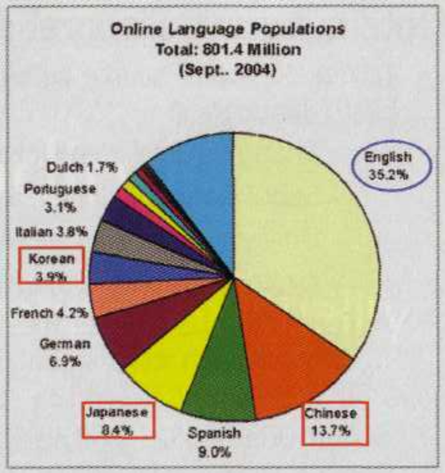
Intercultural Collaboration
Using
Machine Translation

ICE Intercultural Collaboration Experiments

Toru Ishida
Department of Social Informatics
Kyoto University

1. Motivation

- It seems there is no common language anymore in the Internet.
- English = 35.2%
- Asian languages (Chinese, Japanese, Korean) > 26%.
- We need to understand various languages when viewing Web information.



Source: Global Reach (global-reach.biz/globstats)

1. Motivation

- We have serious language barrier among Asian countries.
    - Neighboring languages are not taught in Asia.
    - Asian people cannot think in English and want to make intermediate documents in their first languages.
- Why don't we use Machine Translation?
    - MT will become the key for the IT people to contribute to mutual understanding among the world.

Experiment

ICE

2. Experiment

Intercultural Collaboration Experiment 2002

- To develop open source software in Asian countries in our FIRST languages!
  - Shanghai JiaoTong University (China)
  - Seoul National University, Handon University (Korea)
  - University of Malaysia (Malaysia)
  - Kyoto University (Japan)
- Team members never meet in person, but complete software with multilingual Communication tools: TransWeb and TransBBS.
- April-June 2002: Software Design
- October-December 2002: Software Integration

2. Experiment

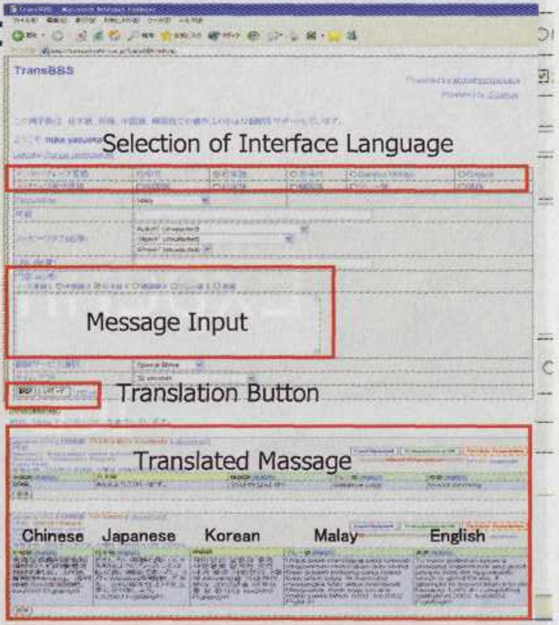
- Each country develops a multilingual tool.
- Japan: TransGroupware
- Malaysia: TransSMS
- China: TransSearch
- Korea: TransChat

Japan Team
ICE2002 Organizers
ICE Intercultural Collaboration Experiments
Malaysia Team
Korea Team
China Team

2. Experiment

- TransBBS
  - Asynchronous Collaboration Tool
  - Translation among Japanese, Chinese, Korean, Malay, and English
  - 31,000 messages during ICE2002.

Developed by Kyoto University

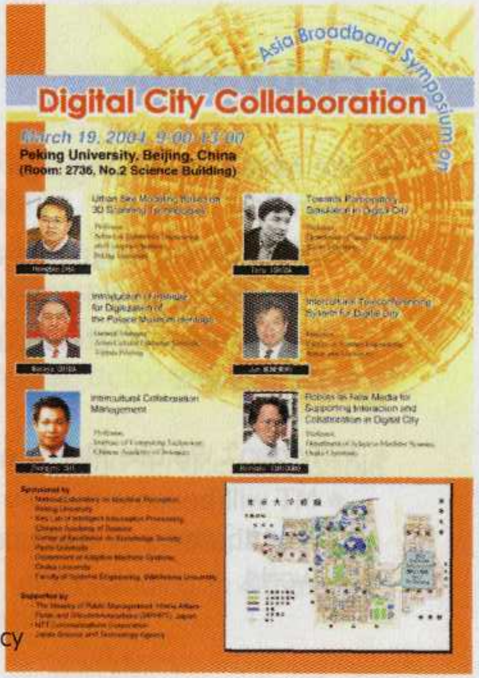

TransBBS
Selection of Interface Language
Message Input
Translation Button
Translated Massage
Chinese
Japanese
Korean
Malay
English

2. Experiment

Asia Broadband Intercultural Collaboration Experiment (ICE2003)
October 2003 – March 2004

Kyoto University
Osaka University
Wakayama University
Peking University
Chinese Academy of Science

The Ministry of Public Management, Home Affairs, Posts and Telecommunication (MPHPT)
NTT Communications Corporation
Japan Science and Technology Agency


Asia Broadband Symposium on
Digital City Collaboration
March 19, 2004, 9:00-17:00
Peking University, Beijing, China
(Room: 2736, No.2 Science Building)

2. Experiment

by Takashi Yoshino

- TalkGear
  - Synchronous Collaboration Tool
  - Video Conference, Multilingual Presentation, Multilingual Chat.

Kyoto, Japan

Peking, China

Developed by
Wakayama University

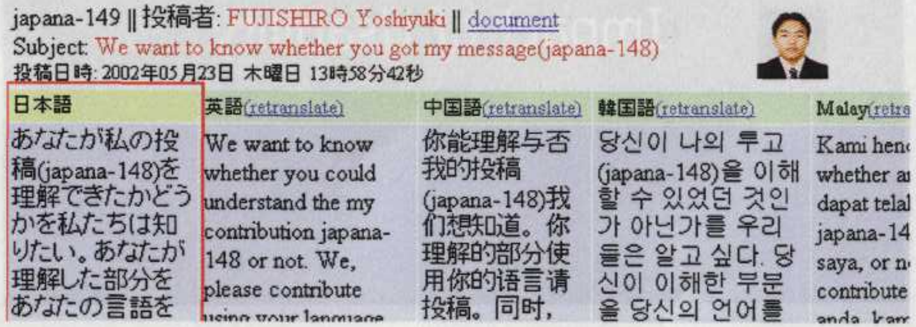
Findings

Impact of Translation Quality

---

3. Findings

- Adaptation emerged...
  - when writing (self-initiated repair)
    - Try translation again and again before posting
  - when reading
    - Browse multiple languages on TransBBS



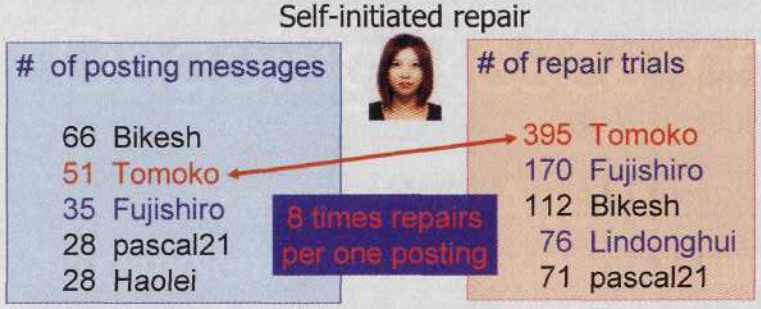
japana-149 || 投稿者: FUJISHIRO Yoshiyuki || document
Subject: We want to know whether you got my message(japana-148)
投稿日時: 2002年05月23日 木曜日 13時58分42秒

日本語	英語(retranslate)	中国語(retranslate)	韓国語(retranslate)	Malay(retra
あなたが私の投稿(japana-148)を理解できたかどうかを私たちは知りたい。あなたが理解した部分をあなたの言語を	We want to know whether you could understand the my contribution japana-148 or not. We, please contribute	你能理解与否我的投稿(japana-148)我们想知道。你理解的部分使用你的语言请投稿。同时,	당신이 나의 투고 (japana-148)을 이해할 수 있었던 것인가 아닌가를 우리들은 알고 싶다. 당신이 이해한 부분을 당신의 언어를	Kami hen whether a dapat tela japana-14 saya, or n contribute

---

3. Findings

Users modified their original messages again and again.



Self-initiated repair

# of posting messages	
66	Bikesh
51	Tomoko
35	Fujishiro
28	pascal21
28	Haolei

# of repair trials	
395	Tomoko
170	Fujishiro
112	Bikesh
76	Lindonghui
71	pascal21

8 times repairs per one posting

Why? simply because they wanted to communicate with each other.

Data on TransBBS from Oct 10 to 30.

3. Findings

by Kentaro Ogura

Effect of Self-initiated Repair
(from Japanese to other languages)

Target Language	Translation Quality			Sum
	increased	no change	decreased	
English	85.2% (46)	11.1% (6)	3.7% (2)	100.0% (54)
Korean	23.7% (14)	54.2% (32)	22.0% (13)	100.0% (59)
Chinese	42.4% (25)	47.5% (28)	10.2% (6)	100.0% (59)

---

3. Findings

by Kentaro Ogura

- Because users adapt to Japanese-to-English translation, self-initiated repair does not effect on Japanese-to-non-English translation.
- Users who have weak English skill find that self-initiated repair is useless.
- To solve this problem, we need a tool that allows user adaptation to their first language.

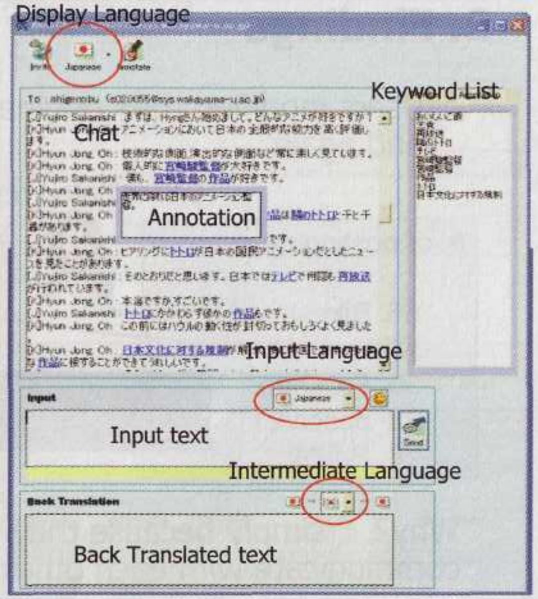
⇒ Ex. Self-initiated repair with J-to-E-to-J translation.

Kentaro Ogura, Yoshihiko Hayashi, Saeko Nomura and Toru Ishida. User Adaptation in MT-mediated Communication. *The First International Joint Conference on Natural Language Processing (IJCNLP-04)*, pp.596-601, 2004

---

3. Findings

- AnnoChat supports back translation with various intermediate languages, and multilingual annotations.


Display Language
Keyword List
Chat
Annotation
Input Language
Input
Input text
Intermediate Language
Back Translation
Back Translated text

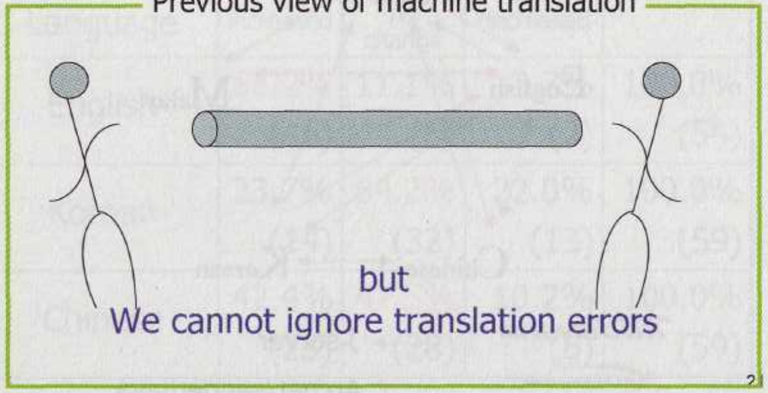
by Takashi Yoshino

Findings

Impact of Usability

4. Ongoing Research

Machine translation should be transparent.
The goal is "accuracy of translation."

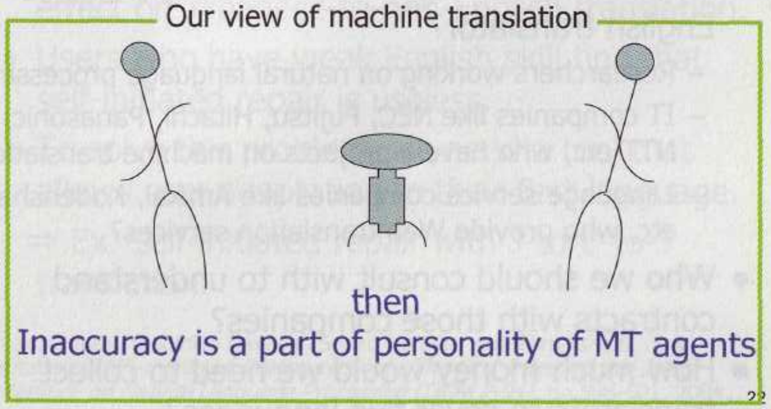


but
We cannot ignore translation errors

---

4. Ongoing Research

Machine translators can be explicit agents.
The goal is "interactivity of translators."

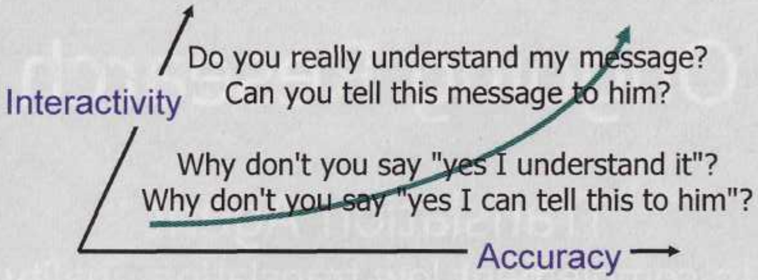


then
Inaccuracy is a part of personality of MT agents

---

4. Ongoing Research

Interactivity can be seen as a
meta level architecture.



Do you really understand my message?
Can you tell this message to him?

Interactivity

Why don't you say "yes I understand it"?
Why don't you say "yes I can tell this to him"?

Accuracy →

知るを知るとなし,知らざるを知らざるとなす,これ知るなり.
知之爲知之、不知爲不知、是知也. -孔子
To know that we know what we know, and to know that we do not know what we do not know, that is true knowledge. - Copernicus

---

4. Ongoing Research

by Kiyotaka Uchimoto

Let machine translator to know how well is a given sentence translated, and to know which part is not translated well.

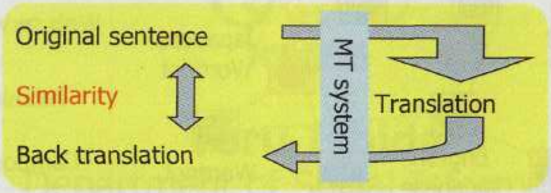
それほど上手く
私の文章も~~あまり上手に~~翻訳されません。
日本語~~の~~勉強~~から始めようかな。~~
私は　を　し直そう。

so well
My sentences, too, aren~~'t too skillf~~ully translated.
~~Will it begin at the Japanese study~~ing?
I will study Japanese once more.

4. Ongoing Research

by Kiyotaka Uchimoto

- Confidence measure (C-measure)
  - Similarity between a source-language sentence and its back translation.
    - Similarity ideally becomes high when given two sentences have the same meaning.



Kiyotaka Uchimoto, Naoko Hayashida, Toru Ishida, and Hitoshi Isahara. Automatic Rating of Machine Translatability. *In proceedings of MT summit X.* 2005.

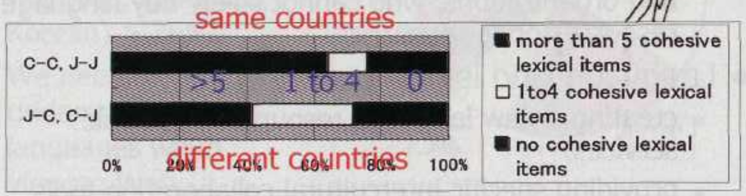
---

4. Ongoing Research

by Naomi Yamashita

Let machine translator to know how well discussion is continuing, and to know exist misconceptions.

- When people cannot fully understand the translated message, they tend to respond based on few words that stuck in their mind. (Subway effect).


same countries
C−C, J−J
J−C, C−J
>5
1 to 4
0
0%
20%
40%
60%
80%
100%
different countries
more than 5 cohesive lexical items
1to4 cohesive lexical items
no cohesive lexical items

Compare discussion pairs posted from the same country and from different countries by the number of shared lexical items.

---

4. Ongoing Research

by Naomi Yamashita

Predicting Misconceptions

- Discussion threads
  - Syntactic thread
    - Generated based on "references" and "in-reply-to" fields.
  - Semantic thread
    - Generated based on lexical similarities between messages.

→ Misconceptions can be predicted when there is a gap between syntactic and semantic threads.

Ongoing Research

Language Grid

to increase MT usability

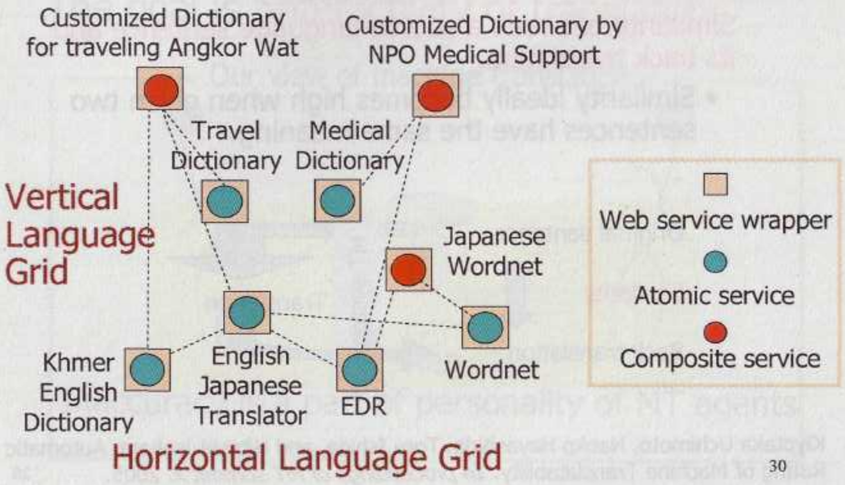
4. Ongoing Research

- Language Grid is for
  - increasing accessibility and usability of language resources/services on the Internet.
  - creating composite services by combining existing language resources and atomic services.
  - encapsulating various intellectual property rules.
  - providing a language infrastructure to individuals and organizations, who cannot solely buy language resources or services.
- Language Grid is not for
  - creating a new language resources or atomic services.
  - providing specific intercultural collaboration tools.

---

4. Ongoing Research

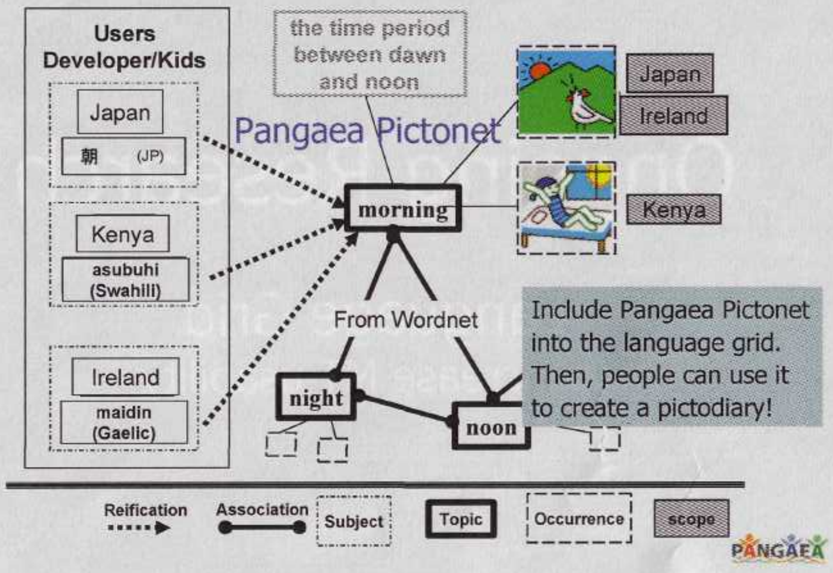
Applying Semantic Web and Multiagent Technologies


Customized Dictionary for traveling Angkor Wat
Customized Dictionary by NPO Medical Support
Travel Dictionary
Medical Dictionary
Vertical Language Grid
Japanese Wordnet
Khmer English Dictionary
English Japanese Translator
EDR
Wordnet
Horizontal Language Grid
Web service wrapper
Atomic service
Composite service

---

4. Ongoing Research

provided by NPO Pangaea


Users
Developer/Kids
Japan
朝 (JP)
Kenya
asubuhi (Swahili)
Ireland
maidin (Gaelic)
the time period between dawn and noon
Pangaea Pictonet
morning
From Wordnet
night
noon
Japan
Ireland
Kenya
Reification
Association
Subject
Topic
Occurrence
scope
PANGAEA

Include Pangaea Pictonet into the language grid. Then, people can use it to create a pictodiary!

---

5. Conclusion

- Machine translation is a powerful tool for intercultural collaboration.
  - Quality of machine translation is low but still useful.
  - Improving interactivity of machine translator is one way to complement its quality.
- We need a language infrastructure on the top of the Internet.
  - To make language resources and services accessible and usable.
  - To support various intercultural activities.
  - To contribute to mutual understanding in different people in different countries.