Automatic language translation for mobile SMS

S. K. Samanta, A. Achilleos, S. R. F. Moiron, J. Woods and M. Ghanbari

School of Computer Science and Electronic Engineering
University of Essex, Colchester, CO4 3NU, UK
ssksam@essex.ac.uk

***Abstract*— Currently mobile users engage the services of a third party provider to translate an SMS text into a different language. The existing services have a number of drawbacks e.g. high cost to the user, not user friendly, they reduce the message space, and are inefficient. To communicate with a foreign person the sender must know the recipients preferred language and device display capability. What is needed is a service where a sender can send message in their native language without regard for the target tongue. Using an open source language translation package and a database server we show that a mobile operator can provide a transparent service where the text message is automatically converted to the recipients preferred language. We show that in comparison to the existing system our implementation is efficient and cost effective and has large implications for commerce, language learning and person-to-person communication.**

I. INTRODUCTION

The need for language translation began when people started travelling from place to place and the first documented cases are between the 3rd and 1st centuries b.c. In modern times, telecommunications have allowed global interaction without the need to travel at all. Currently more than half of world's population have a mobile phone and access to the global facilities. In most countries people use messaging services (e.g. short message and e-mail) more than the conventional voice calls. The mobile short messaging service (SMS) provides a convenient platform where the message can be delivered even if the recipient's phone is engaged in voice communication.

Clearly not all people understand the same language and a textual message in the wrong language may be meaningless. Some messages are generated automatically and generally sent in one language, so many of the recipients may not understand. What is needed is a service that performs language translation according to the preferred language of the recipient, without them even knowing.

The literature has focused on SMS language translation; where the service is implemented either in the mobile device [1] or in the network [4] [7].

For translation in mobile devices an appropriate software interface such as Java Micro Edition (J2ME) is needed. This generally limits the number of languages which can be translated. Therefore devices with larger memory and high processing capability are required (e.g. phones with Symbian or Windows Mobile operating systems) which tend to be costly and can be a barrier to some people. Mobile users can avail the translation service without a costly handset if the translation is performed in the network; this is the focus of this paper.

Currently network SMS language translation is done at a centralized server usually belonging to a third party service provider. Mobile users who want to use the SMS translation service indicate the source and target language (e.g. Hindi to English) along with the actual text and then send the message to the service provider as an SMS message. After receiving the message the service provider translates the message and sends it back to the sender. This translation service is used for learning foreign languages and person-to-person communication; where the sender resends the translated message to the recipient.

The current implementation has a number of drawbacks and makes it difficult to deliver the messages which are automatically generated from applications (e.g. mobile commerce). In the current implementation the sender (e.g. mobile user) must have prior knowledge about the recipient's preferred language and or language display capability of the recipient's mobile phone. This increases the complexity if application generated messages are to be delivered in the recipient's preferred language. Using an open source language translation package and a database server we show that a mobile operator can provide a transparent service where the text messages (both from mobile users and application providers) are automatically delivered in the recipient's preferred language.

II. BACK GROUND

In a mobile network an SMS is delivered using a store and forward mechanism as shown in Fig. 1. Any device capable of sending and receiving a short textual message is designated as a Short Message Entity (SME). A mobile phone or an application server works as an SME. When a mobile user sends a text message, the network first transfers the message to a Short Message Service Centre (SMSC) which stores it for delivery to the recipient. In a similar way, application providers deliver value-added

content (e.g. ring tone, news and weather reports) to the mobile users. The header of the message contains information such as the sender and recipient's phone number. The SMSC stores the message and after collecting the routing information and status of the recipient's handset (e.g. eligible to receive SMS), it forwards the message to the mobile network which delivers to the recipient [2].

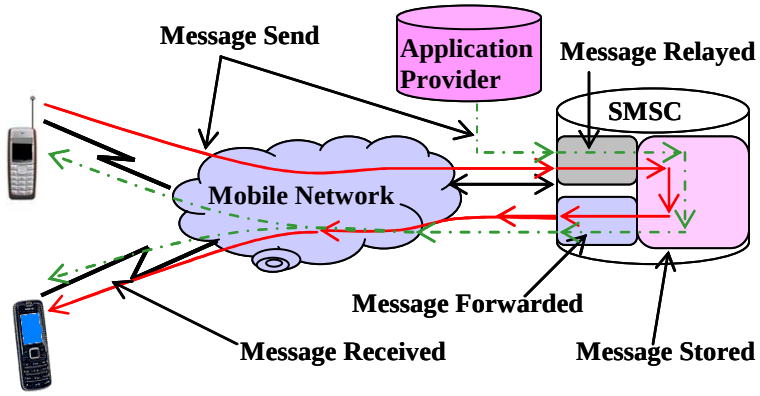

Fig. 1 SMS delivery in a mobile network

In the process of delivery from the sender to the recipient the message content (i.e. text) transparently passes through the mobile network and the SMSC. If two people speak the same language (e.g. English) then they can exchange textual messages in their native language. If one speaks English and other German then there are two ways they could communicate: 1) one could have a phone which converts from one language to another, and 2) one of them could employ a third party translator. If the conversion is done on the phone then either the sender translates (e.g. English to German) before sending, or the text is sent in the sender's native language and the receiver performs the translation. In either case the device must have the capability to translate many languages if it is to be truly international.

To translate a text message on the mobile, users needs to supply the source and target languages. This is not difficult for a sender if we assume that he/she knows the recipient's language preference. However the receivers face additional difficulty since they need to identify the source language from the received text. The recipient may not be able to identify the language not being a linguistic expert. Therefore the phone needs intelligent software to automatically detect the language and display the text in the user's native tongue. This will increase the cost of the phone which may be prohibitive to some people. In the absence of a high end handset either the sender or the recipient can engage a third party (who makes available a centralized server in the network) to provide the translation service [4] [7].

III. Network SMS Translation

Currently third party service providers perform language translation at a centralized server located outside the mobile network. Mobile users connect to the provider through the network and SMSC as shown in Fig. 2. The service is used for learning foreign languages and for person-to-person communication. Users, who learn foreign languages, indicate the source and target languages (e.g. English and German) along with the word or the sentence they wish to convert, this is then sent to the translation service provider as an SMS message. The service provider translates the message and sends it back to the sender.

For person-to-person communication (between people who speak different languages) the service can be used either by the sender or the recipient. If the sender wishes to use the service they send the message indicating their language and the recipient's language in a similar way to the language learning. The service provider translates and sends the message back to the sender who resends the translated message to the recipient. If the sent message is in the native language of the sender, then the recipient can use the service to translate to their preferred language. It should be noted that If the sender does not perform the translation the recipient faces the problem of recognising the source language.

Third party translation has a number of drawbacks; it is not user friendly as users need to indicate the source and target language along with the actual text for each translation. The mechanism is not efficient since it reduces the space available for the actual message and increases the traffic in SMSC and mobile network. For person-to-person communication there is a three fold increase in SMSC traffic and a two fold increase in network traffic.

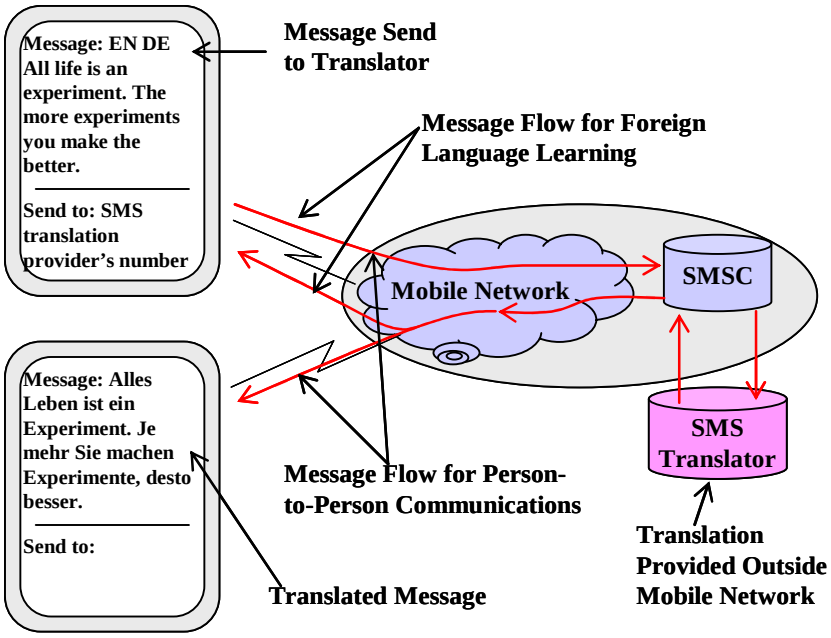

Fig. 2 Current implementation for SMS language translation

In an attempt to reduce traffic some implementations allow the sender to supply the recipient's number along with the source and target language within the text. This in turn further reduces the available character space and therefore short codes for source and target languages (e.g. ED for English to German or two digit numeric codes) are used. The need for short codes increases the complexity and the system becomes less user friendly with the need to remember language code pairs. As a result of including the recipient's number, the mobile network traffic does not increase, but the SMSC traffic still increases two fold (shown in Fig. 3).

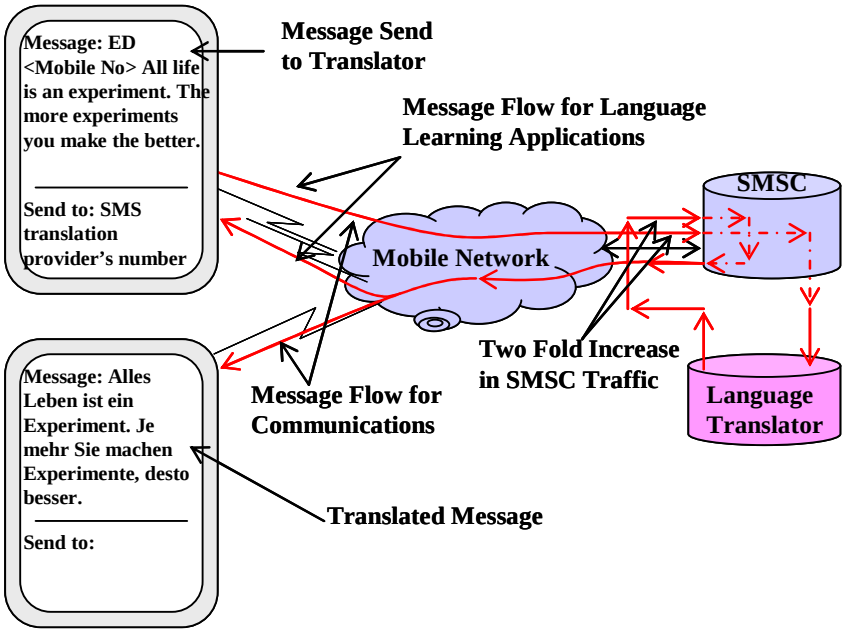

Fig. 3 Increase SMSC traffic in current SMS language translation

Increased SMSC traffic leads to higher cost for the mobile operator and the accompanying higher service charges. The main drawbacks of network based SMS translation services are listed below:

1) High cost to the user; users need to pay a premium fee for each translation request (e.g. Lingtastic LLC, one such service provider charges $1.00 per SMS).The cost can be prohibitively high and therefore can create a barrier for the growth of the translation service.

2) Not user friendly; users need to indicate the source and target languages for each translation request. Also users need to remember the set of codes and rules applicable for language learning and for communication.

3) Reduction of character space; less characters are available for the actual message since the source/ target language, and the recipient's number are included in the text.

4) Inefficient; the network traffic (i.e. traffic on SMSC) is increased two fold.

Under the existing service the sender must have prior knowledge of the recipient's language and/or language display capability of the recipient's mobile phone. As an example a sender may know the recipient's language but may not know whether the phone can display the message. This is a problem for automatically generated messages from application servers such as news, weather reports, security information and notification. Automatic messages are generally sent in a single language (e.g. English). This creates a large barrier since not everyone may understand that language.

What is needed is a service which can deliver the message to the individual recipient in his/her language of preference. If this were the case, the application provider could communicate with any individual regardless of their native tongue. If the provider wished to perform the translation themselves at the source they would need a database containing the language preferences for all users. This would significantly increase the complexity to the sending mechanism. Currently automatically generated messages are sent in batches. The message is composed once and software retrieves the mobile numbers of the recipients from a database. In order to send a message in recipient's language each message has to be sent separately after translation. This increases the cost (i.e. cost for additional software and hardware) for the application provider.

There are other problems in performing translation individually at the application's source. Application providers need to collect the language preference of each individual, but people may not be willing to share their information with every application provider (e.g. due to privacy). In contrast they will have no hesitation (or choice) in providing the same to their mobile operator who already has other information such as sex, age, spending capabilities etc. Therefore it is appropriate for the mobile operator to perform the message translation and deliver it to the recipient in their preferred language [9].

We show that the operator can perform the SMS language translation automatically by adding a standard language translation software package and a database in the network. By adding an Automatic SMS Language Translator (ASLT) in between the SMSC and the mobile network as shown in Fig. 4 we demonstrate that a sender can send the message in their preferred language without needing to consider the language preferences of the recipient. The ASLT automatically translates the SMS to the recipient's preferred language before delivery. The mobile user needs to once only notify the mobile operator as to the language of preference. The message content will pass through the translation process only if the recipient has subscribed to the service, otherwise it will be transparently forwarded. In the event the translation is required, the content language is identified automatically and translated as per the language preference. Our implementation solves many of the current problems and is laid out in the following.

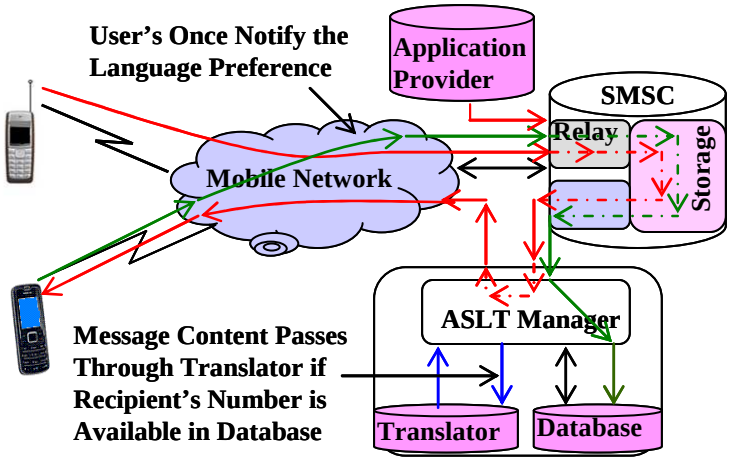

Fig. 4 Implementation of ASLT in real network

IV. IMPLEMENTATION

Using the Google Translate java API (an open source language translation package) and a database server we implement the functionality of ASLT. Our implementation allows message communication in the user's native language and also helps users in language teaching. The implementation is applicable to any messaging service (e.g. SMS, MMS and e-mail) but here we emphasis on SMS. It is assumed that the mobile handsets have the capabilities to input and display the preferred language of the user.

In Fig. 4 we show the architecture of ASLT where a box (containing language translation software and a database server) is inserted in the SMS delivery path from SMSC to mobile users. Mobile users who want to receive their communications in their preferred (e.g. native) language indicate their preferences once only by sending an SMS which is stored in the database of the ASLT. When an SMS is received (from a mobile user or an application server), the ASLT manager searches the recipient's number in the database. If the entry is not found, the message is transparently delivered. If the number is available in the database, the language preference along with the text content is forwarded to the translator. The translator automatically detects the source language and performs the translation. The translated content is forwarded to the mobile network for ultimate delivery to the recipient. In addition to being user friendly, Fig. 4 indicates that in our implementation there is no change in SMSC traffic.

We simulated the functionality of the ASLT on a PC using: The Google Translate java API and our own software also in java. Using J2ME (Java Micro Edition) as the Short Message Entity (SME) we simulated the origin and reception of the SMS. In Fig. 5 we show the screen shoot of our simulation. A text message: "All life is an experiment. The more experiments you make the better." is sent to the recipient (number "8888888") whose language preference is German. The message is automatically translated to "Alles Leben ist ein Experiment. Je mehr Sie machen Experimente, desto besser." in German and delivered to the recipient's phone. The implementation in a real world network is relatively straightforward; it works as a relay box and therefore can work independently of other units. It can scale with the network since multiple threads allow translations in parallel.

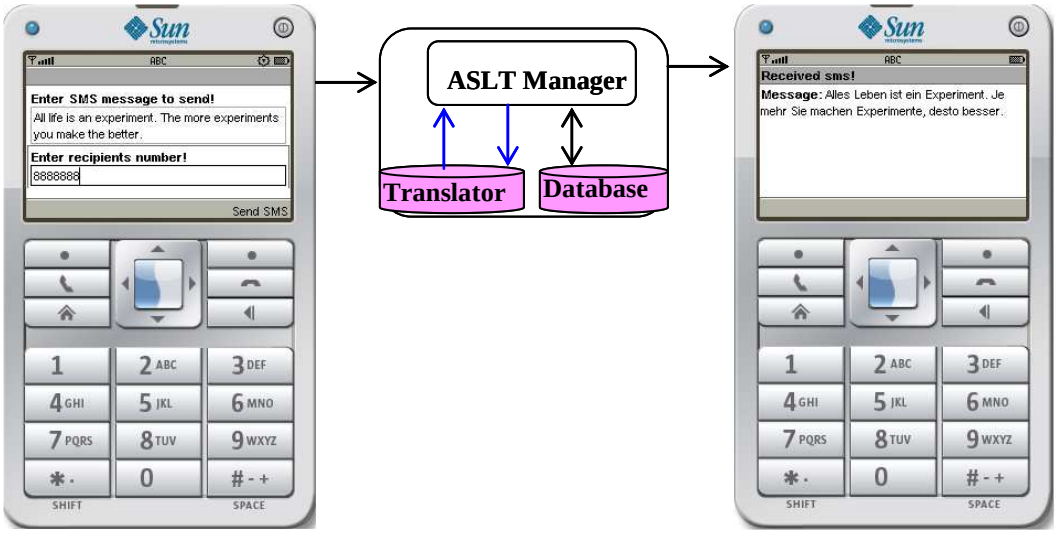

Fig. 5 Screen shoot of ASLT simulation

We acknowledge that there will be a slight delay in searching the recipient's number and to perform the translation. However the delay is reduced if the ASLT is integrated within the SMSC. The ASLT software package can be incorporated easily between the relay and the storage of the SMSC. We are currently engaged in a fully fledged implementation in a network test bed with separate hardware for SMSC and SMEs. A number of mobile operators in India have shown their interest in implementing our ASLT within their real network.

Drawbacks:
There are obvious drawbacks of automatic translation since there are difficulties with abbreviations and euphemisms (e.g. u instead of you) frequently used in SMS messages. Research on SMS texting for 11 European Union countries indicates that even though abbreviations such as “lol,” “u,” “brb,” and “gr8,” are frequently used they make up only 10% of the total words [5]. Therefore abbreviations may not pose as big limitation to SMS translation as first apparent. Using artificial intelligence and neural network technology [6] it is possible to improve translation accuracy. Some developers [7] currently implement commercial products which translate abbreviations from Chinese to English and vice versa.

V. Benefits From Automatic SMS Language Translation

In addition to person-to-person messaging, a large number of applications such as tourism, international business communication, foreign language learning and Business to Consumer (B2C) communications will benefit from our work.

In our implementation, the sender inputs the message in their preferred language (e.g. English) and simply sends the message to the recipient (see Fig. 5). There is no need to indicate the source and target language and no requirement to know the preferred language and display capability of the recipient; the recipient automatically receives the message in their preferred language (e.g. German). The communication is seamless, user friendly and unobtrusive. Our implementation is efficient since there is no reduction of character space for the actual message and no increase in the SMSC traffic. Mobile operators themselves can provide the translation service using our relay box without the need for a third party service provider. There will be additional financial cost for the translation software and database server but the cost can be recovered from additional use generated by applications such as language learning and B2C communications.

Mobile SMS provides potential opportunities for learning foreign languages [8]. In our implementation language learning is user friendly since learners are freed from the burden of selecting the source and target language for each translation. A mobile user first updates the language preferences such as German if they want to learn the German language. In the same way as a normal SMS messaging, a user inputs a word or a sentence in their native language and sends it to their own mobile number. Message is automatically translated and delivered to the sender.

For B2C communications such as: news, weather reports, and commerce applications, messages often come from application servers. A lot of services such as: health care management, advertisement, auctions, sale of goods and services [3] are currently provided using SMS. Our implementation will facilitate the message delivery from applications according to the preferred language of the recipient without them even knowing and without the need to change the necessary hardware and software in each individual application server. Application providers gain benefit from by connecting to a large population who are currently not accessible to them.

Currently more than half of world’s populations have a mobile phone and access to the global facilities. In developing countries people who do not have easy access to commerce applications (e.g. sale of goods through internet) can now do business using a mobile phone. As an example, similar to eBay people in Bangladesh buy and sell via mobile using CellBazzar [3]. CellBazzar allows people to obtain current market prices in order to buy and sell agricultural products, goods and services through SMS. The service is based on a simple searching mechanism; users only need to read the simple text in English (e.g. buy, sell and name of goods, product and services) and type numeric digits.

However, from current (as of April 2009) statistics about the number of goods and services for sale, it is clear that the CellBazzar has failed in enabling people to efficiently sell their agricultural products. One reason may be that most of the people engaged in agriculture cannot read English and need the help of a third party translator. Our implementation will allow the CellBazar and similar application providers to deliver the information to the peoples in their preferred language (e.g. Bengali) automatically and gain benefit from the additional usage.

VI. Conclusion

Using an open source language translation package and a database server we implemented automatic language translation for SMS text. We show that our mechanism will allow mobile operators to provide textual message delivery in a user’s native language without the need for a third party service provider. We show that in comparison to the third party provider our mechanism is user friendly and efficient. We demonstrate that using our platform a large number of services such as advertisement, auctions and sale of goods and services can be delivered to those who are currently not able to access them due to the language barrier.

REFERENCES

[1] Agrawal, A. J. and Chandak, M. B. 'Mobile Interface for Domain Specific Machine Translation Using Short Messaging Service', 4th International Conference on Information Technology, 2-4 April 2007, pp.957-958.

[2] Brown, J., Shipman, B. and Vetter, R. 'SMS: The Short Message Service', Computer, Vol. 40, No. 12, Dec. 2007, pp.106 – 110.

[3] CellBazzar. Available at: http://corp.cellbazaar.com/sms.html

[4] Chava, V., Smith M. R. and Dudley, W. H. 'System and method for in-transit SMS language translation', United States Patent No. 7272406.

[5] Crystal, D. 'Txtng: the Gr8 Db8', Oxford University Press, 2008.

[6] Khalilov, M., Fonollosa, J.A.R., Zamora-Martinez, F., Castro-Bleda, M.J. and Espaa-Boquera, S. 'Neural Network Language Models for Translation with Limited Data', 20th IEEE International Conference on Tools with Artificial Intelligence, 3-5 November 2008, Vol.2, pp.445-451, Dayton, OH.

[7] Moka LLC. 'Moka Partners with China Mobile for Mobile Chinese to English Language Translation and Language Learning', Available at: http://www.moka.com/en/news/news-1.htm.

[8] Saran, M. Cagiltay, K. and Seferoglu, G. 'Use of Mobile Phones in Language Learning: Developing Effective Instructional Materials', Fifth IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education, 23-26 March 2008, pp.39-43, Beijing, China.

[9] Rownok, T., Islam, Md. Z. and Khan, M. 'Bangla Text Input and Rendering Support for Short Message Service on Mobile Devices', Proc. of 9th International Conference on Computer and Information Technology, December 2006, Dhaka, Bangladesh.