Challenges and Solutions of Multilingual and Translingual Information Service Systems

Hans Uszkoreit[1], Feiyu Xu[1], Weiquan Liu[2]

[1]DFKI GmbH, LT Lab, Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
{feiyu, uszkoreit}@dfki.de

[2] Capinfo Co. Ltd., Yu Min Dong Lu, Xi Cheng District
Beijing, 100029 P. R. China
liuweiquan@capinfo.com.cn

Abstract. In this paper, we present a survey of challenges and solutions of multilingual and translingual information service systems. In contrast to the computational linguistics literature on such systems, we are approaching the theme here from an HCI perspective. We will argue for a strategy that reduces reliance on automatic free-text translation, language input and classical information retrieval while not giving up these less reliable technologies altogether. We will also opt for a close situation-driven integration of information and communication functionalities. The described solutions have been incorporated into a novel mobile combined information and communication system for foreign tourists that has been tested under realistic conditions by users from several countries. The system is developed by the German-Chinese cooperation project COMPASS2008[1], a research action within the Digital Olympics framework.

Keywords: human computer interaction, multimodal interaction, multilingual and crosslingual strategies

1 Introduction

The term multilingual and translingual information services encompasses a wide scale of systems, ranging from simple multilingual database front ends and search engines all the way to sophisticated combinations of translingual information and communication services. In this paper we will approach these types of advanced information systems from an HCI perspective. Such an approach could mean that we investigate multilingual and translingual information services with respect to the special requirements that such applications pose to their user interfaces. Indeed, some of our

[1] www.compass2008.org

discussion will cover such considerations. However, the main focus of our overview article will be a view of multilingual and translingual capabilities as advanced interface functionalities that are cleverly combined with other interface capabilities and thus help the users to solve their real problems.

The first approach might for instance ask how an interface to a crosslingual search engine realizes the language selection through the users and how relevant documents in several languages are sorted or marked according to their languages. It may also investigate methods for indicating multiple translations of queries, arising from readings of the search keys that the user had not been aware of. These are all very important concerns giving rise to a multitude of open issues.

The second approach does not consider the linguistic competence as part of the core functionality of a system--as important as it may be for the intended application. It separates the purpose of the system from the language issue and lifts the linguistic functions up to the other interface properties. This may seem a trivial decision and, moreover, not really a novel view, but we will try to demonstrate that such a view if consequently followed can have interesting and far-reaching implications for a new class of communication and information systems.

This view assumes that mastering a foreign language usually is not the main goal of the user and thus not the driving force of the software design. This situation is different for computer-assisted language learning software, but for information and communication services, the central user goal is to obtain and understand relevant information and to communicate with other people. The language barrier can be a problem for reaching this goal, such as missing connectivity is a problem for web information services, a noisy environment for voice communication and a missing keyboard for text input.

In this paper, we will sketch relevant technologies for multilingual information systems and address the problem of communication with human partners. We will show that there is no strict border between information and communication systems. On the contrary, we will demonstrate that the combination of information and communication services in a seamlessly integrated system yields a much more natural and ergonomic application than the separation of the two types of services.

We will show how the services need to be adapted to the expected situations of usage and outline the ramifications for the user interface design.

A central message of our paper concerns the observed lack of robustness and reliability of some advanced language technologies. We will argue for combinations of technologies that achieve the needed reliability by often circumventing some fragile technologies such as speech recognition and machine translation.

2 Multilingual and Translingual Language Technologies

Human language technologies are technologies that are specialized for the generation or processing of natural language. Such technologies usually employ our knowledge about the special properties of human languages for achieving the desired performance. Some of these technologies are especially designed for written languages, while others are suited for dealing with speech. A few technologies are even specialized for sign languages. Typical examples of language technologies are: speech recognition, speech synthesis, natural language text analysis, natural language text generation and machine translation.

Multilingual technologies or applications can deal with more than one language. This is different from technologies that are equally suited for any given language such as certain text compression algorithms. A *multilingual* technology embodies knowledge about several languages. A multilingual open domain question answering system, for instance, can process questions in several languages. It will find the answers in the selected language of the query if such answer can be found in the text base. *Translingual*, also called *crosslingual* technologies involve at least two languages. These technologies support the user in crossing a language boundary. Crosslingual open domain question answering, to stick with our example, would (also) search in documents that are written in a language different from the language used in the question.

The most demanding crosslingual language technology is high quality and fast automatic machine translation. If such an application already existed, all other crosslingual technologies could easily be realized on top of machine translation. If someone used a question answering system for querying texts in a foreign language, the question would simply be translated for a monolingual QA system and the found answers could accordingly be translated in the opposite direction.

We can now easily define multilingual information systems as systems that provide access to digital information in more than one language. Crosslingual information systems provide access to information coded in one or several languages that are different from the language of the user. A translingual communication system, finally, is a system that helps users to communicate with partners who do not master their language.

The ultimate crosslingual communication service could be a speech-to-speech translation system such as the research prototype Verbmobil [8]. However, at this time such a system could reliably only work for a very limited domain and vocabulary. Since today's speech recognition technology is not tolerant with respect to background noise, a speech translation system could not be trusted in many real life situations.

Since we cannot cover the variety of possible and realized systems in this paper, we will discuss the relevant issues on the basis of two typical classes of applications: translingual web-based information systems and integrated translingual mobile information and communication services. For both classes we will draw on examples from our own research.

2.1 Translingual Web-based Information Systems

As explained above, the major goals of the translingual information systems are to help users to access information in a foreign language with a query in their own language. Three general strategies are applied in most of these systems: query translation, online document translation and offline document translation and indexing. A survey of advantages and disadvantages of these strategies is given in [9]. Query translation converts the query string in one language to another. However, because of the lack of context knowledge, there can be various translation options for a short query. Some of them do not belong to the intended meanings. Therefore, the biggest problem for a query translation is the problem of ambiguity leading to incorrect or unwanted information. Google provides online document translation, which helps users to translate documents written in a foreign language. However, it does not help in the search proper. The third strategy is offline translation. The entire document pool is translated into the user language and thus crosslingual search is reduced to monolingual search in the translated documents. The performance of the two methods strongly relies on the availability of machine translation systems for the relevant language pairs and on the quality of these systems. In [9], a general framework was presented, that permits uniform and multilingual access to heterogeneous data sources, e.g., web pages and database contents, for an application domain. Within this framework, it is easy to build up a multilingual information service system for relevant domains such as weather, public traffic, sports information, travel, restaurants, hotels, etc.

The system permits queries and search via cascaded pop-up menues. If the query is selected from menus or if a free-text query can be mapped onto a query to the database, the relevant DB records are retrieved and transformed into natural language sentences of the language used in the query. For this transformation, a special language technology is applied that is termed template-based multilingual generation [2].

The framework also integrates cross-lingual retrieval with information extraction. If some types of information are not maintained in a central database but can be collected from various websites such as dynamic properties of tourism sights and services, these information pieces can be extracted from the websites and fed into the database. These records can then also be accessed and viewed in multiple languages with help of the multilingual generation technology.

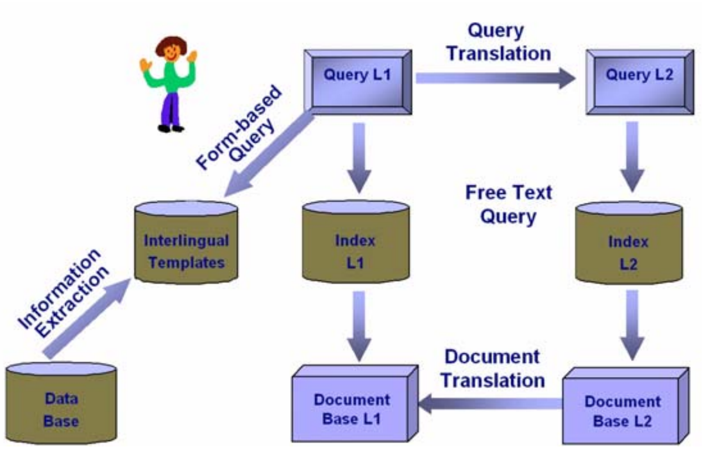

Figure 1. A general crosslingual information system framework for various data resources

The important lesson, we learned concerns the role of translation. Since machine translation is still far from being reliable, we need to design methods to circumvent this brittle technology as much as we can. In the case of crosslingual information access, the solution is a clever and carefully tuned combination of database access allowing multilingual input and output with the automatic translation of queries and documents outside the database.

2.2 Translingual Mobile Integrated Information and Communication Services

In 2.1 we described various crosslingual strategies which ease the users to access information unavailable in their own languages. In this section, we will discuss the technologies which support users to communicate with people who do not share their own languages by integrating information and communication services.

When people travel to places that they do not know well, they experience an increased demand for information on facilities, locations, events, services, goods, food etc. When these strange places are located in regions where their own language is not commonly understood, a language barrier blocks the access to the sources of the strongly needed information. Assume an Irish tourist in China. The tourist wants to see as much as possible during his stay and therefore needs information on sights, opening hours, transportation, cultural events, restaurants and shopping. But the web-based city map services are presented in Chinese and so are many other information sources. But also most people around that could be asked for information, only speak Chinese. Even ordering food in a restaurant and telling a taxi driver a destination can result in frustrating miscommunication.

These are situations in which the tourist would be thrilled if he had a mobile translation device. This application has received much attention recently because several large projects have worked on the required technologies (Verbmobil, Communicator, TC Star).

The largest single project in the area was the German effort Verbmobil [8] with more than twenty partners in industry and research, coordinated by DFKI. Although the successful project resulted in a strong push of face-to-face speech translation, the base technologies could not be lifted up to the level of coverage and robustness required for a product in the demanding tourist market.

The two critical technologies are machine translation and speech recognition. The combination of these two fragile technologies yields application prototypes that are still far too unreliable for use in real life. Machine translation cannot handle partial sentences or slightly ungrammatical input as it is abundant in spontaneous speech. But even completely grammatical utterances may not be correctly analyzed by the automatic speech recognition. As soon as the utterances are produced in a noisy environment, the analysis becomes even worse. Existing speech recognition systems do not yet possess the faculty to filter out irrelevant background noise from the speech signal as human acoustic perception does. Partial analyses of speech input are highly ambiguous. Human listeners can disambiguate by utilizing world knowledge and achieve a good understanding of the situation due to their capabilities of semantic inferencing. The machine translation systems often cannot even properly resolve the inevitable ambiguities found in complete and correct analyses.

Several courageous avant-garde products for speech translation are on the market for the customer category of early adopters such as assistant tools for call centers but none of these products are reliable enough to serve tourists safely in the variety of real communication situations.

One obvious solution would be to decouple the unreliable speech input analyses from the machine translation technology. Since most handheld computers today feature a virtual keyboard, we could simply rely on typed input. Some PDAs and smart phones are even equipped with a miniature QWERTY keyboard. This solution has the disadvantage that entire sentences need to be typed in during the dialogue. Such waiting times can be highly disruptive or annoying, e.g., if other customers are waiting in line or when the taxi driver has taken the wrong turn. The machine translation system is not going to correct typos so any input error will lead to a translation failure by returning no translation or an incorrect one.

Written output, on the other hand, is usually sufficient although there are situations when spoken output is preferred. Such situations include, for instance, the communication with taxi drivers and with waiters in badly illuminated bars and restaurants. In certain other target countries one may also be encountered with the effects of illiteracy.

A serious problem for the typed-input solution is also the handling of responses from the local communication partner. Sales assistants, cab drivers, waiters would have to use a keyboard on a device they are not familiar with. Moreover, in the case of Chinese responses there is the problem of the different methods for Chinese character input. The local partner may not be used to the input method set as default and a negotiation about the appropriate method could be difficult or at least rather disruptive in a real communication situation.

In our own work, we have experimented with a number of combinations for input and output methods and with a combination of machine translation technologies. However, since the number of possible combinations is far too large to try them all, we have utilized experience and results from HCI research plus some common sense thinking to arrive at a combination of interface methods that seems optimal for the given task and for the existing state of hardware, interface software and translation technologies.

In the following we will describe this combination. We will then explain why and how this combination of technologies for translingual communication was extended by multilingual information services.

In order to explain our strategy for coping with imperfect translation technology by a novel combination of translation and input methods, we will first discuss some methods for improving the quality of automatic translation.

One method is to restrict the language input to a sublanguage that the machine translation can reliably handle. This approach is called "controlled language". It is applied in corporate document production, where the professional authors can be instructed to follow certain rules defining the corporate language. Such corporate language does not only support automatic translation but also facilitates comprehension by human addressees. Therefore, it is applied in safety and security critical areas such as aircraft maintenance or military communication.

Another method for improving the reliability of machine translation is the combination of example-based techniques with rule-based translation. As the name suggests, example-based machine translation exploits examples found in human translations. Some systems such as the English-Chinese Huajian translation software integrate example-based translation into a classical rule-based system. Whenever a translation candidate is found in the example base, preference is given to this option, otherwise the system translates by using dictionaries and rules.

A third approach for improving the accuracy of automatic translation is to utilize formalized knowledge about the relevant domains. This approach is based on the observation that human translators have to know the subject domain of a text in order to deliver a high quality rendering of terms and formulations in the target language. The proper translation of terminology and subject-specific formulae, for instance, depends on their exact context.

In our framework which is developed within the Sino-German cooperation project COMPASS2008, we try to combine insights from these approaches in a radical and very simplified way. The application scenario is the 2008 Olympic Games in Beijing, a highly demanding setting for a cross-lingual application because of the heterogeneous nature of the information and the involved languages. First of all, we utilize a phrase-book-like large collection of correctly translated utterances ordered by subject domains. These utterances are assigned to situation types in order to further reduce mistranslations and to facilitate their effective selection in a given situation. Only if some input cannot be matched to any of the pre-translated utterances, we call a free translation system. But before we call this system, we check for proper names of places, people, dishes and special terminology in order to prevent the translation of names such as Bush, Gates, or Madonna and to select the proper translation of special terms such as "track and field", "finals", "goal", etc.

Although our system permits spoken input, we very well know that background noise drastically reduces the recognition accuracy. As we expect that visitors and participants of the Olympic Games will not use the system in quiet places very often, we are not counting on the usefulness of the speech input channel. We may even decide to deactivate this option for the specific application. We will permit typed input through a real or a virtual keyboard, but we will try to limit the need for dealing with the small keyboard to cases, where some input could not have been foreseen in the design phase.

We have tried to model as many situations as possible in which the demand for information and communication arises. Such a situation can be a taxi ride. A taxi ride again consists of sub-situations such as boarding the taxi, the ride itself and the end of the ride when the fare is paid. Each of these situations consists of numerous possible communication steps. If in each of these moments the users can be offered the next intended communication act as part of a short menu, then there is no need to type in a sentence. Even a selection in five steps using cascaded menus will be faster than the typed input of an entire sentence. If every menu only consists of five choices, a five-step selection already gives you 3125 options. However, our menus are organized in such a way that most selections only require one or two clicks.

Finally we are going to discuss the integration of communication and information services. As we already stated above, the foremost goal of the user is not to access information services or to seek communication with a local waiter.[2] Typical goals of foreign tourists are to visit sights, to get food, to buy goods, to attend some

[2] This does not mean that we underestimate or neglect the tourists' interest in communication with local citizens. We simply observe that in most situations, communication serves some practical goals.

entertainment events, or to receive assistance in emergency situations. To achieve these goals, users may need to consult information services. They may also need to communicate with local service personnel or with people in the street. A typical example is the tourist's goal to obtain some tasty meal. She may first consult a restaurant information service to find suitable restaurants in the vicinity. She may then want to find out which restaurants offer her preferred type of food in the appropriate price range. Next she may want to get directions on how to get to the selected restaurant. Getting there may require a taxi ride. Once arrived, she may want to find out exactly which dishes are offered. If these dishes are unknown to her, she might be interested in learning about the ingredients and the mode of preparation, either to gather the facts for an optimal final selection of a menu or simply to avoid certain ingredients she dislikes or is allergic against. Once the decision is made, the selection needs to be communicated to the waiter. After the meal, another round of communication is needed for getting and settling the bill.

Such situation-driven combinations of information and communication services form the core of the COMPASS2008 system. The design has been driven by a careful analysis of a variety of use cases. Much effort has been invested into the optimal combination of input and output modalities suited for the supported situation types. The best choice of input modalities for the responses solicited from local Chinese people was especially demanding. Usually these communication partners do not have any previous experience with our system. Instructional dialogues and help systems are much too time consuming and disruptive for the intended usage. Thus for each type of solicited response, we needed to find a way of presenting the Chinese partner with a selection of choices or some input method that is intuitive and very easy to handle.

For instance, for numbers such as prices and distances, the system presents the user with a virtual numeric keypad that also displays the appropriate unit of measurement. Other GUI templates are used for dates and times, for binary choices or for menu selections.

Figure 2. COMPASS2008 User Interface Examples

In July 2006 the COMPASS2008 system was successfully tested by tourists from several countries in a field test in Beijing under realistic conditions. Equipped with the new software on a palmtop computer the test subjects faced typical challenges for a nonnative visitor. After the test they all praised the system. Without speaking a word of Chinese themselves, they could give directions to cab drivers, search for sights, order food and ask for directions. The field test consists of a list of tasks associated with the special contexts, e.g., restaurant and shopping. A systematic evaluation and assessment of the usability and acceptability of the system functionalities have been worked out taking the filled questionnaires after each task and the whole test as input material. The overall evaluation result is very satisfying and promising [7]. The evaluation results can summarized as follows:

- Task completion is 100% for all tasks. With most tasks, 10 to 20% the users reported little usability problems.
- On the average of 60% subjects did not feel that it takes too much time to complete it. The percentage of subjects, who complain about a long task duration is quite high (40%) in some tasks.
- Most subjects (73%) felt well orientated while solving the different tasks.
- The main factors for the acceptability of the COMPASS2008 services are rated positive by 62%. Still there is potential for optimization. In particular, the controllability and completeness of the service functionality should be improved in order to increase user satisfaction.
- All COMPASS2008 services are judged to be useful or rather useful by more than 80 Percent of the subjects.
- The percentage of users willing to use COMPASS2008 is a above 50-80 % for almost all services, except “Currency Converter”
- There are significant differences between the test groups. Native Chinese and native English tend to rate the services as more useful and show a greater willingness to use COMPASS2008 services than non native English.

3 Conclusion and Outlook

We have presented an HCI-oriented approach to translingual information and communication services. A novel situation-driven combination of various language technologies reduces both the dependence on less reliable technologies and the required user effort. The strategy of circumventing less reliable technologies in the great majority of use cases was followed in the design of both information and communication services. However, these technologies, i.e., open text search, machine translatiion and speech recognition are still part of the architecture and can be called on if needed or when technology matures.

A central property of the system is the seamless integration of information and communication functionalities. Basis for this integration is the situation-driven approach again. A future extension already foreseen in the strategy and architecture of the system is the addition of transaction services. By transaction services we mean services other than information and communication, e.g., the purchase of tickets, the registration for a conference or the reservation of a table in a restaurant. Following our approach, these services would also be offered depending on usage situations by the same mechanisms that trigger the choice of information and communication services.

Acknowledgments. The research reported here has been partly conducted in the EU-funded projects MULINEX, MIETTA, and also the Sino-German cooperation project COMPASS2008, funded by the German Ministry for Education and Research grant no. 01IMD02A, 01IMD02B and 01IMD02C, and by the Chinese Ministry of Science and Technology. Special thanks to German Telekom for their co-funding for COMPASS2008 too.

References

1. Aslan, I., Xu, F, Uszkoreit, H., Krüger, A., Steffen, J.: COMPASS2008: Multimodal, Multilingual and Crosslingual Interaction for Mobile Tourist Guide Applications. In: INTETAIN 2005, Italy, 2005.
2. Busemann, S. , Horacek, H.: A Flexible Shallow Approach to Text Generation. In: Eduard Hovy: 9th INLG '98, August 5-7 pp. 238-247, Niagara-on-the-Lake, Ontario, Canada, 1998
3. Uszkoreit, H.: Cross-Lingual Information Retrieval: From Naive Concepts to Realistic Applications. In Proceedings of Language Technology in Multimedia Information Retrieval, Proceedings of the14th Twente Workshop on Language Technology.
4. Uszkoreit, H., Xu, F.: Modern Multilingual and Crosslingual Information Access Technologies. In Proceedings of Multilingual Information Service System for the Beijing 2008 Olympics Forum. The 7th China Beijing International High-Tech Expo (CHITEC), May 21 2004, Beijing, China.
5. Uszkoreit, H., Xu, F., Steffen, J., Aslan, I.: The pragmatic combination of different cross-lingual resources for multilingual information services. In Proc. of LREC 2006, Genova, Italy, 2006.
6. Uszkoreit, H, Xu, F., Aslan, I., Steffen, J.: COMPASS2008: An Intelligent Multilingual and Multimodal Mobile Information Service System for Beijing Olympic Games. In Proceedings of KI2006 Demo Collection, Germany, 2006.
7. Uszkoreit, H., Xu, F., Liu, W., Steffen, J., Aslan, I., Liu, J., Müller, C., Holtkamp, B., Wojciechowski, M. A Successful Field Test of a Mobile and Multilingual Information Service System COMPASS2008. To appear in Proc. of HCII 2007.
8. Wahlster, W. (ed.): Verbmobil: Foundations of Speech-to-Speech Translation. Springer-Verlag. Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokyo, 2000.
9. Xu, F.: Multilingual WWW --- Modern Multilingual and Cross-lingual Information AccessTechnologies. In Knowledge-Based Information Retrieval and Filtering from the Web. Witold Abramowicz (Ed.), Kluwer Academic Publishers, page 165--184, 2003.