not allow dialogues at all. However, phrasebook translation is more accurate because the phrases have been translated by humans.
In the research project COMPASS2008 [8] a bilingual information and translation service system for mobile devices was developed. The goal of the project was to develop a system to empower foreign visitors during the 2008 Olympic Games that integrates functionalities of a phrasebook, a translation aid and a powerful information system that is connected to various services via the internet [9]. The main ideas behind this system have directed to our research for overcoming language and cultural barriers.

3 System specification

Our cross lingual translation and information system consists of two main components: One component is the system for overcoming language and cultural barriers that runs on the user's tablet. Dialogues and information data are stored and maintained in a database and our user interface as well as the Text-To-Speech (TTS) system also run on the device. Its second component is the system for creating and managing dialogues, cultural information and translations.

3.1 System on the device

Our system is designed according to the Model View Controller (MVC) pattern [10]. The user interface, the database containing the cross lingual dialogues and dictionary, and the interaction controller between these two layers form independent modules which communicate via object messages. It runs as an App on an Apple iPad with the underlying operating system iOS 3.2 and does not need an internet connection.

---

[1] http://babelfish.yahoo.com/
[2] http://translate.google.com

3.2 Functionality

We implemented our solution as a mixed-initiative dialogue system. The dialogues are divided into several main dialogue situations. Each situation provides several phrases and a set of more specific sub-situations, which again contain several phrases. The user initiates a dialogue by selecting a phrase in his language. This phrase is then translated and presented to the partner along with a set of phrases designed as responses. The partner may then select a phrase which in turn is translated and presented to the user along with a relevant set of possible responses and so on.

The system provides three types of phrases:

1. **Static phrases** are sentences that are translated immediately after being selected/touched.
2. **Dynamic phrases** contain slots that have to be completed. These slots have a specific type with expected values. Possible types are (a) part_of_body, (b) diseases, (c) allergies, (d) dictionary and so on. They are represented as clickable buttons within the phrases. There are four distinct types of slots: select lists (like a, b, c), empty fields to fill with values (i.e. measuring units) and special forms to enter time or date expressions, and selectable pictures to enhance communication and understanding.
3. **Information phrases** can either be static or dynamic and they contain an additional information symbol. By touching this symbol the system provides context-aware cultural information.

The system also takes into account that sometimes patients may not be able to respond by using the device. In these cases, the doctor is able to use an appropriate single mode with further options (i.e. phrases like “Wink, if you can hear me!”, “Wink once for no and twice for yes!”, “Everything will be fine, I’ll take care of you.”).
Fig.1 shows an example dialogue between a Turkish speaking senior and a German speaking doctor. The doctor first chooses the desired dialogue situation, in our case “Notfall zu Hause” (Emergency At Home) shown on the left screen. After that the system computes the phrases that will be displayed to the doctor. In our case, the doctor chooses the phrase “Können Sie beschreiben was genau passiert ist?” (“Please describe what happened.”) which will be translated for the senior as “Tam olarak ne olduğunu anlatabilir misiniz?“. Now the senior can answer the doctor by simply touching one of the phrases offered.
The second screen displays the information phrase “Ich muss Sie jetzt untersuchen.” (Engl.: “I need to examine you”). After clicking on the information symbol, the system shows the following hint: A western male doctor should not examine female persons with a Muslim background without any relatives in the room.

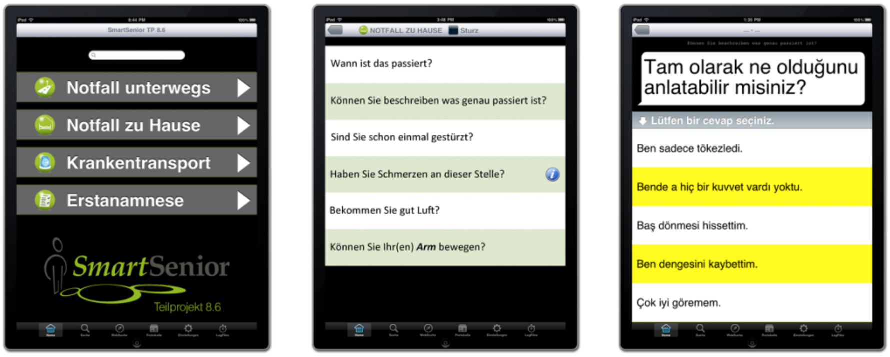

Fig. 1. Screens of the dialogue system on the mobile device

3.3 Additional components on the device

The system architecture allows the integration of additional components. So far we have experimented with two components in order to provide alternative, more comfortable and less restricted forms of user interaction: Text-To-Speech (TTS) synthesis and Automatic Speech Recognition (ASR).

Speech-Output – TTS.

In many situations, it is important for the system to provide not only visual output but also spoken language - especially in situations where the senior is not able to read, e.g. because of impaired vision or illiteracy. Some languages, for example, Chinese, use special symbols for certain technical terms which may be unknown to the senior but are understandable when spoken. Spoken language also helps increase confidence between the doctor and patient.
The fact that our architecture utilizes dynamic phrases makes it impossible to prerecord the phrases and play them when necessary. A phrase like “Please give me [dictionary]” could produce thousands of translations depending on the user’s input for the slot dictionary. That is why we decided to make use of the commercial TTS system products already available on mobile devices. We tried three vendors: Loquendo, Nuance and SVOX. All systems provide voice output, synthesized spoken language, which sounds natural and is easy to understand. However for technical reasons, we chose the Mobile TTS Standard by SVOX [11]. It supports 27 languages including nearly all European languages such as: German, English, French, Spanish, Italian, Danish, Turkish, etc. It also supports languages like Mandarin, Cantonese, Arabic, Taiwanese, and Korean. Further key features that are important for our system are:

- Dynamic switching between languages
- Flexible voice control in terms of speed, pitch and volume: for our target group, the voice should be loud, slow and have a clear pronunciation

- Support of phonetic alphabets: this feature guarantees correct articulation for names of medical products
- Synthesis of mixed languages: this is useful for the names of specific products or people

Speech Input – ASR.

At the moment, it is not technologically possible to perform 100% accurate speech recognition and on portable devices where CPU and memory are restricted, the ability to handle dictation (i.e. to recognize everything spoken) is even less reliable. However if the number of phrases that have to be recognized are limited, recognition accuracy can be improved dramatically. In our case, although we do provide phrases that cover many situations and their sub-situations, the number of these phrases is still limited. Under this condition, an ASR system that recognizes these phrases can be deployed to our target portable device. We have run tests with Nuance, SVOX and Fonix ASR engines on HTC HD mobile phones[3] (528MHz CPU and 288 MB RAM) in both English and German. The best engine reached 72.69% accuracy for recognition. This result relates to a 68.67% recognition rate for outdoor usage and 74.28% for indoor usage. [12].
Taking into account that our tests were carried out by colleagues in a stress-free situation who were willing to speak very clearly and who held the device at the proper distance and in the right direction, we estimate that the result will decrease significantly in real life emergency situations with seniors. So for the moment, we have abandoned the idea of implementing ASR in our final system. In the future we plan to continue testing and evaluating ASR systems on the iPad. (See last chapter for more information)

3.4 Dialogue Management System (DMS)

The system's quality depends heavily on the dialogues and information that are provided. This pertains to both the quantity of the dialogue situations and their sub-situations as well as to the quality of the questions and their answers. Each emergency situation is different in respect to the course of events but also with respect to the people involved and though the dialogues should be as general as possible, it is important that each conversation seem unique to the patient. A further factor is the screen size which plays a very limiting role for the development of the dialogues. The sentences or phrases can not be too long but they have to include the main information needed for the dialogue.

Our DMS has been designed as a browser-based WYSIWYG[4] system that fully supports the dialogue's author, editor and translator in their specific needs. It provides all necessary operations, such as:

---

[3] At the time of our tests, the ASR systems running on the iPad were still under development.
[4] What You See Is What You Get

- adding and deleting languages
- translating phrases, dictionary entries or names of dialogues and sub-dialogues
- as well as adding/deleting/modifying dialogues, sub-dialogues, phrases and slot types and content

All of these different levels - i.e. representation of the system, the dialogue structure and the phrases in their context - are important for developing advanced dialogues and providing context-aware translations. That is why we provide two views for the managing environment. One is an application like view (Fig 2, left) and the other one is a ScalableVectorGraphic[5] representation (Fig.2, right) of a specific dialogue situation which gives a general overview of context, structure and flow. The DMS exports the dialogues, which were created, into the appropriate database format for the system on the device.

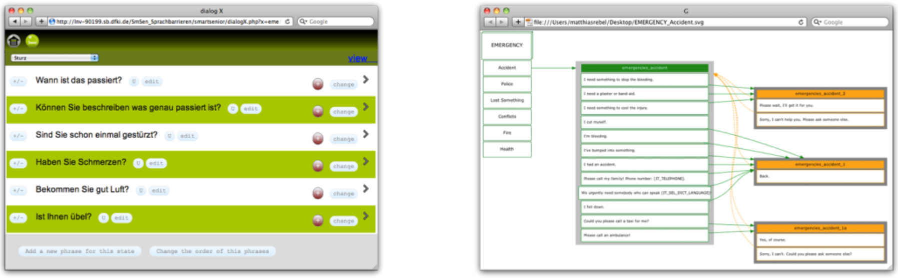

Fig. 2. Screens of the DMS with described functionalities

4 Task-oriented structure of dialogues

In this section, we introduce in more detail some of the dialogues that have been implemented in our system and the rationale behind their development. There are many different places and situations in which medical assistance is carried out for e.g. at hospitals, at home, during transport etc. Each of these situations are represented as a main menu item. Hence at the first level, there are four Situations: *Emergency at Home*, *Emergency On the Way*, *Ambulance* and the situation independent topic *Anamnesis*.

Each situation provides, with respect to its topic, a set of specific phrases and sub-situations with specific phrases. In case of an accident, it is important to understand what happened and why in order to better understand which injures or problems were caused by the accident. The answers to these question lead to the doctor's first assessment. In severe health incidents, the doctor needs to act immediately making it necessary that the number of questions are limited and well organized. They also have to be simple and easy to understand and each question should be answerable by yes or no. In extreme cases, the patient may be immobilized and unable to operate the

[5] SVG support to augment objects in the picture with URLs.

device, so we provide questions that allow alternative ways for the user to respond within the dialogue (like nodding or head shaking or blinking). There are also situations, when no information is expected or the patient can not respond at all. These dialogues are designed to calm down the patient and explain what has happened and is happening. In such situations, the capability of the system to produce TTS synthesis is especially important.
The area of medical examinations is a wide field and exhaustive coverage can clearly not be expected from the limited dialogues provided in our system. Especially the anamnesis of a patient is very difficult and tedious to model. Nevertheless, the British Red Cross offers an emergency phrasebook[6] which covers the most common medical questions and terms to help first-contact-staff communicate with patients who do not speak English. This phrasebook is meant to make initial assessment in case there is no interpreter. For our system, we drew on the phrases contained in the phrasebook and added specialized sentences for seniors.

5 Evaluation

The system will be fully evaluated in field-tests in November 2011. Currently, we have finished the first test series. A second test series specifically to evaluate the UI design for elderly people is planned.

5.1 Setup

Each test was performed by two persons[7] role-playing a doctor-patient dialogue. After a brief introduction to the system and how to use it, the testers were asked to perform three different tasks (easy, medium and hard). The order of the tasks was randomized for each pair of testers. After each task, both testers had to rate several statements on a Likert scale and a general questionnaire had to be filled out after completing the entire test with questions concerning the user interface, dialogue structure, and difficulties encountered during the course of the dialogue etc.

5.2 Results

Most people judged the system as an appropriate means of communication between people who speak different languages (58% agree, 14% totally agree). More than 70% of the testers assessed the level of politeness and naturalness of the presented phrases as good. The rating for the dialogue flow, i.e. dialogue structure and hierarchies, the efficiency of the phrases, and the time needed to respond was mixed. We observed that subdividing a dialogue in sub-dialogues sometimes and too many phrases on one page produce longer latencies per turn. Furthermore, the testers had some problems

---

[6] http://www.dh.gov.uk/ en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4073230
[7] The testers had not seen the system before.

with dynamic phrases, i.e. phrases with slots that had to be completed. We noticed that, independent of task order, these problems became less prominent during the course of the test.
In summary, we gained useful feedback concerning the user interface as well as dialogue and phrase structures. The key issues to improve the system's usability will be determining the optimal weight between static vs. dynamic phrases and the ratio between dialogue hierarchy, number of phrases per page and general expressivity.

6 Conclusion and Outlook

We presented a system that helps overcome language and cultural barriers between doctors or medical staff and seniors with a migration background. The system is designed as a cross lingual, mixed-initiative dialogue and information system running on a Tablet PC. These conditions meet our main requirements: robustness, accurate translation, mobility, scalability with respect to new languages and content as well as appropriate user interfaces for seniors.
The content for the dialogues and information have been designed and developed with the help of doctors and medical staff. We have also built a Dialog Management System which allows us to easy integration of content into the system.
A first in house evaluation has been done which confirmed several aspects concerning the general user interface, the appropriateness of the dialogues and the overall handling of the various system components including the Text-To-Speech module.
From field-tests planned for November 2011 and the beginning of 2012, we expect to gain new insights and results that will be integrated in our system. Even though our dialogues and content are not yet certified for hospitals, we hope our system can be a beginning to help support the integration of people with a migration background into society.
In the near future, we will concentrate on incorporating speech input into our system. As we learned from an interview with a practitioner the capability of recognizing speech would be extremely useful for patients and doctors. Currently, we are evaluating Speech Recognition (ASR) systems on the iPad and hope to find ways to achieve satisfactory recognition rates. We also have to evaluate whether the speed, i.e. the time between speaking and retrieving the results is reasonable in this context. Furthermore, the current rapid development of tablets will no doubt have positive effects on developmental concerns. We expect a vast improvement of hardware and operating systems which will lead to better noise reduction for outdoor usage, better signals for the ASR module and the possibility of accessing larger speech models for improved recognition accuracy.

7 Acknowledgements

This system has been developed in the project SmartSenior. The SmartSenior project consists of a consortium of 29 project partners of either industrial or research background and it is funded by the German Ministry of Research and Education (BMBF, FKZ 16KT0902) within the course of the "High-Tech-Strategie-Deutschland" program.

8 References

1. Statistisches Bundesamt: Bevölkerung und Erwerbstätigkeit; Bevölkerung mit Migrations-hintergrund - Ergebnisse des Mikrozensus 2008. Wiesbaden 2010

2. SmartSenior (FKZ 16KT0902) Project Website: http://www.smartsenior.de

3. Elderkin-Thompson V, Silver RC, Waitzkin H: When nurses double as interpreters: a study of Spanish-speaking patients in a US primary care setting.

4. Siejo R, Gomez H, Friendenferg J: Language as a communication barrier in medical care for Hispanic patients. In Hispanic psychology-critical issues in theory and research. Edited by: Padilla AM. California: Sage Publication; 1995:169-81.

5. Smedley BD, Stith AY, Nelson AR: Unequal treatment. Confronting racial and ethnic disparities in health care. Washington: The National Academies Press; 2003.

6. Flores G: The impact of medical interpreter services on the quality of health care: a systematic review. Med Care Res Rev 2005 , 62(3):255-99.

7. Flores G, Laws MB, Mayo SJ, Zuckerman B, Abreu M, Medina L, Hardt EJ: Errors in medical interpretation and their potential clinical consequences in pediatric encounters. Pediatrics 2003 , 111(1):6-14.

8. Aslan Ihlan, Xu Feiyu, Uszkoreit Hans, Krüger Antonio, and Steffen Jörg, COMPASS2008: Multimodal, multilingual and crosslingual interaction for mobile tourist guide applications: In Proceedings of Intelligent Technologies for Interactive Entertainment (Intetain) 2005.

9. Uszkoreit Hans, Xu Feiyu, Liu Weiquan, Steffen Jörg, Aslan Ilhan, Liu Jin, Müller Christel, Holtkamp Bernhard, Wojciechowski Manfred: A Successful Field Test of a Mobile and Multilingual Information Service System COMPASS2008, In Proceedings of HCI International 2007, 12th International Conference on Human-Computer Interaction, Beijing, 2007.

10. Gammar Erich, Helm Richard, Johnson Ralph, Vlissides John: Design Patterns: Elements of Reusable Object-Oriented Software, Addison-Wesley Longman, Amsterdam, 2004

11. SVOX AG: SVOX Manual: Speech Output Engine SDK 4.3.1; Jul 17, 2009

12. Renlong Ai: Fuzzy Match between Speech Input and Textual Database, Master's thesis, Technical University Berlin, 2010