# Real-Time Machine Translation for Software Development Teams

Rafael Prikladnicki, Tiago Duarte
PUCRS
Porto Alegre, Brazil
rafaelp@pucrs.br, tiagoduarte@gmail.com

Tayana Conte
UFAM
Instituto de Computação
Manaus, Brazil
tayana@dcc.ufam.edu.br

Fabio Calefato, Filippo Lanubile
University of Bari
Dipartimento di Informatica
Bari, Italy
calefato,lanubile@di.uniba.it

## ABSTRACT
Opportunities for global software development are limited in those countries with a lack of English-speaking professionals. Machine translation technology is today available in the form of cross-language web services and can be embedded into multiuser and multilingual chats without disrupting the conversation flow. However, we still lack a thorough understanding of how real-time machine translation may affect communication in global software teams. In this paper, we present a program of research related to real-time machine translation where we aim at investigating how MT technology could be used by software development teams located in countries where professionals are not proficient in one common language. We present the studies executed so far, including text-based and voice-based machine translation, as well as the next steps planned for this research.

## Categories and Subject Descriptors
D.2.9 [**Management**]: Programming teams

H.4.3 [**Communications Applications**]: Computer conferencing, teleconferencing, and videoconferencing.

I.2.7 [**Natural Language Processing**]: Machine translation.

## General Terms
Experimentation, Human Factors.

## Keywords
Controlled experiment; global software engineering; machine translation; requirements meetings.

## 1. INTRODUCTION
Global Software Development (GSD) is characterized by the dispersion of stakeholders across different countries, continents and time zones. Requirements engineering is one of the most communication-intensive activities in software development and, thus, it suffers much from language difficulties in global software projects [1, 2, 3]. Language is indeed an important factor that largely accounts for the success of offshore IT work in countries with strong English language capabilities, such as Ireland, the Philippines, India, and Singapore [4, 5].

However, there are several other countries, considered followers in global competition, which are increasing their presence in the global IT market. Brazil is one real example of this situation [6]. Brazil's IT industry is large – A.T. Kearney consultancy estimates that the sector employs 1.7 million people, including programmers, systems analysts, and managers [7] – and it is growing by 6.5% a year on average since 2005 [8], although the vast majority of the IT companies are focused on domestic clients and do not export. For those who export, US companies are the main clients, accounting for over 80% of demand, followed by Latin America (especially Argentina, Chile, Colombia and Mexico), and Europe (especially Germany, Spain, France, England and Portugal). Nearly 100% of Brazil's IT export clients have time zone overlap with this country [6]. However, in order to take full advantage of the time zone overlap, Brazilian sites should create richer interactions with their foreign partners. This could avoid problems such as coordination breakdown, asynchronous and not so frequent communication, lack of interactive work, among other problems that lack of rich interaction may cause. And one key element for this is more effort on the English. Unfortunately, A.T. Kearney estimates that Brazil has only 10.2 million of English speakers, or 5.4% of the population. Chile, for example, has 34.7% of English speakers; India has 8.2% (which represents 90.6 million). Another study published by KPMG in 2009 indicated that one of the disadvantages of Latin American countries is the lack of English speaking professionals [9]. In this context, there are several initiatives going on, for example, in order to include English in the qualification of the IT professionals in Brazil [6]. However, this may be not enough and, to stay competitive in the global IT market these countries we will have to search for alternative solutions. For this reason, distributed project meetings, such as requirements workshops, can benefit from machine translation, as this technology is today available in the form of cross-language chat services and it might be used in countries, such as Brazil, where there are at the same time opportunities for global projects and the lack of English speaking professionals.

Machine translation (MT) is an established technology that uses software to translate text or speech from one natural language to another. The idea of using digital computers for translation of natural languages was proposed 50 years ago [10]. The technology available today – i.e., real-time, online conversation – is experiencing tremendous growth of interest, mostly because of the Internet continuous expansion. The rise of social networking has also contributed to this growing interest, allowing users of social media to speak different languages to communicate with each other. Despite the recent progress of the technology, we still lack a thorough understanding of how real-time machine translation affects communication [27].

The remainder of this paper is structured as follows. Section 2 presents the background on Machine Translation. Section 3 described the ongoing research related to MT and the studies conducted so far, including the simulation, a controlled experiment and a replication of the controlled experiment. Section 4 presents the next steps in this program of research, concluding the paper.

## 2. MACHINE TRANSLATION BACKGROUND

Machine translation (MT) is an established technology, some 50 years in the making, which may be defined as the use of a computer to translate a text from one natural language, the source language, into another one, the target language [11]. The technology available today – i.e. real-time, online conversation – is experiencing tremendous growth of interest, on the heels of the Internet continuous expansion.

MT is difficult mainly because translation *per se* involves a huge amount of human knowledge that must be encoded in a machine-processable form. In addition, natural languages are highly ambiguous, as two languages seldom express the same content in the same way [12]. Although hybrid approaches also exist, MT systems can be broadly classified into two main categories, corpus-based and rule-based, according to the nature of the linguistic knowledge being used. The *rule-based* MT systems use knowledge in the form of rules, explicitly coded by human experts, which attempt to codify the translation process. Instead, *corpus-based* MT systems use large collections of parallel texts (i.e. pairs consisting of a text in a source language and its translation into a target language) as the source of knowledge from which the engine learns how to perform translations.

Compared to the rule-based approach, the corpus-based approach is particularly appealing to researchers because systems can be trained automatically, without any direct human intervention. Google Translate (http://translate.google.com) is an example of corpus-based MT system that applies statistical learning techniques to build language and translation models from a large number of texts, both monolingual text in the target language and text consisting of examples of human translations between the source and the target languages. The Google Translate service can be used by third-party applications because it exposes a RESTful interface [13] that returns responses encoded as JSON results (http://json.org). As of this writing, Google Translate supports the translation between any two pairs of over 50 languages, although not all at the same quality level. In our previous work [14], according to a set of human raters, Google Translate was found to produce better (i.e. more accurate) automatic translation than the rule-based Apertium service (www.apertium.org).

Accurate computer translation is particularly appealing because it is quicker, more convenient, and less expensive than human translators are. An interesting research study was conducted by Yamashita *et al.* [15, 16] who investigated the effects of machine translation on mutual understanding. The study found that shared understanding is affected by the asymmetry of machine translation since the sender of a message does not know how well it has been translated to the target language. A limitation of this study is that the researchers employed picture description as the experimental tasks in one-to-one chat communication.

Aside from research prototypes or projects (e.g. for further reading, see [17, 18, 19, 20]) also commercial tools that offer cross-language chat services are available, such as IBM Lotus Translation Services for Sametime (www-01.ibm.com/software/lotus/sametime) and, lately, VoxOx (www.voxox.com), which provide cross-language translations for most of the existing instant messaging networks. Recently Google has even pushed MT goal further releasing a Google Translate app for Android [21], which integrates automatic translation with voice recognition for the English-Spanish pair.

## 3. THE MT PROGRAM OF RESEARCH

In order to study how real-time machine translation could benefit software development teams, we have planned a program of research in several steps, as following:

### 3.1 Simulated study on text-based MT

We first run a simulated study on text-based real-time machine translation in order to compare the efficiency and the effectiveness of two MT services, Google Translate and apertium-service, in translating the messages exchanged during four distributed requirements engineering workshops. The results show that Google Translate produces significantly more adequate translations than Apertium from English to Italian and that both services can be used in text-based chat without disrupting real-time interaction. The complete details and results can be found here [14]. This study proved that state-of-the-art machine translation services, such as Google Translate, could be embedded into synchronous text-based chat with a negligible extra time. However, the simulation could not say anything about completing complex group tasks while communicating with multiple native languages.

### 3.2 Controlled Experiment on text-based MT

Our second step involved the design of a controlled experiment to investigate whether real-time machine translation could be successfully used instead of English in distributed multilingual requirements meetings [22]. Thus, we proposed the following research questions for study:

RQ1: *Can machine translation services be used in distributed multilingual requirements meetings, instead of English?*

RQ2: *How does the adoption of machine translation affect group interaction in distributed multilingual requirements meetings, as compared to the use of English?*

The complete details and results can be found here [22]. Regarding the RQ1, based on the data collected, we have found evidence that the use of MT is accepted with favor by participants and is not disruptive of the conversation flow, even during the execution of complex group tasks, such as distributed requirements meetings. Such finding is interesting because, as already shown by our previous study [14], state-of-the-art MT services are still far from 100% accuracy. Thus, the point is what an acceptable error rate is for automatic translation to be effective. We expect such rate to vary largely, depending on the criticality of the task to execute. In addition, data confirm that MT interaction is faster when it comes to contributing utterances, since native language is used, but overall it takes longer to complete the task, due to repairs (i.e. extra sentences) needed when mistranslations occur. Such findings are in line with results obtained by previous studies on MT (e.g. see [16, 23].

With respect to RQ2, we could not find any evidence of differences between MT and English interactions so far, although there are some clues (e.g. increase of participation of least proficient subjects) suggesting that differences might become evident with basic levels of English skills, but we still don't have concrete results at this time. In summary, we could observe that real-time machine translation was not disruptive of the conversation flow and, therefore, accepted with favor by participants. However, since we involved only groups with high English proficiency, we concluded that stronger effects could be expected to emerge when language barriers are more critical.

## 3.3 Replicated Experiment on text-based MT

Our third step involved a replication of the former study by means of a controlled experiment, which involves participants who are not proficient in English, that is, they are not able to communicate in English as in their mother tongue. From the initial experiment, we reused the research questions, the experimental plan, the variables and the instrumentation. Since a better command of language provides better opportunities of steering communication during meetings, one could reasonably argue that machine translation is more useful to those who are not proficient in English. Therefore, we add the following research question:

RQ3: *Do individuals with a low English proficiency level benefit more than individuals with a high level when using their native language, assisted by real-time translation*

We have investigated these research questions by means of a replication of the original controlled experiment. The complete details and results can be found here [24].

Regarding RQ1 in both the original study and in this replication the frequency of presented messages (measured by utterance per minute rate – upm) is substantially similar between EN and MT runs. Moreover, no matter what their English proficiency level is, members of multilingual groups participate in more balanced discussions when using their native language with the help of MT, instead of English. Overall, these findings from the two studies allow us to affirm that machine translation is not disruptive of the conversation flow, even during the execution of complex group tasks, such as distributed requirements meetings, and that it is accepted with favor independently of subjects' English proficiency level.

With respect to the research question RQ2, one of the results from our original study was the definition of a coding schema that emerged from the inspection of meeting logs in the original experiment. In this replication, we can observe a higher number of utterances that could not be coded because the meaning was unclear, during the two runs with native language. Such finding suggests that inaccurate translations may impair the development of shared understanding more than low English skills. In addition, a percentage as high as the 4% of utterances that cannot be coded due to poor performance of the MT service raises questions on the feasibility of supporting multilingual groups with real-time translation in professional contexts for executing crucial tasks. More specifically, although such inaccuracies neither break the communication flow nor impair interaction to the extent that a task cannot be carried out, they force participants to fix them nonetheless. And, even if such a lack of common ground can be resolved by exchanging further utterances, this requires extra time, thus decreasing the efficiency of a meeting

With respect to the research question RQ3, in terms of the levels of satisfaction and comfort perceived during the experimental runs, questionnaire analyses failed to reveal any difference, which, on the one hand confirm findings from the original study with highly proficient subjects. On the other hand, however, these results (surprisingly?) suggest that, as of now, state-of-the-art MT technology is no more beneficial to individuals with low English proficiency than it is to people with high skills in a foreign language. The only statistical significant difference observed is that people with low English skills are more incline to use MT again in multilingual group interaction, despite some flaws of the current technology

In summary, the results of this replication confirmed that real-time machine translation is not disruptive of the conversation flow, is accepted with favor, and grants a more balanced discussion. In this case, we planned a fourth step, in order to start evaluating voice-based MT technology. However, the findings also show that state-of-the-art MT technology is no more beneficial to individuals with low English proficiency than it is to people with high skills in a foreign language. Content analysis suggests that this might be due to machine translation inaccuracies, which slow down the development of a common ground.

## 3.4 Simulated study on voice-based MT

The fourth step in our study involves the study of voice-based technology, by replicating the simulation we developed for text-based technology [14]. The use of voice-based MT technology is one of the trends one could find in the Gartner Hype Cycle for emerging technologies. This is a special report that provides an assessment of the maturity, business benefit and future direction of more than 2,000 technologies, grouped in 98 areas.

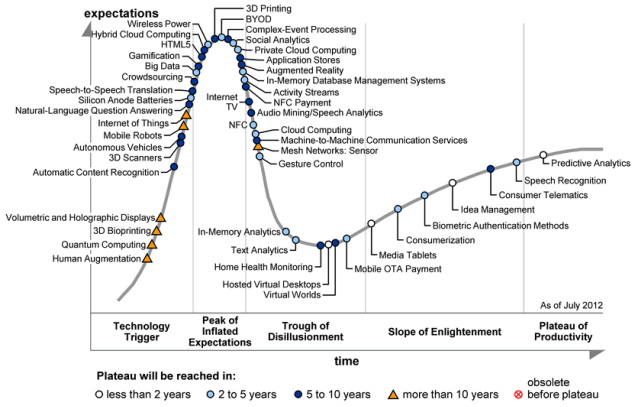


**Figure 1. Hype Cycle for Emerging Technologies, 2012 [25].**

Both in 2012 (Figure 1) and in 2013 (Figure 2), we could see speech-to-speech recognition and speech recognition as important emerging technologies. With this in mind, we developed a study in order to better understand technologies available for voice-based real time machine translation, in order to replicated the simulation study using voice. We found several technologies availabe such as Microsoft Speech API (SAPI), Microsoft .NET System.Speech, IBM Via Voice, Julius, HDecode, and Google Web Speech API, but only few of them were available for Brazilian Portuguese.

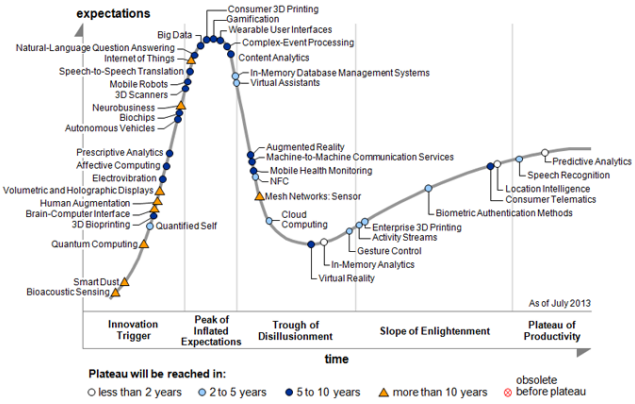


**Figure 2. Hype Cycle for Emerging Technologies, 2013 [26].**

For this reason, this fourth step will involve a simulated study on voice-based MT using English as the main language, with the main purpose of evaluating the available technology for future use by software development teams.

## 4. CONCLUSIONS

The work presented here is part of an ongoing research where the purpose is understanding to what extent real-time machine translations can be beneficial for distributed, multilingual teams located in countries where professionals are not proficient in one common language. As future work, we plan (a) new runs of the text-based MT experiments to obtain more data points and strengthen the conclusion validity; (b) new studies also involving professionals, and (c) studies in order to evaluate existing voice-based MT technologies.

## ACKNOWLEDGMENTS

This research is partially funded by the Rio Grande do Sul State funding agency (FAPERGS), and by the European Territorial Cooperation Operational Programme "Greece-Italy 2007-2013" under the project Intersocial. Rafael Prikladnicki is a CNPq researcher (309000/2012-2).

## REFERENCES

[1] D. Damian and D. Zowghi, "Requirements Engineering Challenges in Multi-Site Software Development Organizations", *Requirements Engineering Journal*, 8-3, 2003, pp. 149-160.

[2] D. Damian, "Stakeholders in Global Requirements Engineering: Lessons Learned from Practice", *IEEE Software*, 24-2, 2007, 21-27.

[3] B. Nuseibeh, and S. Easterbrook, "Requirements engineering: a roadmap," ICSE '00, pp. 35-46, June 2000.

[4] E. Carmel, and R. Agarwal, "Tactical Approaches for Alleviating Distance in Global Software Development," IEEE Softw., vol. 18, no. 2, pp. 22-29, Mar. 2001.

[5] Y. Hsieh, "Culture and Shared Understanding in Distributed Requirements Engineering," *ICGSE'06*, Florianopolis, Brazil, Oct. 2006.

[6] E. Carmel and R. Prikladnicki, "Does Time Zone Proximity Matter for Brazil? A Study of the Brazilian I.T. Industry." Technical Report, 2010, available at http://ssrn.com/abstract=1647305.

[7] A.T. Kearney. "Destination Latin America: A Nearshore Alternative, Technical Report, 2007.

[8] Brazil IT-BPO Book 2008-2009, published by Brasscom, Brazilian Association of Information Technology and Communication Companies, São Paulo, SP, Brazil, 2010.

[9] KPMG, Nearshore Attraction: Latin America Beckons as a Global Outsourcing Destination, Technical Report, 2009.

[10] M. Höst, B. Regnell, B. and C. Wohlin. "Using Students as Subjects - A Comparative Study of Students and Professionals in Lead-Time Impact Assessment." *Empirical Software Engineering*, Vol. 5, No. 3, 2000, pp. 201-214.

[11] N. Hara, C.J. Bonk, and C. Angeli, "Content Analysis of Online Discussion in an Applied Educational Psychology", *Instructional Science*, vol. 28, no. 2, pp. 115-152, 2000.

[12] D. Arnold, "Why translation is difficult for computers", In *Computers and Translation: A translator's guide*. Benjamins Translation Library, 2003.

[13] C. Pautasso, O. Zimmermann, F. Leymann, "RESTful Web Services vs. Big Web Services: Making the Right Architectural Decision", *WWW '08*, pp.805-814, 2008. doi=10.1145/1367497.1367606

[14] F. Calefato, F. Lanubile, and P. Minervini, "Can Real-Time Machine Translation Overcome Language Barriers in Distributed Requirements Engineering?", *ICGSE'10*, Princeton, NJ, USA, Aug. 23-26, pp. 257-264, 2010.

[15] C. Wohlin, P. Runesson, M. Höst, M.C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in Software Engineering, An Introduction*. Kluwer Academic Publishers, 2000.

[16] N. Yamashita, R. Inaba, H. Kuzuoka, and T. Ishida. "Difficulties in establishing common ground in multiparty groups using machine translation." *CHI '09,* Boston, USA, April 4-9, 2009, pp, 679-688.

[17] Bangalore, S., Murdock, V., and Riccardi, G. "Bootstrapping bilingual data using consensus translation for a multilingual instant messaging system." *COLING*, Taipei, Taiwan, Aug. 24 – Sep. 1 2002, Volume 1, doi:10.3115/1072228.1072362.

[18] C. Hogan and R. Frederking, "WebDIPLOMAT: a Web-based interactive machine translation system." *Proc. 18th Int'l Conference on Computational Linguistics - Volume 2*, Saarbrücken, Germany, Jul. 31 – Aug. 04, 2000, pp. 1041-1045, doi:10.3115/992730.992801.

[19] S. Jones and G. Parton. "Collaboration Across the Multinational Battlespace in Support of High-stakes Decision Making - Instant Messaging with Automated Language Translation", Technical report, The Mitre Corporation, 2008.

[20] W. Odgen. "A Task-Based Evaluation Method for Embedded Machine Translation in Instant Messaging Systems," in *Advanced Decision Architectures For The Warfighter: Foundations and Technology* (P. Mcdermott And L. Allender eds.), chapter 19, pp. 341-357, Aug. 2009.

[21] A.Verma, "A new look for Google Translate for Android", The Official Google translate Blog, January 12, 2011, http://googletranslate.blogspot.com/2011/01/new-look-for-google-translate-for.html

[22] F. Calefato, F. Lanubile, and R. Prikladnicki, "A Controlled Experiment on the Effects of Machine Translation in Multilingual Requirements Meetings", *ICGSE'11*, Helsinki, Finland, August 15-18, 2011.

[23] N. Yamashita and T. Ishida. "Effects of machine translation on collaborative work." *CSCW '06*, Banff, Alberta, Canada, November 04-08, 2006, pp. 515-524, doi:10.1145/1180875.1180955.

[24] F. Calefato, F. Lanubile, T. Conte, R. Prikladnicki, "Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment", *ESEM'12*, Lund, Sweden.

[25] Gartner's 2012 Hype Cycle for Emerging Technologies, online at http://www.gartner.com/newsroom/id/2124315, accessed on September 2013.

[26] Gartner's 2013 Hype Cycle for Emerging Technologies, online at http://www.gartner.com/newsroom/id/2575515, accessed on September 2013.

[27] H.C. Wang, S. R. Fussel, D. Cosley. "Machine Translation vs. Common Language: Effects on Idea Exchange in Cross-Lingual Groups." CSCW 2013, US.