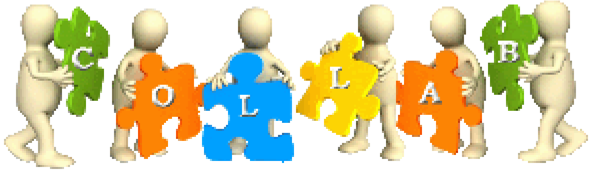
Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment

Fabio Calefato, **Filippo Lanubile**
University of Bari, Italy

Tayana Conte
Universidade Federal do Amazonas, Brazil

Rafael Prikladnicki
PUCRS, Brazil

Motivation



- Global software projects challenged by language differences
  - especially requirements meetings
- Machine translation technology for remote meetings in countries with
  - Opportunities for global projects
  - Lack of English speaking professionals

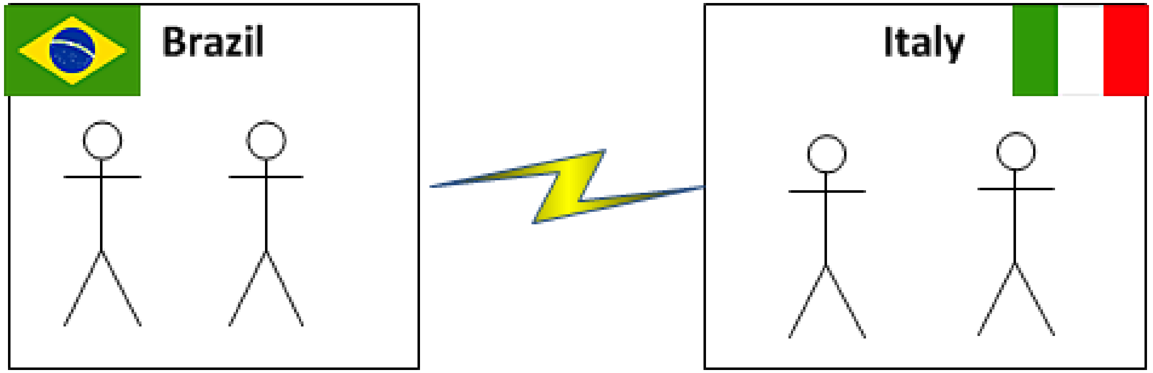
Research questions

- RQ1: *Can MT services be used in distributed multilingual requirements meetings?*
  (instead of English)
- RQ2: *How does the adoption of MT affect group interaction?*
  (in distributed multilingual requirements meetings)

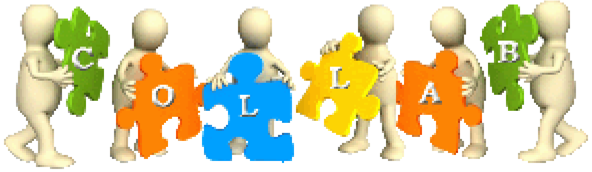
Original experiment

F. Calefato, F. Lanubile, R. Prikladnicki. "*A Controlled Experiment on the Effects of Machine Translation in Multilingual Requirements Meetings*“, ICGSE 2011.

- Participants: 16 students from Bari (Italy) and PUCRS, Porto Alegre (Brazil)
- Multilingual groups highly proficient in English



Experimental tasks



T1 – requirements prioritization (30 min.)

– Customer’s perspective

1. Assign 16 mobile phone features to 3 piles: *very important, important, less important*
2. Rank the features within piles

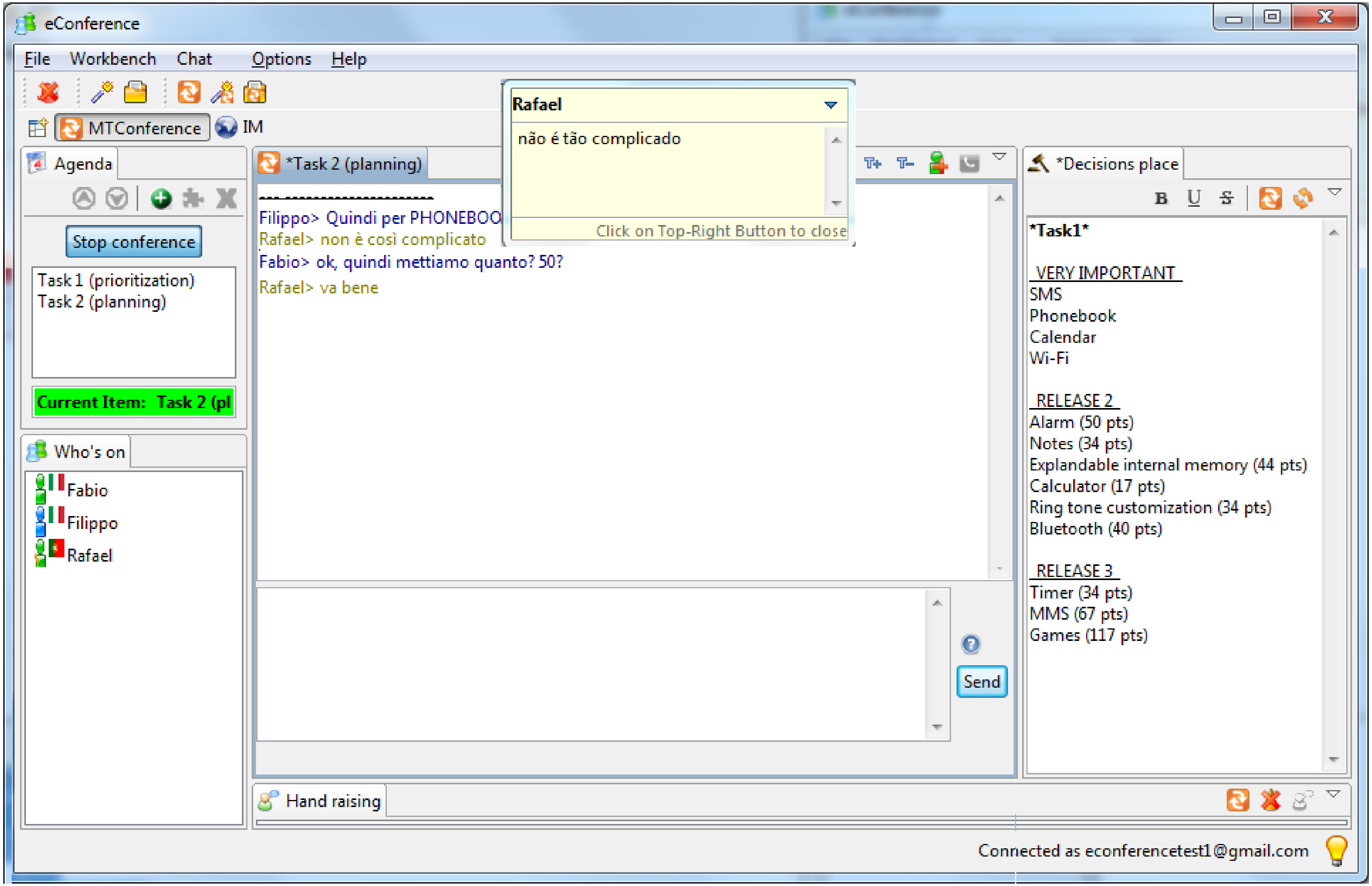
T2 – release planning (60 min.)

– Developer’s perspective

1. Distribute 1000 story points to each feature as an estimate of implementation costs
2. Plan 3 releases based on priorities (T1) and cost estimates

Instrumentation

eConferenceMT http://code.google.com/p/econference-mt-plugin



Task 1 (prioritization)
Task 2 (planning)

Current Item: Task 2 (pl

Rafael

não é tão complicado

Filippo> Quindi per PHONEBOO

Rafael> non è così complicato

Fabio> ok, quindi mettiamo quanto? 50?

Rafael> va bene

Task1

VERY IMPORTANT
SMS
Phonebook
Calendar
Wi-Fi

RELEASE 2
Alarm (50 pts)
Notes (34 pts)
Explandable internal memory (44 pts)
Calculator (17 pts)
Ring tone customization (34 pts)
Bluetooth (40 pts)

RELEASE 3
Timer (34 pts)
MMS (67 pts)
Games (117 pts)

Connected as econferencetest1@gmail.com

Original experiment: Findings

- MT can be used without disrupting the conversation flow
- Generally accepted with favor

RQ1

- More balanced discussions when using MT
- Differences might be more evident with lower levels of English skills

RQ2

Replicated experiment

RQ3: *Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?*

- Participants: 16 students from Univ. Bari (Italy) and Fed. Univ. of Amazonas, Manaus (Brazil)
- Multilingual groups
  - Same tasks
  - Same instrumentation
  - **Lowly proficient in English**

Experimental design

	Original experiment (high proficiency)		**Replicated experiment (low proficiency)**	
	MT	EN	MT	EN
Run 1	Gr1, Gr3 execute T1	Gr2, Gr4 execute T1	Gr6, Gr8 execute T1	Gr5, Gr7 execute T1
Run 2	Gr2, Gr4 execute T2	Gr1, Gr3 execute T2	Gr5, Gr7 execute T2	Gr6, Gr8 execute T2

Data sources:

- post-task questionnaires
- meeting logs

Questionnaire analysis

- Satisfaction with performance
  - No significant differences (over 4 items)
- Engagement and comfort during interaction
  - No significant differences (over 6 items)
- Perceived usefulness
  - No significant difference:
    “Group activity benefited from using *MT/EN*”
- Communication mode preference
  - One significant difference:
    “Another time, I would rather communicate using MT/EN”

Log analysis: frequency & delay

Group		Comm. mode	# Utterances	Frequency (upm)	Delay (sec.)
Gr1 (High)	Run 1	MT	159	3.95	15
	Run 2	EN	322	5.28	11
Gr2 (High)	Run 1	EN	68	4.25	15
	Run 2	MT	346	5.86	10
Gr3(High)	Run 1	MT	190	6.33	10
	Run 2	EN	462	6.90	8
Gr4 (High)	Run 1	EN	52	3.25	20
	Run 2	MT	169	3.13	14
Gr5(Low)	Run 1	EN	92	5.41	11
	Run 2	MT	358	6.17	10
Gr6(Low)	Run 1	MT	140	4.38	14
	Run 2	EN	164	2.83	21
Gr7 (Low)	Run 1	EN	264	6.44	9
	Run 2	MT	405	6.75	9
Gr8 (Low)	Run 1	MT	240	5.58	11
	Run 2	EN	354	5.28	11

Little extra delay (1.6 sec) with EN

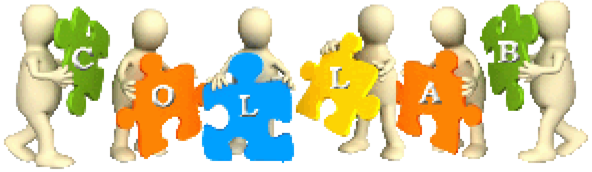
Most active groups better both with MT and EN

Log analysis: equality of participation

Group (level)	Least proficient subject (nationality)	% of utterance	
		EN	MT
Gr1 (High)	Student #7 (Brazilian)	19%	27% ↑
Gr2 (High)	Student #4 (Brazilian)	22%	26% ↑
Gr3 (High)	Student #16 (Brazilian)	32%	23%
Gr4 (High)	Student #12 (Brazilian)	10%	14% ↑
Gr5 (Low)	Student #17 (Italian)	21%	36% ↑
Gr6 (Low)	Student #22 (Italian)	20%	27% ↑
Gr7 (Low)	Student #27 (Brazilian)	15%	14%
Gr8 (Low)	Student #32 (Brazilian)	23%	26% ↑

Gain in participation of least proficient subjects with MT

Log analysis: coding



- Clarification requests as an evidence of lack of common ground
- Relevant categories:
  - Check misunderstanding *(e.g., “I didn’t get your question", "What?")*
  - Check provisional *(e.g., "So we go for color screen, right?")*
  - Unknown *(i.e., cannot be coded by raters)*

	EN (Run 1)			**MT (Run 2)**		
	Check misunderstanding	Check provisional	Unknown	Check misunderstanding	Check provisional	Unknown
Gr5 (Low)	0%	2.2%	0%	2.9%	5.9%	4.3%
Gr7 (Low)	1.9%	3.8%	0.9%	1%	1.2%	3.2%

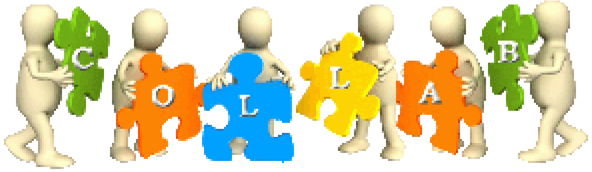
- Contrasting results
- More meaningless utterances from inaccurate translations rather than poor English

Conclusions: RQ1

Can MT services be used in distributed multilingual requirements meetings?	**Original experiment (high proficiency)**	**Replicated experiment (low proficiency)**
Satisfaction with performance	MT = EN	MT = EN
Engagement and comfort during interaction	MT = EN	MT = EN
Frequency of messages and delay between utterances	MT = EN	MT = EN
Perceived usefulness	MT = EN	MT = EN
Communication mode preference	MT = EN	MT > EN

- Confirmation that machine translation is not disruptive of the conversation flow and is accepted with favor

Conclusions: RQ2



How does the adoption of MT affect group interaction?	**Original experiment (high proficiency)**	**Replicated experiment (low proficiency)**
Equal participation	MT > EN	MT > EN
Clarification requests	-	MT = EN

- Confirmation of more balanced discussions when using native language with MT

Conclusions: RQ3

Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?

so far, **NO**

however

- people with low English skills are more prone to use MT again
- messaging is easier than talking for a non-native English speaker

Current & Future work

- Apply coding schema to remaining groups
- Assess the effects of typos on MT accuracy
- Gather more data
  - Double the # of high and low proficiency groups
- Compare with groups including native English speakers
- Replicate with other languages
  - e.g. Chinese, Japanese, Turkish, …
- Replicate with voice conferences