Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Informática
Programa de Pós-Graduação em Ciência da Computação

Real-Time Machine Translation for Software Development Teams

Lucas Hilgert [1], Fabio Calefato [2], Fillipo Lanubile [2], Rafael Prikladnicki [1], Renata Vieira [1], Maria José Finatto [3], Susana Termignoni [3]

[1] Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, RS, Brazil
[2] Università degli Studi di Bari "A. Moro", Bari, Italy
[3] Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

Technical Report Nº 068

Porto Alegre, November 2012

Abstract

This technical report presents an analysis of synchronous, textual communication occurred between Brazilian and Italian teams involved in agile development planning tasks, using machine translation. The work here presented occurs in the context of the project "*The effect of Natural Language Processing on the development of Brazilian capability in the Global Software Development Market"* as an initial investigation of the linguistic issues concerning cooperative distributed multilingual requirement engineering. The aim is to discuss problems occurring in machine translation for this type of discourse and specific task domain.

Keywords: Machine Translation, Global Software Engineering

1 Introduction

Global Software Development is characterized by the geographic dispersion of development teams and stakeholders across different countries. This dispersion, besides its advantages, tends to involve different issues as, for example, distance, time zones and language [DAM08b] [LAN10].

Language is indeed one of the main difficulties associated to the execution of such distributed collaborative tasks [YAM06][LUT09] that accounts for the success of offshore IT works in countries such as Ireland, the Philippines, India and Singapore [CAR01].

In this scenario, there are countries considered followers in the global competition that lack English proficient professionals. Brazil, for example, with an estimated 1.7 million people employed in the sector (growing on average 6.5% a year since 2005) [CAR10][BRA10], is considered to have a small number of English speakers (10,2 million approximately 5,4% of the population) compared to countries such as India (90.6 million speakers).

Aiming to find alternative solutions for this problem, researchers in software engineer are evaluating the use of real-time machine translation services during the execution of collaborative distributed tasks [CAL10] [CAL12][YAM06][YAM09]. However, available real-time machine translation services are not tailored for this type of communication and domain. From this perspective, this paper presents and discusses translation problems observed in communication log files, as a road map for the improvement of translation services quality in this scenario.

This work occurs in the contexts of the project "*The effect of Natural Language Processing on the development of Brazilian capability in the Global Software Development Market"* as an initial investigation of the main problems relating language issues.

The referred project aims to contribute to Brazilian development teams inclusion in the Global Software Development Market due the utilization of Natural Language Processing (NLP) methods, techniques and tools.

The remainder of this paper is structured as follows. Section 2 presents the controlled experiment from which the log files were obtained. Section 3 presents the analysis procedure and the problems identified for this specific use of MT. Section 5 presents the conclusions and future works.

2 Real-time MT Experiments

The analysis presented in this technical report was conducted over Calefato *et al.* (2012) experiment communication logs. The referred authors conducted two controlled experiments aiming to investigate the impact of the utilization of real-time machine translation during requirements meetings. A brief description of the experiments is given later in this section.

Aiming to investigate how real-time machine translation can effectively be used during multilingual distributed requirements engineering meetings, Calefato *et al.* (2011), initially, proposed the following research questions:

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?"*

Lately, during a replication of the experiment [CAL12] trying to determine if the machine translation is more benefic to participants with lower English proficiency levels, a new research question was added:

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?*"

2.1 Participants

The experiment participants were graduate and undergraduate students from Brazil and Italy. In the original experiment [CAL11], the Brazilian participants where students from PUCRS (Porto Alegre - RS) and for the replication [CAL12], students from the Federal University of Amazonas (Manaus – AM) were chosen. On the Italian side, in both experiments the students were from the University of Bari.

For each instance of the experiment 16 students were selected as participants and distributed into 4 groups (composed by 4 people, two from Italy and two from Brazil) according to their nationalities and English proficiency levels.

The students English proficiency levels were measured using the public available Cambridge University proficiency test which includes 40 questions to be answered in 20 minutes. This test classifies the participants in four levels of proficiency: 1 (poor), 2 (basic), 3 (average), 4 (advanced).

Each participant's proficiency was then mapped in low (levels 1 and 2) and high (levels 2 and 3) categories. For the original experiment students with high proficiency levels were chosen, while for the replication experiment, due to the objective of identifying if less proficient participants benefit more from this technology (*RQ3*), only students with low proficiency levels participated.

2.2 Tools

The communication between the participants during the experiments was supported by the *eConference* tool [CAL09], a distributed meeting system which provides functionalities such as group chat, agenda, meeting minutes editing and typing awareness capabilities.

For the referred experiments, a plug-in integrating the tool's chat functionality to Google Translate translation service was developed so the participants could choose the source and target languages to be used during the meetings. The translation process was carried out in real-time.

2.3 Tasks

The tasks given to the participants consists of Planning Game (*eXtreme* Programming practice) activities, adapted from Barender's work (Barender, 2004) as it was publicly available and the chosen domain (mobile phones) was familiar to the students (which typically have equal knowledge during daily use).

The following tasks were presented to the students:

- **(T1)** Interpreting the customer role, the participants had to separate, collaboratively, a few requirements from a set of previously elicited ones and classifying them accordantly to their importance;
- **(T2)** Interpreting the developer role, the participants had to, collaboratively, complete a release plan, attributing *history points* (effort metric) to the previously separated requirements;

2.4 Design

The experiments were designed so each group executed both tasks (T1 and T2), however, alternating the strategies of communication in a way that the group that used machine translation services for Task 1, used the English language for Task 2 execution. Table 1 presents the different run configurations for the replicated experiment.

As shown in the table, during the experiment replication, groups 6 and 8

executed Task 1 using machine translation services and Task 2 using the English language as *Lingua Franca* (common language), while groups 5 and 7 used the English language for Task 1 execution and the machine translation service for Task 2. The groups were numbered from 5 because the authors considered the replicated experiment as a continuation of the original one.

Table 1: Experiment run configurations

	MT	**EN**
Run 1	Gr6, Gr8 execute T1	Gr5, Gr7 execute T1
Run 2	Gr5, Gr7 execute T2	Gr6, Gr8 execute T2

2.5 Previous Analysis

The validation conducted in [CAL11] and [CAL12] had a different objective than the analysis presented in this paper. The experiments objetive was to investigate the impacts of real-time machine translation in requirements meetings. With this objective the following qualitative criteria were observed:

- **Time:** Time taken by the groups to conclude the tasks;
- **Number of Messages:** Number of messages necessary to fulfill the tasks;
- **Frequency of Messages:** Number of messages exchanged per minute;
- **Delay:** Average time between the messages;

Analyzing the presented criteria, the referred authors aimed to investigate if the utilization of machine translation was causing some kind of delay in the communication process (time), or interfering in establishing a common ground (number of messages) as well as the time needed to write a message (frequency and delay).

A content analysis (also called coding) was also carried out. This kind of analysis consists of a mix of quantitative and qualitative data that, through the application of a coding schema, transforms qualitative in quantitative data.

The coding schema uses textual characteristics to classify utterances. Utterances providing evidence that a previous message was not fully accepted like "*what?*", for example, were classified in the *Check Misunderstanding* category, while utterances asking for confirmation, such

as "*So we decided for color screen, right?"* were categorized as *Check Provisional*. The total number of utterances in each category can be later used to get information about the communication. The complete list of categories can be seen in [CAL11].

During the previous analysis, Calefato *et al.* [CAL11] [CAL12] found out that the utilization of real-time machine translation is not disruptive of the conversation flow, even during the execution of complex tasks (*RQ1*). The results, however, showed no significant difference between the machine translation and English language utilization (*RQ2*) [CAL11][CAL12] besides the fact that some participants feel more comfortable to engage in the discussions using their own mother language.

Relating the hypothesis that participants with lower English proficiency levels would benefit more from the utilization of machine-translation (*RQ3*), no significant differences could be found in comparison with the use of English as a common language [CAL12].

The main causes associated to this low performance of real-time machine translations were associated, as in Yamashita *et al.* [YAM06][YAM09] studies, to translation errors that difficult and delay the communication process (as later presented and discussed).

3 Translation Logs Analysis

Evaluation of translation quality is an essential task, although quite subjective, and disagreements about the evaluation methodology used are rampant [CAL10]. This section presents the procedure used for our analysis of the log files.

3.1 Log Files

Each group executed at least one of the tasks using automatic machine translation, producing two different log files, one for the Brazilian participants and another for the Italian. Brazilian log files contain messages originally written in Portuguese and their translation to Italian. Italian log files contain messages in Italian translated to Portuguese. Those files were manually aligned according to their group and task in a way that each message was aligned with its respective translation.

Table 1 gives an example of the logs alignment. The highlighted messages are the original ones, English translation (between parenthesis) is included by the authors for the sake of clarity.

Table 2: Example of a log file alignment from Italian to Portuguese

Portuguese Log File	Italian Log File
“Assim começamos...” (*Thus we can start ...)*	“**allora iniziamo...”** (*We can start now...)*
“você tem idéias sobre como proceder?” (*do you have an idea of how should we proceed?*)	**“avete idee su come procedere?”** (*do you have an idea of how should we proceed?)*

3.2 Message Analysis Procedure

The message analysis task was conducted emphasizing the following central question (for the automatic translated messages): *"Can this message be understood?"* To make the analysis less subjective, the adequacy scale used in [CAL10] was adapted to determine the categories of messages, as shown in Table 2.

Table 3: Adequacy scale. Adapted from [CAL10].

Categories of message
1. Completely adequate
Translation clearly reflects the information contained in the original sentence. It is perfectly clear, intelligible, grammatically correct, and reads like ordinary text.
2. Fairly adequate
Despite some inaccuracies or infelicities, the translation generally reflects the information contained in the original sentence. One can understand (almost) immediately what it means.
3. Poorly adequate
Translation poorly reflects the information contained in the original sentence, containing grammatical errors and/or poor word choices. The general idea of the translation is intelligible only after considerable study.
4. Completely inadequate
Translation is unintelligible being impossible to obtain the information contained in the original sentence. Studying the meaning of the translation is hopeless.

The present analysis considered the messages classified in categories 2, 3 or 4 of the adequacy scale presented in Table 3, accordantly to the question previously presented.

4 Common Translation Errors

In this section the most common problems found during the log files analysis are presented and discussed, as well as some of the communication problems that they can lead to.

4.1 Numeral Reordering

One of the most frequently observed errors was the reordering of messages containing numerical values, as demonstrated in Table 4. Highlighted messages correspond to the original ones, English translations are included for the sake of comprehension.

Table 4: Example of numeral reordering problem

Portuguese	Italian
"Controle de voz Jogos 170 16 0 120 Cor da tela" (*Voice Control Games 170 16 0 120 Screen Color)*	**"Voice Control 16 0 Games 170 Color Screen 120"** (*Voice Control 160 Games 170 Screen Color 120*)
"não ponha em tipo 200 jogos" (*do not attribute to type 200 games*)	**"no metti tipo 200 a giochi"** (*do not attribute like 200 to games*)

Two types of sentence reordering are presented in Table 4. In the first example numbers are randomly grouped at the center of the message, changing the correct order of attributions. The second kind of reordering consists in placing numerals in front of the nouns as if they were quantifying it.

The grouping of numbers affects the right requirement/effort points attributions, delaying the communication and leading to misunderstandings, as can be seen in Table 4 first example where the "*Games*" functionality effort value can be wrong assumed as "170", when "160" is the right one.

Placing numerals before nouns can also lead to misunderstandings. For example, in Table 4 second row, where the statement can be understood as "*200 games*" or "200 *history points* for the games functionality" (right choice).

As could be observed during the analysis, messages that used separators (commas, for example) between value attributions (requirement/value) presented a lower rate of reordering, as can be seen in Table 5.

Table 5: Examples of correct attribution translation

Portuguese	Italian
"Alarme 50, 40 SMS, Agenda Wi-Fi 5 0 40, de memória 40" (*Alarm 50, SMS 40, Phonebook Wi-fi 50 40, of memory 40*)	"**Alarm 50, SMS 40, Wi-fi 5 0 Phonebook 40, memory 40**" (*Alarm 50, SMS 40, Wi-fi 50, Phonebook 40, memory 40*)

One of the possible solutions to kind of problem is the utilization of separators between attributions. The proposed solution can be incorporated as a writing good practice or as system warnings, alerting the authors before sending the messages to the translation services.

4.2 Use of Unexpected Languages

The use of words and expressions in other languages can also be a source of errors, as seen in Table 6, where English words are mixed in Portuguese and Italian messages.

Table 6: Examples of errors involving unexpected languages

Portuguese	Italian
"o menos importante" (*the less important)*	**"i less important"** (*the less important)*
"removê-lo apresenta uma"	**"toglilo ad una features"**

The examples in Table 6 show an inconsistent behavior of the translation service, once the English term "*features*" is discarded while other English words like "*important*", for example, are correctly translated. This behavior can lead to unpredictable problems as many terms used during this kind of meeting have no equivalent for most of the languages. Worth noting that in the first example the sentence subject number was changed from plural to singular, fact that can lead to misunderstandings.

Therefore, another good practice is to avoid the use of other languages, allowing it in special cases (technology or domain terminology for instance). The use for domain vocabulary containing usual English terms that do not require any translation should work together with the translation service.

4.3 Abbreviations

In language discourse and dialogues, as two people repeatedly refer to

the same entities, they tend to shorten the referential expression used [YAM06]. One way to shorten these expressions is to abbreviate them. The problem with abbreviation is that there are words from which abbreviations might have different meaning and, by consequence, different translations than its original forms. Table 7 presents some examples of problems caused by such abbreviations.

The abbreviation of the term *"bluetooth" ("blue"*) (Table 7), for example, when used in English, remains similar to the original word (or, at least, resembles it). However, when translated to Portuguese, "*blue*" becomes "*azul*" (the blue color), different from the term "*bluetooth*" in Portuguese that has no equivalent, so is used in its original form. The same problem occurs for the term "*ring tone*".

So, avoiding the abbreviation (good practice), the utilization of auto-completion functionalities in the communication tool and the utilization of a list containing the most frequent used terms with their possible abbreviations (at least the most frequent ones) were the identified solutions.

Table 7: Examples of abbreviation errors

Portuguese	Italian
"MMS calculadora 70 calendário 40 40 60 35 anel azul vibrando temporizador" (*mms calculator 70 calendar 40 40 60 35 blue ring vibrating, timer)*	**"mms 70 calendar 40 calculator 40 blue 60 ring 35 vibrating 40 timer 30 notes 35"** (*mms 70 calendar 40 calculator 40 blue 60 ring 35 vibrating 40 timer 30 notes 35)*
"MMS calculadora 70 calendário 40 40 60 40 azul anel vibrador temporizador 40 40 40 notas" *(mms calculator 70 calendar 40 40 60 40 blue ring vibrating, timer 40 40 40 notes)*	**"mms 70 calendar 40 calculator 40 blue 60 ring 40 vibrating 40 timer 40 notes 40"** *(mms 70 calendar 40 calculator 40 blue 60 ring 40 vibrating 40 timer 40 notes 40)*

The use of automatic completion functionalities allows the participants to write their messages faster (one of the main reasons for abbreviation). These functionalities can use an initial base vocabulary base or construct one dynamically during the communication process (or both).

4.4 Conversational Markers

Conversational markers are other source of problems. These kinds of formation sometimes are not correctly translated, as can be observed in

Table 8.

Table 8: Example of expression translation error

Portuguese	Italian
Lucas é bom... (*Lucas is good...)*	**va bene Lucas....** (*ok Lucas...)*

There might be misunderstanding once the communicative intention of the participant is not correctly expressed in the target language. Once this kind of expression is direct related to pragmatic and cultural aspects of the language, its treatment is not an easy task.

4.5 Typing Errors

Typing errors (also known as typos) is a kind of error, frequently observed during the analysis that can directly influence in the translation processes as demonstrated in Table 9.

Table 9: Examples of typing errors

Portuguese	Italian
"VaBene é a soma é de 1000" *(VaBene is the sum result is 1000)*	***"si vabene la somma è 1000"*** *(ok, the sum result is 1000)*
"SFOR 20"	***"sforo di 20"***
"por exemplo, a RME bom para você?" *(for example, are the RME good for you?)*	"**pe rme a bene per voi??**" *(all righ for me, and for you?)*

In Table 9 third row, for example, the typing error in which the expression "per me" (for me) is written as "*pr rme*" creates a new referent (*"RME"*) that could be understood as a new element, completely changing the sentence sense.

The problems caused by this errors are unpredictable. The results can be something meaningless like "*SFOR 20*", only delaying the communication process, or can cause some misunderstanding as in the third line of the table where the sentence sense is completely changed.

The simplest solution for this kind of problem is the utilization of spell-checking methods (frequently used by text editors) and auto-complete functionalities adapted for the software's domain (specific vocabulary).

4.6 Inconsistent Translations

Another major error found during the log files analysis were the inconsistent way in which some terms were translated. For this paper, translation inconsistency is roughly considered to be the attribution of multiple translations for the same source word. Table 10 demonstrates examples of inconsistencies in the translation of term "*release*".

Besides the 3 different translation presented by Table 10, another 3 were found, accounting 6 different translations: *"lançamento", "entrega",* "liberar", "reles", *"versão" "release"*.

One problems caused by this errors is the creation of multiple referents for the same object. This multiple referents can be identified as different objects causing confusion in the communication process.

The simplest solution for this kind of problem is the adoption of a single translation as the standard one [NAK10]. However, to determine the most adequate one (once the communication is done in real time) a bilingual vocabulary is necessary.

Table 10: Examples of inconsistent translations

Portuguese	Italian
"Lucas não aumentar as do segundo lançamento" (*Lucas do not increase the ones of the second launching)*	**"no Lucas aumenta quelli della seconda release"** (*no Lucas,* you should *increment the quantity of the second release"*
"mas podemos acrescentar, por exemplo, a chamada versão 3 vibração de alerta com 50 pontos" (*but we can add, for example, the called release 3 vibrating alert with 50 points)*	**"ma possiamo aggiungere per esempio nella 3 release vibrating call alert con 50 punti"** (*but we can add, for example, in the third release vibrating call alert with 50 points)*
"devemos também dividir releases" (*we must split them in releases too*)	**dobbiamo suddividerli anche in release** *(we split them in release)*

5 Conclusion

In this technical report the communication log files from a previous experiment [CAL12] were analyzed with the objective of identifying the main problems related to the application of real-time machine translation to multilingual distributed software requirement engineering, from a machine translation point of view.

As a result, six frequently translation problems could be identified. These errors were then classified as structural (numeral reordering) and vocabulary related (typing error, abbreviation, inconsistent translations, different languages and conversational markers).

We propose as solutions for these problems the use of message preprocessing such as spell-checking and automatic-completion functionalities, which can be added in the communication tool. An advantage of these preprocessing tasks is that they are time efficient (being frequently used in text editors, for example).

As future work, the required linguistic resources (such as bilingual vocabularies) for the development of the proposed solutions will be generated from domain specific corpora aiming to improve the translation results for this specific domain.

References

[BRA10] Brazil IT-BPO Book 2008-2009. (2010), published by Brasscom, Brazilian Association of Information Technology and Communication Companies, São Paulo, SP, Brazil.

[CAL09] Calefato F. and Lanubile F. “Using frameworks to develop a distributed conferencing system: an experience report”. In: Softw. Pract. Exper. 2009.

[CAL11] Calefato F., Lanubile F. and Prikladnicki R. “A Controlled Experiment on the Effects of Machine Translation in Multilingual Requirements Meetings”. In: Global Software Engineering (ICGSE), 2011 6th IEEE International Conference, 2011, pp. 94 -102.

[CAL10] Calefato F., Lanubile F. and Minervini P. “Can Real-Time Machine Translation Overcome Language Barriers in Distributed Requirements Engineering?”. In: Global Software Engineering (ICGSE), 2010 5th IEEE International Conference, 2010, pp. 257 264.

[CAL12] Calefato F., Lanubile F., Conte T. and Prikladnicki R. “Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment”. In: 6th Int'l Symposium on Empirical Software Engineering and Measurement (ESEM’12), 2012, pp. 254-260.

[CAR01] Carmel, E. and Agarwal, R. “Tactical Approaches for Alleviating Distance in Global Software Development”. In: IEEE Softw., vol. 18, no. 2, 2001, pp. 22-29.

[CAR10] Carmel, E. and Prikladnicki, R. “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.

[DAM08a] Damian D., Lanubile F. and Mallardo T. “On the Need for Mixed Media in Distributed Requirements Negotiations”. In: IEEE Trans. Software Engineering, 34(1), 2008a, pp.116-132.

[DAM08b] Damian, D., Sengupta B. and Lanubile F. In: “Global software development: where are we headed?”. Software Process: Improvement and Practice, Volume 13, Issue 6, Nov-Dec, 2008b, pp: 473-475.

[LAN10] Lanubile F., Damian D. and Heather O. L. “Global software development: technical, organizational, and social challenges”. In: SIGSOFT Softw. Eng. Notes 28, 6, 2003.

[LAN06] Lanubile F., Ebert C., Prikladnicki R. and Vizcaino A. “Collaboration Tools for Global Software Engineering”. IEEE Software, Volume 27, 2010, pp.52-55.

[LUT09] Lutz, B. “Linguistic Challanges in Global Software Development: Lessons Learned in an International SW Development Division”. Proceedings of the 2009 Fourth IEEE International Conference on Global Software Engineering. IEEE Computer Society. 2009, pp.249-253.

[NAK10] Nakatsuka M.;Yasunaga, S., Kuwabara, K. “Extending a multilingual chat application: Towards collaborative language resource building”. Cognitive Informatics(ICCI) 2010. 9th IEEE International Conferenceon. 2010. Pages: 137 –142.

[YAM06] Yamashita N. and Ishida T. “Effects of machine translation on collaborative work”. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work CSCW 06. ACM Press, 2006. pp. 515-524.

[YAM09] Yamashita N., Inaba R., Kuzuoka H. and Ishida T. “Difficulties in establishing common ground in multi-party

groups using machine translation". In: Proceedings of the 27th international conference on Human factors in computing systems. 2009, pp.679-688.