BuzzTrans: Evaluating a Translating Instant Messenger Client

Rahul Nair, Kedar Shiroor
GVU Center, College of Computing
Georgia Institute of Technology
85 5th Street NW
Atlanta, GA 30332-0760
rahulnair@acm.org, kedarshiroor@gmail.com

Justin Godfrey
School of Psychology
Georgia Institute of Technolgy
654 Cherry Street
Atlanta, GA 30332-0170
gtg470s@mail.gatech.edu

ABSTRACT
The ability of instant messaging (IM) to instantly connect geographically disparate people in real-time communication is breaking many cultural barriers. In this paper we present BuzzTrans - an IM system that uses automatic machine translation to transparently connect users who do not speak the same language. The results of an evaluation are also presented along with a set of design recommendations for future work in this space.

Keywords
Instant messaging, chat, machine translation, CSCW, design recommendations, qualitative user study

INTRODUCTION
The popularity of Instant Messaging (IM) has been rising rapidly and it has reached an estimated 180 million [13] users worldwide. Surveys [2] have shown that an approximately 90% of American teens and young adults who use the internet are also IM users. This familiarity with IM makes it an ideal vehicle to introduce younger audiences to collaborative systems and groupware.

This proliferation of IM is corroborated by Grinter and Palen [11], who conducted an ethnographic study into the IM habits of teenagers and found that IM was a very important part of their daily lives. All their study participants reported using IM as an important collaborative resource to complete schoolwork. One of the participants also reported using IM to practice foreign language skills by having conversations in a different language.

MOTIVATION
The primary motivation behind BuzzTrans was to build a collaboration support system for the Global Classroom Project [8] – a long-term, global, distance learning project. This project was designed to provide a forum for cross-cultural collaboration with the aim of producing digital artifacts. In its current form it is a class that is simultaneously taught to two groups of students; one in an American university and the other in a Russian university. A large proportion of the class is virtual and uses the internet to connect the Russian and American students into one cohesive group. To do this, they utilize WebBoard [1]; a commercial bulletin board style conferencing software which allows posting and reading of messages and files, similar to many other HTML–based user groups and the Usenet protocol.

As we studied the project, we found that there was a lack of real time communication between the Russian and American students. The static nature of a web-based newsgroup is not conducive to informal, interpersonal communication. With a few exceptions, users can read every message posted and any replies. Since professors can read the content, the students are more formal about their posting and think twice before posting any “fun” content. Also, the process of posting messages is quite lengthy when compared to other computer supported tasks—the user must first log onto WebBoard, find the thread they wish to post or reply to, compose and then send their message. Finally, the user must then wait for a variable length of time while reloading the webpage to check if their message was posted correctly.

Since most of the students are in the 16-22 age group, we felt that IM would be an ideal way to support their informal communication. IM would allow students to collaborate on an assignment, transfer files, have a quick chat, etc..., all which are somewhat out of place in the formal academic nature of WebBoard. However, the language barrier is somewhat of an obstacle to this idea, given that few, if any, of the American students speak Russian. On the other hand, Russian students tend to speak passable English and usually are quite fluent.

Further interviews with the professors who teach the course showed that while the Russian students could fluently read English, they have more difficulty in composing statements

and responses in English. It appeared to us that creating a translating IM client would benefit the students by freeing them from the language barrier, while still allowing them to engage in real time communication. Providing additional support to the existing practice of IM would lead to faster adoption when compared to introducing a completely new practice.

SYSTEM DESIGN

We designed BuzzTrans with the express intention of allowing users to transparently conduct an IM conversation with users in other languages. Our prototype was designed to use free software and open protocols so that we could test the feasibility of the translating IM client without any of the legal issues that accompany the use of a proprietary standard.

Infrastructure Components

Our goal while designing the system was to enable advanced translation services while maintaining interoperability with established Instant Messaging platforms. The following sections explain the various components of the BuzzTrans infrastructure.

Messaging Protocol

We decided to use the Jabber XML protocol [14] as the basis for our project. The secure, extensible and decentralized architecture of the Jabber network along with the ready availability of freeware servers would allow the universities to set up their own jabber server to provide private communication between students. While Jabber has a much smaller installed base than popular commercial IM services such as Yahoo Messenger, MSN Messenger and AOL Instant Messenger, the ability to create gateways on the jabber server enables users to communicate with the commercial networks with minimal effort.

Messaging Client

In order to reduce the development time of the project we used the TechJab jabber client [21] as the starting point for BuzzTrans. TechJab is a .Net based Jabber client that was designed to be used in Tech support environments. While it does not support all the features of the Jabber protocol, it is a very well written and stable client, making it an ideal starting point for BuzzTrans. The finished BuzzTrans messaging client consists of a user interface module, translation module and a communication module.

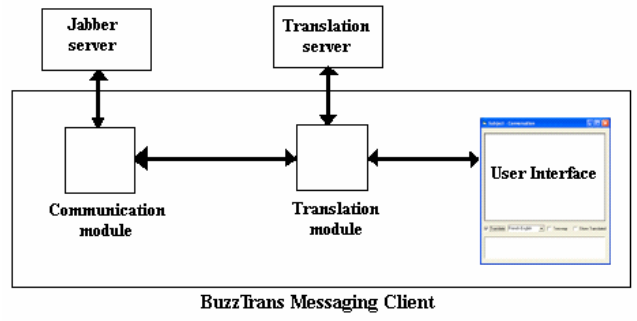

Figure 1: BuzzTrans system architecture

Translation Module

The primary modification that we made to the TechJab client was to add a translation module between the user interface and the underlying communication protocols. This module takes the input messages and sends them to the translation server to be translated into the appropriate language. Once the translation is complete it passes the message to communication layer and/or the user interface as appropriate.

The translation module is designed to use a multithreaded model with one thread for each active conversation. This allows the user to simultaneously have conversations with different contacts in different languages. Since the translation is accomplished within the BuzzTrans system, it is not necessary for the users' contacts to have any special software installed on their machine. In fact, they can use any Jabber compatible IM client.

Translation Server

The translation function of BuzzTrans relied on the free translation service provided by Systran Language Translation Technologies [20]. The decision to use a web based translation service was based on the following factors:

- Small footprint on the client machine.
- Ability to transparently upgrade the translation engine and database.
- Requires less processing power from client machine.
- The use of an IM service implies the presence of a reliable internet connection.

The Systran translation service is presented as a Common Gateway Interface (CGI) and can be accessed using a properly formatted HTTP POST operation which will return a web page with the translated text.

User Interaction

The user interface (Figure 2) of the BuzzTrans client is similar most IM clients with the addition of a few translation specific components. When the user is in the single language mode, the only new component in the chat window is the presence of a "Translate" checkbox. This checkbox is present in each individual chat window and allows the user to explicitly turn the translation feature on and off for each chat conversation.

If the user checks the Translate checkbox the chat window displays the translation options palette on the chat window. The palette is intentionally displayed in between the users' text entry screen and the chat display screen so that the users will always be aware of the status of translation. The translation palette gives the user the following options

Translation Language

The user can select the translation language using a list box which lists all the translation combinations that are

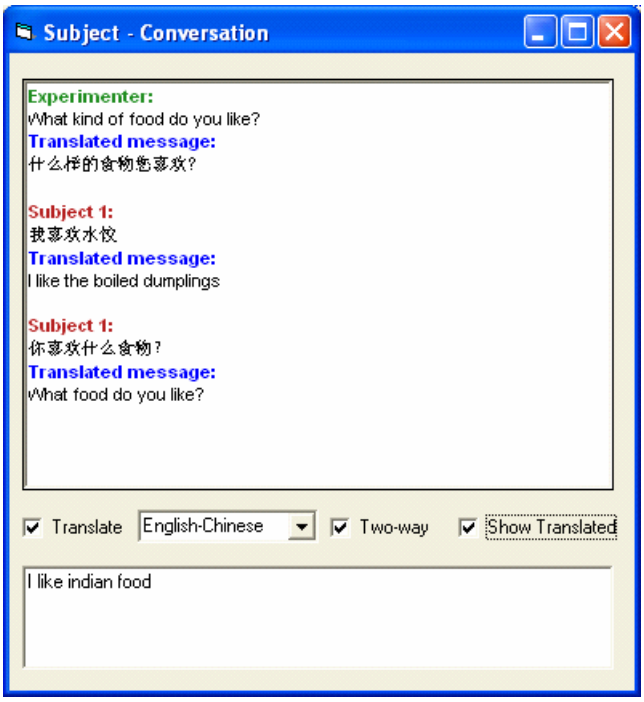

Figure 2: Screenshot of BuzzTrans user interface. The system is set to use English to Chinese, two-way translation while showing the translated text.

supported by BuzzTrans. The combinations are listed with the original language first and the translated language second. Once a language combination is selected it will be constantly visible on the translation palette.

Two-way Translation

The two-way translation check box allows users to selectively enable and disable the two way translation capabilities of BuzzTrans. When two-way translation is enabled, the users can carry out a conversation with another subject without any need of knowing the other persons language. The user need only type in his/her language while the remote users will use his/her own (different) language. If the two-way translation is disabled, BuzzTrans will only translate the local users' text while directly displaying the remote contacts messages with no translation

Show Translated Text

This checkbox allows the user to optionally view the both the original text and the translated text that is being sent. In the case of two way translation, the original text received from the contact is also displayed along with the translated text

Remote Users Screen

BuzzTrans was specifically designed such that users' contacts would not need any special software in order to talk to a BuzzTrans user. Since the translation is transparent and the contacts can use any jabber compatible IM client there is no visible difference between a BuzzTrans user and any other contact. As far as the contact is concerned, the BuzzTrans user can both understand and

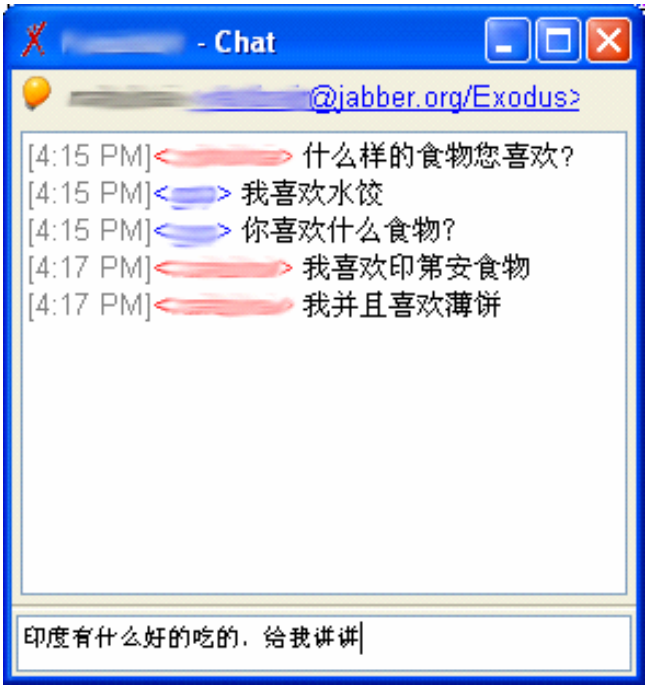

Figure 3: Screenshot of an Exodus IM client user in a Chinese-English conversation with a BuzzTrans user.

respond in the translated language. Figure 3 is a screenshot of a contacts screen while talking to a BuzzTrans user. The screenshot is taken while using a popular Jabber IM client called Exodus and is the other half of the conversation shown in figure 2.

STUDY DESIGN

Since we did not have physical access to the Russian students, we had to develop an alternative strategy to test the feasibility of BuzzTrans. To ensure that all subjects received exposure to similar content, we created a scenario where the subject would engage in a loosely scripted IM conversation with one of the experimenters. Subjects also filled out a questionnaire to assess demographics, language proficiency, and impressions of BuzzTrans. The following is a detailed explanation of our study design:

Participants

Participants in our evaluation were recruited from the campus of the American university. The only requirement for participation was that they were native speakers of a language (other than English) that BuzzTrans could support. Eight subjects (7 graduate students, 1 post doctorate) with an average age of 32 participated in the study. All of them were regular IM users and the average amount of daily computer use was 5.6 hours. Each participant considered themselves fully fluent (reading, writing, speaking) in their native language and on average, ranked their English skills as an 8 on a 10-point scale. In order to test the system without being overly dependant on the specifics of translating individual languages, we tested the system with six different languages: Korean, Portuguese, French (2 subjects), Chinese (2 subjects),

Russian, and Spanish. We specifically included two Asian languages in the study to test how the system would deal with the complexities of translating between completely different language families.

Method

The subjects conducted an IM conversation in their native tongue with an experimenter who would only use English. Care was taken to ensure that the experimenters did not speak the languages that they were testing. While the participants were told that the conversation was being translated into English, they were unable to see any of the translated text.

In order to control the content of the conversations, the experimenters were given a set of 26 questions that they would work into the conversation. The questions were designed to be common conversational topics that might be used while talking to someone for the first time. They included several context dependent and pop culture based questions to see if the basic meaning of the questions would translate correctly. Finally, the experimenters were instructed to be flexible and allow the conversation to flow in the direction that the subjects wanted to go.

Subjects were encouraged to continue the conversation in as realistic manner as possible, including asking the experimenter to rephrase if they did not understand the translated text. During the conversation, another researcher sat with the participant to log any mistakes in the translation and instances when the subject became confused or frustrated. The researchers also encouraged subjects to "think aloud" about the translated text and meaning of the received message.

Data collection

To facilitate review of the conversations, each participant's session was logged to a computer file. After each conversation, the log file was reviewed in conjunction with the experimenters' notes to identify areas of confusion and aspects of the system that needed improvement.

Participants were also given a post-task questionnaire to gather impressions about their experience using BuzzTrans. The data collected from the post-task questionnaire was complied and measures of central tendency and variance were calculated both individually and as a group.

RESULTS

After speaking to our subjects and going over the conversation logs we have come to the following conclusions:

Translation Errors are Inevitable

The translation system encountered several errors while trying to translate the messages. While some of these errors could be attributed to misspelling and slang in the input text the majority were due to the fact that the translation systems are not yet mature enough to perfectly translate between languages.

The errors we encountered could be split into three categories – grammatical errors, translation errors and translation failures. Grammatical errors are errors where the translated message has some grammatical inaccuracies but is still understandable. A phrase error is a more serious flaw in the translation that is usually caused by mistranslation of a word or phrase. A message with a phrase error is usually impossible to understand. Translation failures are messages that the system rejected because it was unable to translate them at all. If the system encountered a translation failure, it rejected the message and instead transmitted an error message saying "Unable to translate".

Users Can Communicate Despite Errors

Though the system introduced many errors into the translation we found that our subjects were able to work around the errors without too much trouble. In most cases they considered grammatical errors to be so insignificant as to ignore them. We believe that this is due to the informal nature of instant messaging conversation where users will often type in abbreviated text in order to save time. Since users do not expect to get grammatically perfect messages in an IM conversation they are not bothered by the grammatical errors. When the subjects encountered messages that they could not understand, they would usually ask the other person to repeat themselves. In our post task questionnaire we asked the users how easy it was to communicate using the BuzzTrans software. One a scale of 1 to 5 with 1 being very difficult and 5 being very easy the users gave it an average score of 4 with a standard deviation of 0.5.

Continuity is Key

The continuity of the conversations is critical. If users had a flowing conversation, they were able to use contextual information from the previous exchanges to compensate for errors in translation. Similarly users often found it difficult to follow a conversation if it suddenly changed track, this was exacerbated by persistent translation errors since subjects would think that the newest message was incorrectly translated instead of being about a different topic.

Users Are Forgiving of Machine Errors

Users are quite forgiving of translation errors because they do not expect a machine to be able to translate perfectly. In certain situations the translated messages are extremely formal but users are willing to overlook this since they do not have very high expectations.

```
Original Message:
How do you do?
Translated message:
Como você faz?
```

Excerpt 1: Translated into a colloquial question asking if the person was gay (English to Portuguese)

One example of phrase error is shown in excerpt 1 where the system translated the English statement "How do you do?" into a colloquial Portuguese for "Are you gay?" However since the subject knew that this was machine translated he looked further to guess its true intent and was able to answer correctly. The subject later said that if he had received the same message from a real person, he would have considered it to be quite offensive.

Users Adapt to Avoid Translation Errors

Once users learn about the strengths and weaknesses of the translation system they begin to adapt their conversational style to reduce errors. The most common sign of this is the avoidance of words that are consistently mistranslated. The use of shorter and simpler sentences is another indicator of adaptation. While there were noticeable changes in the users' communication style, they did not feel that the changes were very large. When asked if they had to modify their messages to improve translation (Likert Scale 1-5 with 1 being not at all and 5 being All the time) they gave an average score of 3.5 with a standard deviation of 0.5

Words Are Translated to Their Primary Meaning

A recurring translation error was the translation of "rock music" into "music of stones" or "boulder music" (excerpt 2). This particular example was noticed in almost all the languages and was intentionally given to all our subjects to observe their reaction. All of our users were able to understand the meaning of the statement and most were merely amused by it.

Original Message:

I like rock music.

Translated message:

Tengo gusto de la música de la roca.

Excerpt 2: Translation error while translating "rock music" from English to Spanish

While that particular example did not cause any problems, excerpt 3 shows the translation of the term "wearable computing" in Chinese to "may dress the computation" in English. Since wearable computing is a technical term that is not in common use, it was incorrectly translated into the equivalent words for clothing. This will make it especially difficult to carry out a technical conversation, since many technical terms are not included in dictionaries and may have different colloquial uses.

Subject:

可穿戴计算

Translated message:

May dress the computation

Excerpt 3: Translation error while translating from Chinese to English

Users Sometimes Want to Avoid Translation

On occasion, users wanted to send messages without any translation. This is most evident when names are translated into their literal meaning. An example is shown in excerpt 4 where the name "Howard Stern" is translated into "Howard rear of a ship". Since they could not control the translation features from their own client they (excerpt 5) tried to get around the translation system by enclosing the message in escape sequences that would cause the translation to fail.

Original Message:

do you listen to "the howard Stern" show on the radio?

Translated message:

¿usted escucha "la demostración severa del howard" en la radio?

Excerpt 4: Literal translation of the English name "Stern" into "ship stern" in Spanish

Experimenter:

Do you enjoy your academic major?

Translated message:

Você aprecía seu major academic?

Subject:

o que é <<major academic>>

Translated message:

what is << major academic > >

Excerpt 5: Using an escape sequence to demarcate a faulty translation (Portuguese to English)

Experimenter:

what type of computing are you doing?

Translated message:

您做着什么样计算?

Subject:

可穿戴计算，可穿戴means wearable

Translated message:

May dress the computation, may dress means wearable

Excerpt 6: Typing in the English term because the translation did not occur correctly (Chinese to English). Continued from excerpt 3

One subject got around the translation problem by typing the correct translation in English. This (excerpt 6) was a continuation of the conversation in excerpt 3. These workarounds were usually attempted because the users were attempting to explain a concept that the computer could not translate correctly.

Shared Context Enhances Understanding

The users understanding of a statement is greatly enhanced if both the participants have a shared context. If the subjects spoke about a topic that they both knew well there were fewer misunderstandings in the conversation.

Original Message:

What do you think about the new FCC indecency standards for radio broadcast?

Translated message:

你认为新FCC 猥亵标准怎么样为无线电广播?

Excerpt 7: Translation of the English word "indecency" into the Chinese word for "abuse"

Anecdotal evidence of this is show in excerpt 7 where the question "What do you think about the new FCC indecency standards for radio broadcast?" was translated into Chinese with the word "indecency" being translated into "abuse". This message referred to a controversial news story that was receiving extensive news coverage in the American media at the time. When the same message was sent to two Chinese subjects we noticed very different reactions. One was initially confused by the use of the word "abuse" but was able to guess the true meaning since he had been following the news story in the media. The other subject had not heard of the news story and had to ask the experimenter to clarify what he meant by abuse.

Translation Delays Can Cause Problems

Since we were using a free web-service to handle our translation requirements, the system response was heavily dependant on the network conditions at the time. Occasionally the system took several seconds to process, translate and transmit a message which the users found to be quite misleading. On several occasions the users would retype the message using different phrases thinking that the delayed response meant that the other person had not understood the last statement.

DESIGN RECOMMENDATIONS

Based on our experiences with BuzzTrans we created a set of design recommendation for researchers who wish to explore the same space:

Explicitly Mention Machine Translation

Any system that uses automatic machine translation should explicitly mention the fact to all users. This is necessary to avoid misunderstandings and awkward social situations (e.g. excerpt 1) that can be caused by incorrect translations. It is especially important to mention the automatic translation if the user is using the system to communicate with a regular user i.e. a user without any translation software.

Escape Sequence and Transliteration Support

The translation system should have an escape sequence that the users can use to explicitly demarcate sections of a message that should not be translated. In the case of languages with different alphabets, the system should support transliteration of the words between languages. This is especially important when using names and abbreviations.

Feedback during Translation

If the translation component of the system adds any delay to the message transmission, the system should give the user feedback about the state of translation. This will reduce user frustration with long delays between messages.

Optionally Display both Raw and Translated Text

The users should have the option of being able to view both the raw and translated text of each message. This is especially important if the users have at least a basic grasp of the other language. Users will be able to use the raw text to help understand the meaning of incorrect translations.

Support for Alternate Word Usage

One of the biggest problems that users were facing with the current BuzzTrans setup was the failure to translate alternate usages of a word. Future projects should include much greater support in this area. One possible solution would be to ask the user to disambiguate between possible usages of a word. The ideal system would learn from users' choices and automatically update its dictionary to suit the users' conversational style.

Support for Slang and Abbreviations

The system should provide support for slang and short form terminology that is commonly used in IM conversations (e.g., l8r for later). These abbreviations are primarily used to reduce the amount of time needed to type a single message. Since most of these are not accepted into the general practice they will not be included in standard translation dictionaries. This behavior of using abbreviations was also noticed by Grinter and Eldridge [10] in their study of the text messaging habits of teenagers.

RELATED WORK

Our work is not the first to attempt to use machine translation to help users communicate. There are several commercial and open source efforts that attempt to give much of the same functionality of our system. However we are the first conduct a formal study of the effectiveness of such translation technology.

Popvich et al. [17] created the ALTo machine translation system that translates the closed captioning information that is available in North American television broadcasts. Since the closed captioning text usually represents human speech, it has characteristics which are very similar to the context of IM messages. Despite this there has been no attempt to adapt their technology for use in a real time interactive chat system.

The WebDIPLOMAT system developed by Frederking et al [5, 6, 12] added speech recognition and machine translation capabilities to internet chat rooms. However, this system required that the end user have at least a rudimentary grasp of the language to select the best translation choices. There also appears to have been no formal user study of the system.

The GAIM [7] and AYTTM [3] multi-protocol IM clients support automatic translation via plugins that use the Babelfish [4, 22] translation service. While these plugins are quite popular they have never been studied in a controlled environment.

APPLICATIONS TO INDIA

There is a significant amount of research on how to bring computing to the Indian masses. A large proportion of this has been focused on creating

- low cost computing [19] and
- input/output support for Indian languages [15, 18]

While both of these are crucial to bring IT to the average Indian, the language barrier is an equally difficult obstacle. It is estimated that over 56% of all internet websites [16] are written in english and that over 35% of internet users [9] are native English speakers. In order for the power of the internet to be fully utilized by India, it will be crucial to make the content accessible in terms of language as well.

While BuzzTrans did not use any Indian languages as part of the evaluation, we believe that our results are generalizable and can be applied to Indian languages as well. A similar translating IM client with Indic languages support would allow Indians to easily communicate with the majority of internet users who do not speak any Indian languages.

CONCLUSION

In this paper we have presented BuzzTrans, an automatic translating IM client that allows users to transparently chat with people in a different language. We also present the results of a formal evaluation of the system along with a set of design recommendations for any future work in this space. It is clear from our study that while machine translation is not yet mature enough for wide deployment, the informality of IM conversations make it an attractive first step application for machine translation. Future work on the project includes a trial deployment of the system to new class of distance learning students to study the effects of long term usage.

ACKNOWLEDGMENTS

We would like to thank Dr. TyAnna Herrington and Dr. Yuri Tretyakov for their help and advice on this project. Our thanks also go to all the subjects who participated in our study. Finally we thank the TechJab development team and Systran Language Translation Technologies for the use of their software and technology.

REFERENCES

1. Akiva Technologies. http://www.webboard.com/
2. AOL Instant Messaging Trends Survey. http://media.aoltimewarner.com/media/newmedia/cb_press_view.cfm?release_num=55254160
3. AYTTM. http://ayttm.sourceforge.net/
4. Babelfish Translation. http://world.altavista.com/
5. Frederking, R., Rudnicky, A., and Hogan, C. Interactive Speech Translation in the DIPLOMAT Project. *Presented at the Spoken Language Translation workshop at ACL-97*. Madrid, Spain. 1997.
6. Frederking, R., Rudnicky, A., Hogan, C. and Lenzo, K. Interactive Speech Translation in the Diplomat Project, *Machine Translation, Volume 15, Issue 1-2*, p27 – 42
7. GAIM. http://gaim.sourceforge.net/
8. Global Classroom Project. http://www.lcc.gatech.edu/projects/globalclassroom/index.html
9. Global internet user statistics (by language). http://www.glreach.com/globstats/
10. Grinter, R.E. and Eldridge, M.A. y do tngrs luv 2 txt msg?, in *Proceedings of ECSCW 2001.*, ACM press, p219--238.
11. Grinter R.E. and Palen, P. Instant messaging in teen life, *Proceedings of CSCW 2002,* ACM Press p21-30
12. Hogan, C. and Frederking, R. WebDIPLOMAT: a Web-based interactive machine translation system, *Proceedings of the 17th conference on Computational linguistics*, 2000, Saarbrücken, Germany
13. IDC research group. http://www.idc.com
14. Jabber Software Foundation. http://www.jabber.org
15. Joshi, A., Ganu, A., Chand, A., Parmar, V. and Mathur, G. Keylekh: a keyboard for text entry in Indic scripts. *Extended Abstracts of CHI 2002*, ACM Press, p928-942.
16. Language distribution on the Internet (2002). http://www.netz-tipp.de/languages.html
17. Popowich, F., McFetridge, P., Turcato, D. and Toole, D. Machine Translation of Closed Captions , *Machine Translation, Volume 15, Issue 4*, Pages 311 – 341
18. Rathod, A. and Joshi, A. A Dynamic Text Input scheme for phonetic scripts like Devanagari. *Proceedings of Development by Design (DYD) 2002.*
19. Simputer home page. http://www.simputer.org/
20. Systran Language Translation Technology. http://www.systransoft.com/
21. TechJab project wesite. http://techjab.jabberstudio.org/
22. Yang, J., and Lange, E.D. SYSTRAN on AltaVista: A User Study on Real-Time Machine Translation on the Internet, *Proceedings of AMTA 1998*, p.275-285