Meta
Journal des traducteurs
Translators' Journal

META

Multilingual Chat through Machine Translation: A Case of English-Russian

Mehmet Şahin et Derya Duman

Volume 58, numéro 2, août 2013

URI : https://id.erudit.org/iderudit/1024180ar
DOI : https://doi.org/10.7202/1024180ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Şahin, M. & Duman, D. (2013). Multilingual Chat through Machine Translation: A Case of English-Russian. *Meta*, *58*(2), 397–410. https://doi.org/10.7202/1024180ar

Résumé de l'article

Les développements récents dans le domaine de traduction automatique laissent entrevoir la possibilité imminente d'une communication sans aucune barrière linguistique, les fonctions de traduction automatique fournies par des logiciels gratuits permettant déjà à des locuteurs de différentes langues d'entrer en conversation en temps réel. La présente recherche vise à mesurer le niveau d'intelligibilité et d'exactitude grammaticale des messages de clavardage traduits instantanément par des robots de traduction intégrés au logiciel de messagerie instantanée de Google : GoogleTalk. Cette étude se fonde sur un corpus de messages échangés pendant 12 séances de clavardage en temps réel entre trois paires de participants, chacune formée d'un locuteur natif anglais et d'un locuteur natif russe n'ayant aucune connaissance de la langue de l'autre partie. L'analyse des messages échangés et des questionnaires remplis par les participants montre que les participants ont pu maintenir la conversation sur des thèmes divers sans rupture de communication majeure. Environ les trois quarts des messages se sont révélés à la fois intelligibles et exacts selon l'évaluation effectuée par des humains. Les participants aussi ont formulé des commentaires positifs sur l'efficacité de cette forme de communication interlinguale, en particulier dans le cadre d'une expérience de clavardage informel.

Multilingual Chat through Machine Translation: A Case of English-Russian*

MEHMET ŞAHİN
Izmir University of Economics, Izmir, Turkey
mehmet.sahin@ieu.edu.tr

DERYA DUMAN
Izmir University of Economics, Izmir, Turkey
derya.duman@ieu.edu.tr

RÉSUMÉ

Les développements récents dans le domaine de traduction automatique laissent entrevoir la possibilité imminente d'une communication sans aucune barrière linguistique, les fonctions de traduction automatique fournies par des logiciels gratuits permettant déjà à des locuteurs de différentes langues d'entrer en conversation en temps réel. La présente recherche vise à mesurer le niveau d'intelligibilité et d'exactitude grammaticale des messages de clavardage traduits instantanément par des robots de traduction intégrés au logiciel de messagerie instantanée de Google : GoogleTalk. Cette étude se fonde sur un corpus de messages échangés pendant 12 séances de clavardage en temps réel entre trois paires de participants, chacune formée d'un locuteur natif anglais et d'un locuteur natif russe n'ayant aucune connaissance de la langue de l'autre partie. L'analyse des messages échangés et des questionnaires remplis par les participants montre que les participants ont pu maintenir la conversation sur des thèmes divers sans rupture de communication majeure. Environ les trois quarts des messages se sont révélés à la fois intelligibles et exacts selon l'évaluation effectuée par des humains. Les participants aussi ont formulé des commentaires positifs sur l'efficacité de cette forme de communication interlinguale, en particulier dans le cadre d'une expérience de clavardage informel.

ABSTRACT

Recent developments in machine translation give hope for the possibility of communication without language barriers, as real-time interlingual conversations facilitated through automatic translation are already possible using free applications. This study aimed at measuring the level of intelligibility and accuracy of real-time chat messages translated instantly by translation bots embedded in GoogleTalk. The data consisted of chat scripts of a total of 12 sessions conducted between three pairs formed each of a native English- and Russian-speaker. The participants also answered a questionnaire about their chat experiences. The results suggest that even without any knowledge of the language of the other party, participants were able to conduct conversations on various topics without encountering any serious communication breakdown. About three fourth of the translated propositions were intelligible as well as accurate based on the human evaluation. Participants also reported positive comments on the effectiveness of this kind of interlingual communication, especially for informal chat.

MOTS-CLÉS/KEYWORDS

traduction automatique, clavardage interlinguistique, traduction automatique en temps réel, évaluation de traduction automatique
machine translation (MT), interlingual chat, real-time machine translation, machine translation evaluation

> There are no 'translating machines' which, at the touch of a few buttons, can take any text in any language and produce a perfect translation in any other language without human intervention or assistance. That is an ideal for the distant future, if it is even achievable in principle, which many doubt.
>
> (Hutchins and Somers 1992: 1)

1. Introduction

Starting its journey in the early second half of the 20th century, machine translation (MT) has gone through significant progress (Goutte, Cancedda *et al.* 2009). Now, MT is being used in many settings with various purposes. Hutchins (2003: 161-162) lists three purposes for the use of machine translation: (a) dissemination (translating text for publication in other languages), (b) assimilation (translating text for the purpose of understanding its essential content), and (c) interchange (translation for cross-language correspondence). Yang and Lange list five functions of online MT: as an assimilation tool, as a dissemination tool, as a communication tool, as an entertainment tool, and as a learning tool (Yang et Lange 2003: 201-202).

Web MT has been available on the Internet for about two decades and is being widely used for various purposes (Koehn 2010). Although some researchers reported inefficiencies of MT in specific domains, such as Yates (2006), who stated that Babel Fish can hardly be used as a translation tool in a law library, more studies investigating online translation systems reported positive results unless the users have unrealistic expectations (i.e., much more than grasping the intent of the original). Kit and Wong (2008) focused on the use of MT for the translation of legal texts from 13 popular languages to English comparing six Web MT systems, namely Babel Fish, Google, ProMT, SDL free translator, SYSTRAN, and WorldLingo. Although they argue that the performance of a MT program depends on the genre of texts as well as on language pairs, their quantitative evaluation of MT of legal texts with six different systems suggests that MT can be a useful aid even for legal texts, which are quite sophisticated with respect to syntactic complexity. Aiken, Vanjani and Wong (2006) conducted a research on the comprehensibility of Spanish to English translations made by SYSTRAN. The evaluators' responses showed that 10 out of 12 texts could be understood, which suggests an 83% accuracy. In their study, Aiken, Vanjani and Wong equated accuracy with understandability (i.e., intelligibility).

In their discussion on teletranslation and teleinterpretation as new modes of translation required for interlingual computer-mediated communication (CMC), O'Hagan and Ashworth (2002: 58) mention the use of "chat MT" for the translation of "interactive text." Yang and Lange (2003: 207) discuss the possibility of "chatting multilingually" with a focus on a then-popular application AmiChat where users can chat in any one of eight languages. Rather than being stand-alone applications, current multilingual chat environments usually work through embedding Web MT systems into chat applications such as GoogleTalk or Microsoft Lync.

Although there is a satisfactory amount of research on Web MT systems, there is as yet little experimental evidence concerning the effectiveness of the use of such systems in chat conversations, that is, interactive texts. One of the earliest studies was of Aiken *et al.* (2002; cited in Aiken, Ghosh *et al.* 2009), who conducted a research on the accuracy and intelligibility of a web-based machine translation (SYSTRAN)

of an electronic meeting between four conversers speaking English, German and French. Two objective reviewers evaluated the overall accuracy and the understanding accuracy of chat messages: the former was about 50%, while the latter was 95%.

Aiken and Ghosh (2009) conducted a research on automatic translation in multilingual business meetings. They used Polyglot II, a prototype system operating on Microsoft Windows with a server that calls Google Translate[1,2] for translation. The research was conducted on 41 languages. Their findings suggest that overall accuracy was 86% for all languages. But machine translation worked better with some languages such as Italian, Serbian and Russian, than with some others like Filipino, Japanese and Hindi.

Calefato, Lanubile and Minervini (2010) investigated "the adoption of machine translation (MT) services in a synchronous text-based chat in order to prevail over language barriers when stakeholders are remotely negotiating software requirements." They used two real-time MT services: Google Translate and Apertium. They adopted a 4-point Likert scale as a scoring scheme, anchored with values, 4 being "completely inadequate," and 1 being "completely adequate." They found that Google Translate produces significantly more adequate translations from English to Italian than Apertium and that both services can be used in text-based chat without disrupting real-time interaction. Aiken and Balan (2011) also focused on the accuracy of Google Translate by reviewing several studies comparing Google Translate with other MT systems, and they concluded that although accuracy rates vary across languages, "translations between European languages are usually good."

This study aims at providing further empirical evidence as to whether such real-time translation tools can be used effectively in daily chat conversations between chat partners who have no knowledge of each other's language. For this experiment, the English-Russian language pair was chosen because of the following reason: English and Russian come from two different language families and therefore have different scripts, those facts being likely to minimize the possibility of the participants guessing the intent of the messages based on loans and/or cognates. We seek to provide answers to the following research questions for the English-Russian language pair:

1. What percentage of the machine-translated chat messages is deemed intelligible?
2. What percentage of the machine-translated chat messages is deemed accurate?
3. Are there any differences across different tasks and target languages in the intelligibility and accuracy of the messages?
4. What are the common sources of translation problems in such settings?

2. Method

This study used qualitative and quantitative data to investigate the effectiveness of the real-time chat between two languages through machine translation.

Chat scripts were the main source of data and were analyzed using declarative evaluation criterion outlined by Nagao, Tsujii and Nakamura (1985) that investigate the translated texts along two aspects: accuracy and intelligibility.

Based on the percentage rates of intelligibility and accuracy, a declarative evaluation, a type of MT evaluation that focuses on the capability of a MT system to meet users' communicative needs (i.e., maintaining a fluent chat conversation via translated messages), was conducted in order to "test for the attributes of *intelligibility* (how

fluent or understandable it appears to be) and *fidelity* (the accuracy and completeness of the information conveyed)" (White 2000: 104).

An approach with a dual focus – intelligibility and accuracy – has several advantages for the purposes of the present study. Unlike some other criteria developed (e.g., Yates 2006; Aiken, Vanjani and Wong 2006), this one not only focuses on accuracy, but also on intelligibility, which is of great importance for the dynamic texts such as those produced by instant messaging, in which intelligibility sometimes supersedes accuracy for the sake of keeping the conversation going. As clearly put by Tatossian (2010: 290), language used in instant messaging environments is a hybrid form of communication composed of the characteristics of both written and oral discourse. This very hybrid form of chat conversation examined in this study reflects the importance of intelligibility as much as that of accuracy. Such distinction is also useful in translation quality assessment since intelligibility does not always guarantee accuracy, and vice versa.

The two scales developed by Nagao, Tsujii and Nakamura (1985) are as follows:

a) Intelligibility: Intelligibility refers to an evaluation of the extent to which the translated text can be understood by a native/native-like speaker of the target language without any reference made to the original text. Intelligibility is evaluated on a five-level scale (from 1 to 5) and each translated utterance was rated by the evaluators.
   1. The meaning of the sentence is clear, and there are no questions. Grammar, word usage, and style are all appropriate, and no rewriting is needed.
   2. The meaning of the sentence is clear, but there are some problems in grammar, word usage, and/or style, making the overall quality less than 1.
   3. The basic thrust of the sentence is clear, but the evaluator is not sure of some detailed parts because of grammar and word usage problems. The problems cannot be resolved by any set procedure; the evaluator needs the assistance of a English/Russian evaluator to clarify the meaning of those parts in the English/Russian original.
   4. The sentence contains many grammatical and word usage problems, and the evaluator can only guess at the meaning after a careful study, if at all. The quickest solution will be a re-translation of the English/Russian sentence because too many revisions would be needed.
   5. The sentence cannot be understood at all. No amount of effort will produce any meaning.

b) Accuracy: Accuracy criterion refers to the degree to which the translated text conveys the meaning of the original text is evaluated, and a measure of the amount of difference between the input and output sentences. Nagao, Tsujii and Nakamura (1985: 104) employed a seven-level scale (from 0 to 6) in the evaluation of accuracy.
   0. The content of the input sentence is faithfully conveyed to the output sentence. The translated sentence is clear to a native speaker and no rewriting is needed.
   1. The content of the input sentence is faithfully conveyed to the output sentence, and can be clearly understood by a native speaker, but some rewriting is needed. The sentence can be corrected by a native speaking re-writer without referring to the original text. No Russian/English language assistance is required.
   2. The content of the input sentence is faithfully conveyed to the output sentence, but some changes are needed in word order.
   3. While the content of the input sentence is generally conveyed faithfully to the output sentence, there are some problems with things like relationships between phrases and expressions, and with tense, voice, plurals, and the position of adverbs. There is some duplication of nouns in the sentence.

4. The content of the input sentence is not adequately conveyed to the output sentence. Some expressions are missing, and there are problems with the relationships between clauses, between phrases and clauses, or between sentence elements.
5. The content of the input sentence is not conveyed to the output sentence. Clauses and phrases are missing.
6. The content of the input sentence is not conveyed at all. The output is not a proper sentence; subjects and predicates are missing.

(adapted from Nagao, Tsujii and Nakamura 1985)

2.1. Participants

There were a total of 6 participants involved in the study:

a) 3 adult native Russian-speakers:
   - 1 female, 2 males;
   - all with text-chat experience;
   - all with no ability to use the English language beyond possibly a few isolated words;
   - all with knowledge of at least one foreign language;
   - all members of Vkontakte (the Russian version of Facebook);
   - 2 of which had previously used MT.

b) 3 adult native English-speakers:
   - 3 males;
   - all with text-chat experience;
   - all with no ability to use the Russian language beyond possibly a few isolated words;
   - all with knowledge of at least one foreign language;
   - all members of Facebook;
   - none of which had previously used MT.

2.2. Procedure

Google Talk – a free downloadable chat application by Google – was used as the software program for the chat sessions. Chat messages were translated by translation bots embedded in Google Talk, which uses the database of Google Translate, a web MT service that provides translations between 80 languages, that is, between 3160 language pairs. (At the time of the study, only 58 language pairs were supported.) Google Translate is based on statistical machine translation, which focuses on "discover[ing] the rules of translation automatically from a large corpus of translated text, by pairing the input and output of the translation process, and learning from the statistics over the data." (Koehn 2010: xi). As explained in detail on the Google Translate Web site, "[b]y detecting patterns in documents that have already been translated by human translators, Google Translate can make intelligent guesses as to what an appropriate translation should be." The texts come from already translated documents that exist in Google's database and from translated documents uploaded by human translators through Google Translator Toolkit – a free translation memory system. It should also be noted that humans can also contribute to Google Translate by correcting the current translations and suggesting alternative translations to the system.

The participants were randomly paired and they were given no information about their chat partners except for their contact information. The first chat session was appointed by the researchers, while the following ones were arranged by the conversers themselves. Participants completed four chat sessions. For each session, they were asked to chat on different topics. Task descriptions were sent to participants via e-mail in their native language. All of the chat sessions, each lasting at least 40 minutes, were observed by the researchers, and translation bots were added to the conversation where needed to ensure that the conversation flow was not interrupted (sometimes bots left the conversation unexpectedly). Translation bots are special translation agents that use Google Translate database in order to translate typed messages within Google Talk when added to the conversation as a third party.

Each participant logged in with his/her personal account to the chat program and added the translation bot to the conversation, which translated each turn automatically as they hit the Enter key. The turn-around time (the total time taken between the submission of a program for execution and the return of the complete output to the customer for the translation) was quite quick as there was almost no lagging between the moment the participants entered a message and the time it was translated into the target language and displayed both on the receiver's and the sender's screen. Below is an excerpt from one of the conversations.

Figure 1

An Excerpt from an Interlingual Chat Session through Machine Translation

A: do you work, or are you studying?
en2ru: Вы работаете, или вы учитесь?
9:11 PM **Y**: Я выучился, работал преподавателем, поступил в аспирантуру. Сейчас учусь в аспирантуре и работаю
ru2en: I learned, worked as a teacher, enrolled in graduate school. Now I study in graduate school and work
9:12 PM **A**: I work in the textile industry, and that's why i came to Turkey - textiles is finished in England
en2ru: Я работаю в текстильной промышленности, и именно поэтому я приехал в Турцию - текстиль закончена в Англии
9:13 PM **Y**: В Англии нет текстильной промышленности?
ru2en: In England there is no textile industry?
A: there is, but it' very small now and difficult for me to find work
en2ru: есть, но она очень мала и теперь мне трудно найти работу
9:14 PM **Y**: Почти тоже можно сказать и про науку у нас на Украине
ru2en: Almost the same can be said about the science we have in Ukraine
9:15 PM *ru2en@bot.talk.google.com has left*
Y: Поэтому наши ученые сотрудничают с европейской наукой
ru2en: Therefore, our scientists are collaborating with the European science
A: yes, the world has changed a lot over the last ten years
en2ru: Да, мир сильно изменился за последние десять лет
9:16 PM **A**: we are part of the Global Village
en2ru: мы являемся частью глобальной деревне
Y: В этом я нахожу больше хорошего чем плохого
ru2en: This I find more good than bad
9:17 PM **A**: me too, and we have to adapt to all the changes that come along
en2ru: мне тоже, и мы должны адаптироваться к все изменения, которые приходят вместе
9:19 PM **Y**: Я думаю, что мы должны не только адаптироваться, но самим проявлять необходимую инициативу для большего блага мира
ru2en: I think that we should not only adapt, but the very exercise appropriate initiative for the greater good of the world

Figure 1 shows that the participant A types his message in his native language, i.e. English. In the next line the translation bot, En2Ru instantly translates his message into Russian. Similarly, participant Y types his message in Russian, which is then translated into English by the second translation bot Ru2En. It should be noted that sometimes translation bots automatically leave the conversation and need to be added again.

2.2.1. Tasks

The conversers' activities were prompted through four different tasks. The use of different tasks with varying degrees of language functions was useful in coming up with a heterogeneous database in which a variety of topics appear with the relevant vocabulary.

Task 1 was designed as a free conversational environment where the participants were encouraged to introduce themselves and converse with their partners on some details of their lives (e.g., job, family, favorite spare time activities, etc.). Task 1 aimed to encourage the participants to maintain a chat conversation involving basic aspects of conversation like introducing oneself, talking about likes and dislikes, etc.

Task 2 was based on 14 pairs of pictures, each of which was accompanied by a question about the participants' preferences concerning the images shown in the presentation (e.g., are you a cat person or a dog person?). This task, in which the participants were to express their opinions and feelings about topics given beforehand, was expected to promote the use of a rich vocabulary with an expressive and emotive language output.

In Task 3, each pair of participants was given a picture story composed of six pictures. The participants were given three random pictures and were supposed to arrange them in order to have the picture story completed. The participants had to explain the content of each picture and then put them in a correct order. Task 3 promoted a sophisticated linguistic output where conversers had to use complex sentences and specialized vocabulary.

For Task 4, the participants were provided with a variety of subjects, most of them having a political or technical component, and were asked to maintain a chat conversation on the topics they had agreed upon. This task also called for sophisticated language characterized by the frequent use of technical terminology and vocabulary. Task 3 and 4 were less personal in comparison with tasks 1 and 2.

2.2.2. Instruments

The main sources of data for this study were chat scripts and questionnaires. Chat scripts from each session were saved right after they took place. The scripts included information about user names, temporal information concerning the time of logging in/out and turn-taking as well as original sentences along with their translation.

Participants were asked to complete two questionnaires, one before the chat sessions started and another after the chat sessions ended. The pre-session questionnaire was administered in order to get demographic and background information about the participants, whereas the post-session one aimed to gather information about the participants' views and opinions about the overall effectiveness of the experience, their perceptions of the comprehensibility of the messages and suggestions, etc. The results of these two questionnaires also served as qualitative and complementary data for the research.

2.3. Analysis

The total duration of the chat sessions was 811 minutes, which equals approximately 13.5 hours. During the sessions, 1808 sentences were produced and translated. This number does not include the following kind of data:

- Greetings (hello, hi, bye), discourse markers (well), yes/no replies and other expressions of dis/approval (OK, alright) and personal names as address terms were excluded when they appeared as the only utterance in a single turn. This exclusion was motivated by a concern to avoid swelling the statistics, since the translation of such occurrences was hardly challenging (148 instances);
- Meta-comments like "your robot is disconnected" were excluded from the database since these messages did not directly relate to the communicative situation and they tend to recur repeatedly during the chat sessions, which in turn may swell the statistics.

During the evaluation process, sentences were rated on the basis of the effectiveness of the translation. In order to have a more accurate evaluation, coordinated sentences were divided into clauses, and each clause was rated individually by two independent raters. The evaluators were:

- A native speaker of Russian with native-like fluency in English;
- A speaker of Russian and English with an advanced level of proficiency in Russian and a native-like proficiency in English.

Interrater reliability was calculated for the constructs of *intelligibility* and *accuracy* according to the methods of Ebel (1951), using the open-access calculator developed by Solomon (2004). The tool was used because "[t]he Rating Reliability Calculator is appropriate for use where multiple judges rate each subject being rated using a scale that constitutes interval level measurement" (Solomon 2004). The reliability score for *intelligibility* was 0.73, whereas the score for *accuracy* was 0.86. The slightly lower score for intelligibility can be explained by the relatively more subjective nature of the comprehension process. Pearson's r was also computed to test the interrater reliability. For intelligibility, Pearson's r was 0.721, whereas for accuracy it was 0.771. These correlation scores support the reliability of the rating process.

3. Findings and Discussion

Based on the evaluation of the machine-translated propositions by two independent raters, the current study provided positive empirical evidence regarding the potential of real-time machine translation. Below is a quantitative discussion of the findings based on the overall success of the program, on the direction of the translation (Russian to English and English to Russian) and on the distribution of MT performance across tasks.

Table 1 displays the overall evaluation of the translated chat messages between the two languages. In total, 50.55% of the messages was evaluated I1 A0 (Intelligibility = 1 and Accuracy = 0), which means that the messages were accurate and intelligible and did not need to be rewritten. Based on the scales used in the analysis, I1, I2, and A0, A1, and A2 meant that the translation could still be understood by the chat parties without any outside help. When such ratings are included, intelligibility and accuracy levels go even higher: 75.20% of them were rated as *accurate*, and

83.57% were rated as *intelligible*. When taken together, the percentage of the propositions rated both as intelligible and accurate at the same time within the range of I1, I2 and A0, A1, A2, (i.e., excluding any rating beyond I2 and A2, such as I2;A4 or I3A2) was 74.72%. In other words, this last figure refers to the percentage of the propositions that were deemed to be acceptable translations, which facilitated the interlingual conversation between the chat partners despite the fact that they did not know each other's language.

Table 1

Intelligibility and Accuracy of Translated Chat Messages

INTELLIGIBILITY	ACCURACY								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**TOTAL**	**Percentage of TOTAL**
1	914	40	2	1	0	0	0	957	**52.93%**
2	3	356	36	128	27	4	0	554	**30.64%**
3	1	1	6	36	127	5	1	177	**9.79%**
4	0	0	1	6	27	27	15	76	**4.20%**
5	0	0	0	0	2	12	30	44	**2.43%**
TOTAL	918	397	45	171	183	48	46	**1808**	
Percentage of TOTAL	**50.77%**	**21.95%**	**2.48%**	**9.45%**	**10.12%**	**2.65%**	**2.54%**		

As the Tables 1, 2, and 3 suggest, as the level of intelligibility increases, so does the level of accuracy, and vice versa. To put it another way, low levels of intelligibility seem to presuppose low levels of accuracy.

Table 2

Intelligibility and Accuracy of English to Russian Translation (En2Ru)

INTELLIGIBILITY	ACCURACY								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**TOTAL**	**Percentage of TOTAL**
1	427	0	0	0	0	0	0	427	**48.46%**
2	0	211	7	89	16	3	0	326	**37.00%**
3	0	0	1	8	61	4	1	75	**8.51%**
4	0	0	0	1	7	13	15	36	**4.08%**
5	0	0	0	0	0	3	14	17	**1.93%**
TOTAL	427	211	8	98	84	23	30	**881**	
Percentage of TOTAL	**48.46%**	**23.95%**	**0.90%**	**11.01%**	**9.53%**	**2.61%**	**3.40%**		

The total number of the sentences translated from English to Russian was 881. Almost half of them (48.46%) were rated as 'perfect' in terms of accuracy and intelligibility, which means that almost half of the messages were translated into the Russian language and the outcome did not need to be edited.

When taken separately, the percentage of the messages that were intelligible (I1 and I2) in translations from English to Russian was 85.46%. This rate was 73.31% for accuracy (A0, A1 and A2). These figures refer to the percentage of the translated messages that were expected to be intelligible and/or accurate enough to carry on the communicative event despite some grammar and word usage problems. On the other

hand, 14.52% of the sentences were rated as partially or totally unintelligible (I3, I4, I5) and 26.55% of them were rated as partially or totally inaccurate (A3, A4, A5 and A6).

Table 3
Intelligibility and Accuracy of Russian to English Translation (Ru2En)

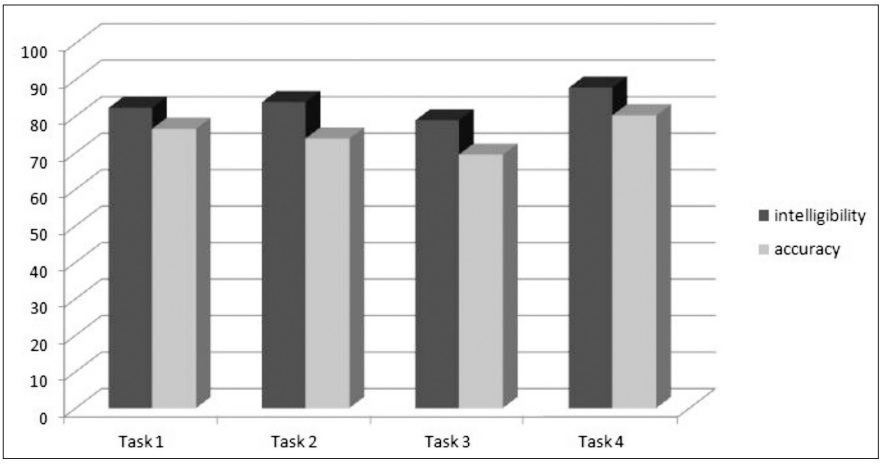
INTELLIGIBILITY	ACCURACY								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**TOTAL**	**Percentage of TOTAL**
1	486	40	2	1	0	0	0	529	**57.06%**
2	3	146	29	39	11	1	0	229	**24.70%**
3	1	1	5	28	66	1	0	102	**11.00%**
4	0	0	1	5	20	14	0	40	**4.31%**
5	0	0	0	0	2	9	16	27	**2.91%**
TOTAL	490	187	37	73	99	25	16	**927**	
Percentage of TOTAL	**52.86%**	**20.17%**	**3.99%**	**7.87%**	**10.68%**	**2.69%**	**1.72%**		

In the database, 927 sentences appeared as Russian to English translations of the chat sessions. The evaluators' ratings show that 486 of them (52.42%) were rated I1/A0, which suggests that more than half of the translations produced by the MT was of high quality in terms of accuracy and intelligibility and that the translated sentences did not need to be rewritten or retranslated. Considering the fact that the translated sentences rated between I1-I2 and A0-A2 could successfully convey the meaning and enable chat partners to perform uninterrupted communication, Table 3 suggests that 84.25% of the sentences made sense to the conversers, while 15.74% of them caused some kind of interruption/failure in the communicative act. When taken separately, propositions rated I2 constituted 24.70% of the whole conversation, whereas propositions rated A1 and A2 accounted for 20.17% and 3.99%, respectively.

Tables 2 and 3 show that when evaluated on the basis of the direction of translation, Google Translate produced similar results for the English to Russian and the Russian to English language pairs. Likewise, when evaluated on the basis of tasks, MT's success at intelligibility and accuracy produced similar results. Figure 2 shows the distribution of intelligible and accurate translations with respect to tasks.

Figure 2 represents the sum of the percentages of the translations rated A0, A1, A2 and I1 and I2. This is because beyond A2 and I2, the translations were more likely to create communication breakdowns. To put it another way, the figure above shows the percentage of the translations that enabled the chat parties to perform their communicative tasks without any serious communication problems. There seems to be little difference with respect to the four different tasks assigned to parties, only Task 3 showed a considerable decrease in intelligibility level: the intelligibility level for this task was 14% lower than the average of the three, whereas the accuracy was rated only 6% lower. The decrease in the level of intelligibility and accuracy may be due to relatively complicated nature of the task: a rather pedagogical, information gap task that requires sequencing images of a story. This finding might be supported by the nature of the conversations that took place during Task 3, in which conversers seemed to have requested clarifications and confirmations from each other much more than during the other tasks.

FIGURE 2
The distribution of intelligible and accurate translation across tasks (%)



The evaluators' ratings were supported by the respondents' answers in the post-session questionnaire. When asked about the overall effectiveness of the communication through translated chat, four participants (3 English and 1 Russian) considered it to be good, and two participants (both Russian) very good. Four of the participants asked for further explanation when a message from their chat partner was unclear, whereas one Russian participant tried to guess what the message was about, based on the context. Only one of the participants (English) reported that he or she needed to ask for re-wording from his or her chat partner because the translation was not clear, whereas all other participants reported that they almost never needed such an intervention. The same results applied for the answers to the question: "How often do you think your messages were not conveyed to your partner properly (i.e. translation was not correct)?"

Participants also provided a score of their global impression of the comprehensibility of the translated messages on a 5-point Likert scale. All Russian-speaking participants and one of the English-speaking participants reported that the translated messages were comprehensible. Two of the English-speaking participants thought that the translated messages were somewhat comprehensible (giving a score of 3 out of 5). Half of the six participants (two Russian-speaking and one English-speaking) reported that based on their experience with the tool within the framework of this research study, real-time translation produced *very effective* results for the purposes of informal chat, and the other half reported *effective* results.

The participants were asked about their overall impression concerning the use of such translation tools and about their good and bad sides. All participants had positive comments, and one of the replies to this question seems to summarize well the overall picture, supporting our quantitative findings:

(1) This was all very new to me but I thought that in general the translation tool served its purpose well. One good side is that it proves that two people with no knowledge of one another's language can communicate with each other with relative ease

thanks to this tool. On the other hand, it could lead to miscommunication quite easily; therefore, it needs to be used with a lot more care in some situations or environments.

Another important finding out of the analysis of the chat scripts was the common sources of translation problems. In their article on real-time translation on the Internet, Yang and Lange (2003: 199) list some challenges of online MT based on users' feedback: name handling, idiomatic expressions and context sensitive translations. In the current study, intelligibility and accuracy rates of the translations were observed to be relatively lower in turns involving:

- implicit use of personal pronouns in conjugated verbs in Russian;
- possessive adjective "свой" (one's own) in Russian;
- the use of personal pronoun "он" in Russian for both personal pronouns "he" and "it" in English;
- homonyms;
- one-letter expressions like references to the alphabet;
- proper names;
- abbreviations.

Another point, however, should be added to this discussion: the careful use of language by the participants. It is expected that having known they were participating in an experiment, the participants might have used language more deliberately and carefully than they would ordinarily do. Besides, the participants had been warned against substandard language use and misspellings at the very beginning of the experiment.

4. Conclusion

As for the usefulness of MT, Hutchins (2005: 11) rightly points out to the fact that the success of MT strictly depends on users' expectations. So much so that MT is "incapable of translating any text on any subject and producing unaided a good translation." But still, the studies focusing on the usefulness rather than the drawbacks of MT overweigh, and our findings suggest that this study is no exception. In the current study, the users expected MT to provide a fluent conversation on some informal topics, with a chat partner speaking a different language.

The overall evaluation suggests that 50.55% of the chat messages were translated into the target language perfectly and the outcome was structurally and semantically well-formed. On the other hand, an even higher percentage was observed when added the percentage of the translated messages that were intelligible and accurate enough not to cause any communication breakdowns or to require outside help to keep the conversation fluent: in the overall evaluation, 83.57% of the translations were intelligible and 75.20% were accurate. The findings of the present study comply with other studies that evaluate MT's success. Although these studies employed different evaluative procedures and were conducted on different language pairs, the researchers achieved similar results.

The efficiency of MT does not seem to be affected by the task or topic assigned to chat partners. Four different communicative tasks had been designed to measure slightly varying communicative situations. Among the tasks, number three was the most challenging one since partners had to describe the scenes of a picture story and

make the story complete. In this task, the accuracy and intelligibility levels of the translated messages were slightly and proportionally lower. But still the intelligibility and accuracy levels of the translated sentences were above 65%. It is, however, important to note again that intelligibility and accuracy scores were proportional across all tasks.

Despite its efficiency in instant translation, Google Translate runs into some problems, among which the incapability of the program to recognize homonyms, abbreviations, proper names and typos. Also, structural differences between Russian and English lead to problems in translation. For example, verb conjugation and pronoun system of Russian are common sources of problems in machine translation. Besides, in real chat contexts, participants tend to produce highly ill-formed structures, which can create poor results for machine translation of such structures. As Hutchins and Gaspari (2007) suggest, more research in this field is still needed.

The findings suggest that when Russian and English are concerned, Google Translate (as a free service) works as an efficient tool for translating instant messages. Looking at the rapid development of machine translation services and their constantly increasing database, it would not be too bold to speculate that thanks to such instant translation programs, language barriers all around the world are likely to be transcended in the near future. This would pave the way for a more globalized world in which communicative problems caused by language barriers in every aspect of human experience, let it be business or personal messaging, are being alleviated.

5. Suggestions for Further Research

Beyond a declarative evaluation of multilingual chat environments as we attempted to carry out in the current study, a diagnostic approach with a focus on the ineffective examples of translated chat would also illuminate a number of issues about the translated chat programs: on the one hand, it may explain the shortcomings of the program, which in turn may be used for the betterment of the software. On the other hand, a diagnostic approach may deal with the linguistic structures that are potentially disadvantaged for machine translation (e.g. subordination, figures of speech, idiomatic expressions, etc.), thus such findings may be used to improve the efficiency of the programs.

A pragmatic approach, on the other hand, may explain the communicative strategies used by conversers in dealing with the inefficiencies of the translated chat programs. The findings suggest that despite the shortcomings of the translated chat programs, participants are able to overcome these problems. Translated chat as a medium of communication seems to have its own conversational maxims and it appears as a new area of study.

This study focused on instant communication, which is closer to spoken form of language with its characteristic features like simpler sentence structure and informality. Further studies may address written genres like textbooks, legal texts, etc., which in turn would evaluate the capacity and usability of MT programs for such contexts.

ACKNOWLEDGMENTS

We would like to express our gratitude to the participants and evaluators who voluntarily devoted their valuable time for this experiment and to two anonymous reviewers for their valuable feedback on the earlier version of this manuscript. We would also like to thank Marie-Claude Hébert for editing the abstract of this manuscript in French.

NOTES

* The study was conducted in January 2011.

1. *Google Translate* (Last update: 16 February 2012): Visited on 22 February 2012, <http://translate.google.com/about/index.html>.
2. *Inside Google Translate* (Last update: 2011): Visited on 1 August 2011, <http://translate.google.com/about/intl/en_ALL/>.

REFERENCES

Aiken, Milam, Rebman, Carl, Vanjani, Mahesh B., and Robbins, Tim (2002): Meetings without borders: A multilingual Web-based group support system. *In*: *Proceedings of the America's Conference on Information Systems* (AMCIS, 9-11 August 2002, Dallas, Texas).

Aiken, Milam, Vanjani, Mahesh B. and Wong, Zachary (2006): Measuring the accuracy of Spanish to English translations. *Issues in Information systems*. 7(2):125-128.

Aiken, Milam, Ghosh, Kaushik, Wee, John, *et al.* (2009): An evaluation of the accuracy of online translation systems. *Communications of the IIMA*. 9(4):67-84.

Aiken, Milam and Ghosh, Kaushik (2009): Automatic translation in multilingual business meetings. *Industrial management and data systems*. 109(7):916-925.

Aiken, Milam and Balan, Shilpa (2011): An analysis of Google Translate Accuracy. *Translation journal*. 16(2). Visited on 8 December 2011, <http://translationjournal.net/journal/56google.htm>.

Calefato, Fabio, Lanubile, Filippo and Minervini, Pasquale (2010): Can real-time machine translation overcome language barriers in distributed requirements engineering?. *In*: Patrick Kellenberger, ed. *2010 IEEE International Conference on Global Software Engineering*. (*International Conference on Global Software Engineering*, Princeton, 23-26 August 2010). Los Alamitos: IEEE Computer Society, 257-264.

Ebel, Robert L. (1951) Estimation of the reliability of ratings. *Psychometrika*. 16:407-424.

Goutte, Cyril, Cancedda, Nicola, Dymetman, Marc, *et al.*, eds. (2009): *Learning machine translation*. Cambridge: MIT Press.

Hutchins, John (2003): The commercial systems: The state of the art. *In*: Harold Somers, ed. *Computers and translation: A translator's guide*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 161-174.

Hutchins, John (2005): Current commercial machine translation systems and computer-based translation tools: System types and their uses. *International journal of translation*. 17(1-2):5-38.

Hutchins, John and Somers, Harold. L. (1992): *An introduction to machine translation*. London: Academic Press.

Hutchins, John and Gaspari, Federico (2007): Ten years of online machine translation: origins, developments, current use and future prospects. *In*: Maegaard Bente, ed. *MT Summit XI*. (*MT Summit XI*, Copenhagen, 10-14 September 2007). Copenhagen: University of Copenhagen, 199-206.

Kit, Chunyu and Wong, Tak Ming (2008): Comparative evaluation of online machine translation systems with legal texts. *Law library journal*. 100(2):199-322.

Koehn, Phillip (2010): *Statistical machine translation*. Cambridge: Cambridge University Press.

Nagao, Makoto, Tsujii, Jun-ichi and Nakamura, Jun-ichi (1985): The Japanese Government Project for Machine Translation. *Computational linguistics*. 11(2):91-110.

O'Hagan, Minako and Ashworth, David (2002): *Translation-mediated communication in a digital world: facing the challenges of globalization and localization*. New York: Multilingual Matters.

Solomon David J. (2004): The rating reliability calculator. *BMC Medical Research Methodology*. 4:11. Visited on 6 February 2012, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC411037/>.

Tatossian, Anaïs (2010): Vers une classification générale des variants graphiques des dialogues en ligne? Le cas du français, de l'anglais et de l'espagnol. *Études de linguistique appliquée*. 159:289-308.

White, John S. (2000): Contemplating automatic MT evaluation. *In*: John S. White, ed. *Envisioning machine translation in the information future: 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000*. 1st ed. Pittsburgh: Springer, 100-109.

Yang, Jin and Lange, Elke (2003): Going live on the Internet. *In*: Harold L. Somers, ed. *Computers and translation: A translator's guide*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 191-210.

Yates, Sarah (2006): Scaling the tower of Babel Fish: An analysis of the machine translation of legal information. *Law library journal*. 98(3):481-491.