

The Language Grid

Multi-Language Service Platform for Intercultural Collaboration

Part 1: Getting Started with Machine Translator: ICE2002

Part 2: The Language Grid Project

Part 3: Intercultural Collaboration Project in Vietnam

Part 4: Ongoing PhD Research

Toru Ishida

Department of Social Informatics

Kyoto University

Part 1

Getting Started with Machine Translator ICE2002

At the beginning of the new millennium,
we proposed the concept of *intercultural collaboration*
where participants with different cultures and languages
work together towards shared goals.

Intercultural Collaboration Experiment 2002

Develop open source software using machine translation



Multi-Language BBS

Translation among Japanese, Chinese, Korean, Malay, and English

31,000 messages in one year.

TransBBS - Microsoft Internet Explorer

Powered by arcnets/sangenjaya
Powered by J_Server

この掲示板は、日本語、英語、中国語、韓国語での書き込みおよび翻訳をサポートしています。

ようこそ, mika yasuoaka **Selection of Interface Language**

[logout](#) || [change preferences](#)

インターフェイス言語	<input type="radio"/> 中文	<input checked="" type="radio"/> 日本語	<input type="radio"/> 한국어	<input type="radio"/> Bahasa Melayu	<input type="radio"/> English
メッセージ表示言語	<input type="checkbox"/> 中国語	<input type="checkbox"/> 日本語	<input type="checkbox"/> 韓国語	<input type="checkbox"/> マレー語	<input type="checkbox"/> 英語

Discussion: lobby

件名: _____

メッセージタグ(必須):
Action? (unselected) _____
Object? (unselected) _____
Where? (unselected) _____

URL (任意): _____

内容 (必須)
ソース言語 || 中国語 || 日本語 || 韓国語 || マレー語 || 英語

Message Input

翻訳サービス選択: Special Blend

タイムアウト: 30 seconds

Translation Button

現在, lobby ディスカッションを表示しています。

japana-256 || 投稿者: FUJISHIRO Yoshiyuki || document

件名: Requirement toward authentic...
Approach: Requirement toward authentic...
Document: Collaboration Diagram
Korea Team
投稿日時: 2003年01月14日 火曜日 10時30分54秒 (+0900)

Translated Message

中国語(再翻訳)	日本語	韓国語(再翻訳)	マレー語(再翻訳)	英語(再翻訳)
早啊。	おはようございます。	안녕하십니까?	Selamat pagi.	Good morning.

返信

japana-255 || 投稿者: lim sanry || document

件名: Good chance
App: ...
Doc: ...
Korea Team
投稿日時: 2002年12月17日 火曜日 17時30分12秒 (+0900)

Chinese **Japanese** **Korean** **Malay** **English**

中国語(再翻訳)	日本語(再翻訳)	韓国語	マレー語(再翻訳)	英語(再翻訳)
获得与有趣的实验知 道的人们的事是对 我好的机会。3月时 期待Meeting。流利 地做2002年完成吧。 Ice2002 Fighting!!!	おもしろい実験と良い人々 を知るようになったことは 私に良い機会であった。3 月にMeetingを期待してみ る。2002年を仕上げを上 手にやります。Ice2002 Fighting!!!	재미있는 실험과 좋은 사람들을 알게된 것은 내게 좋은 기회였다. 3월 에 Meeting을 기대하여 본다. 2002년을 마무리 를 잘 합니다. Ice2002 Fighting!!!	Untuk telah mendapat tahu sebuah eksperimen melucukan dan orang baik adalah peluang yang mana baik untuk saya. Ia mencuba menjangka Mac untuk membuat Mesyuarat. Ayuh siap secara mahir pada tahun 2002. Ice2002 Fight!!!	To have gotten to know a amusing experiment and good people was the opportunity which is good for me. It attempts to expect March to do Meeting. Let's do completing skillfully in 2002. Ice2002 Fighting!!!

返信

Example Discussion

japana-149 || 投稿者: FUJISHIRO Yoshiyuki || [document](#)

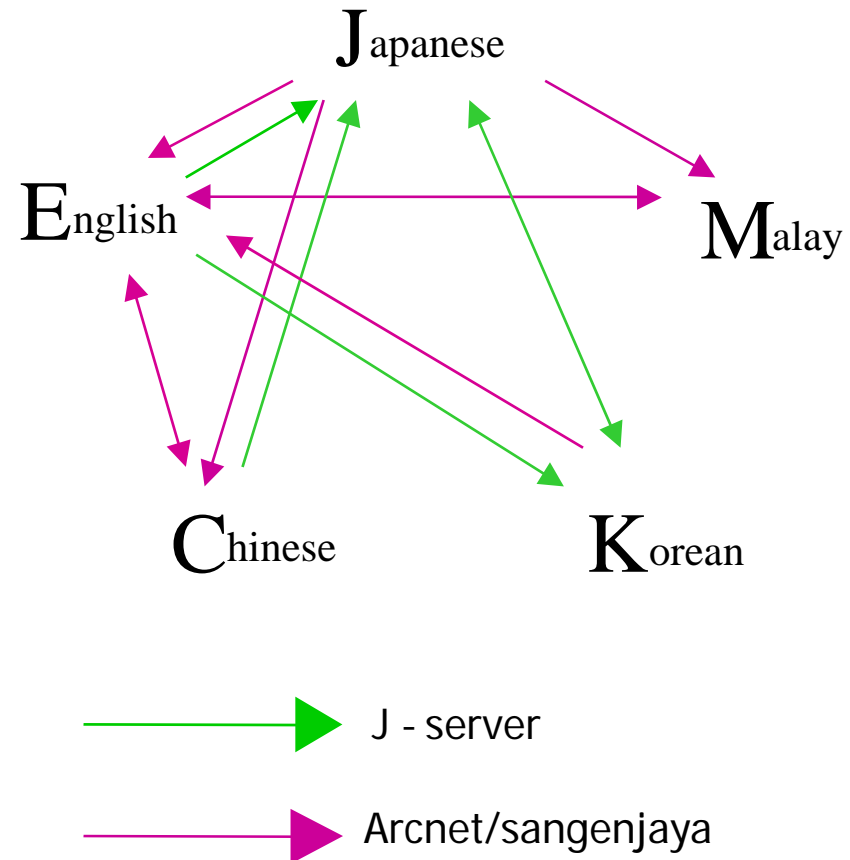
Subject: We want to know whether you got my message(japana-148)

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(retranslate)
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使って投稿してください。また、(japana-148)に対するあなたの意見も投稿して欲しい。	We want to know whether you could understand the my contribution japana-148 or not. We, please contribute using your language the part that you understood. I want us to also contribute you to japana-148's opinion.	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,对(japana-148)的你的意见也请投稿。	당신이 나의 투고(japana-148)을 이해할 수 있었던 것인가 아닌가를 우리들은 알고 싶다. 당신이 이해한 부분을 당신의 언어를 써서 투고해 주십시오. 또한, (japana-148)에 대한 당신의 의견도 투고해주었으면 싶다.	Kami hendak knows whether anda, kamu, dapat telah understand japana-148, (sumber saya, or not. Kami, t contributes using language anda, kamu, awak bahagian yang anda, awak telah understand Saya mahu kami also contributes pendapat kamu, awak) japana (

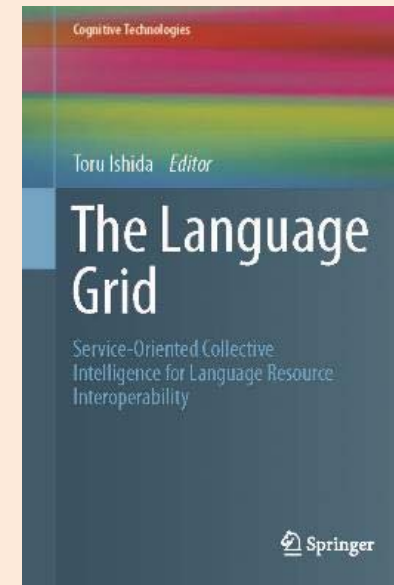
Translation Pentagon for ICE2002

- Hard to collect translation engines to cover five languages.
- Hard to understand their contracts.
- Hard to evaluate their services.
- Hard to get budget to cover expenses for machine translations.
- Hard to customize provided services.



Part 2

The Language Grid Project



We believe that *fragmentation and recombination* is the key to creating a full range of customized language environments for different types of user activities.

The Language Grid



Language Support for Multicultural and Global Societies

Language Grid

Sharing language resources such as dictionaries and machine translators around the world



From Language Resources to Language Services

Dictionary (Data)



Dictionary

Wrapping



Dictionary Service



避難場所

Accurate
word
translation

disaster shelter

Parallel Text (Data)



Parallel Text

Wrapping



Parallel Text Service



避難場所は、家から近い学校です。

Approximate
translation

In the case of disaster, people should be evacuated to a school nearby their house.

Machine Translation (Software)



Machine Translator

Wrapping



Machine Translation Service



避難場所は、家から近い学校です。

Low quality
high speed
translation

Disaster shelter is school close from a house.

Human Interpreter (Human)



Human Translator

Wrapping



Human Translation Service



避難場所は、家から近い学校です。

High quality
low speed
translation

Your disaster shelter is the school closest to your house.

Participants and Services

- **Participants (22 countries, 170 groups)**

- **University/Research Institute**

- Kyoto Univ. (Japan), Univ. of Indonesia, ITB (Indonesia), Shanghai Jiao Tong Univ. (China), Univ. of Stuttgart (Germany), IT Univ. of Copenhagen (Denmark), Princeton Univ. (U.S), DFKI (Germany), CNR (Italy), Chinese Academy of Sciences (China), NECTEC (Thailand), and more.

- **NPO/NGO/Public Sector**

- NGOs for disaster reduction, NGOs for Intercultural exchange, Public Junior-high schools, City Boards of Education, and more.

- **Corporate (CSR activities/language resource providers)**

- NTT, Toshiba, Oki, Google, Kodensha, Translution, and more.

- **Language Services (225 services)**

- **Machine Translator**

- J-Server (ja/en/ko/zh), Web-Transer (ja/en/ko/zh/fr/de/it/es), Toshiba (en/zh), Parsit (en>th), Google Translate (51 languages), and more.

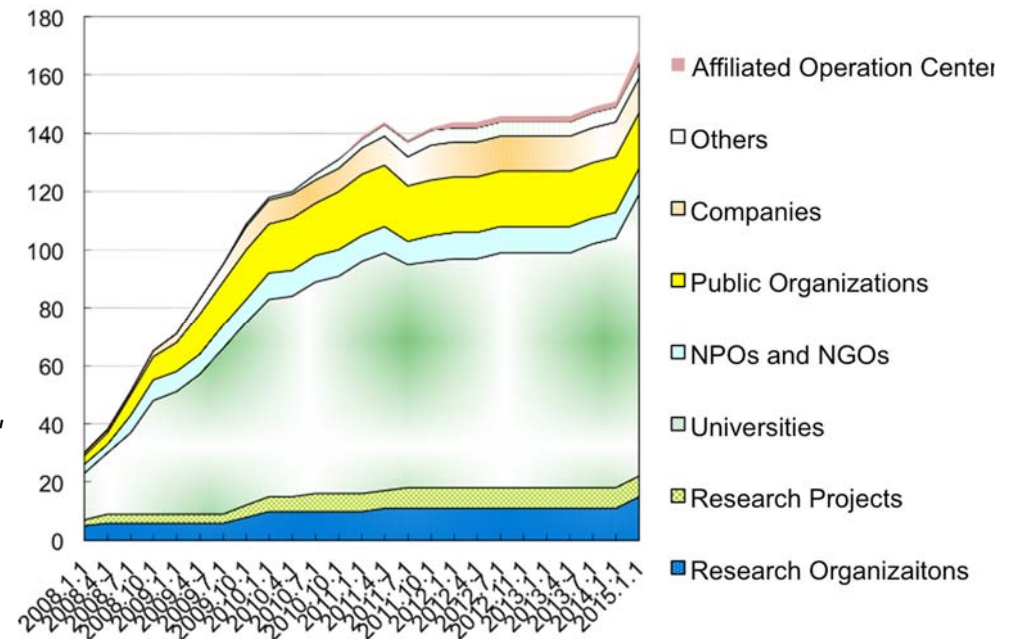
- **Bilingual Dictionary, Concept Dictionary**

- EDR , Wordnet, Life Science Dictionary, Multi-language Glossary on Natural Disasters, and more.

- **Parallel Text**

- **Morphological Analyzer**

- **Dependency Parser**



Operation from December 2007

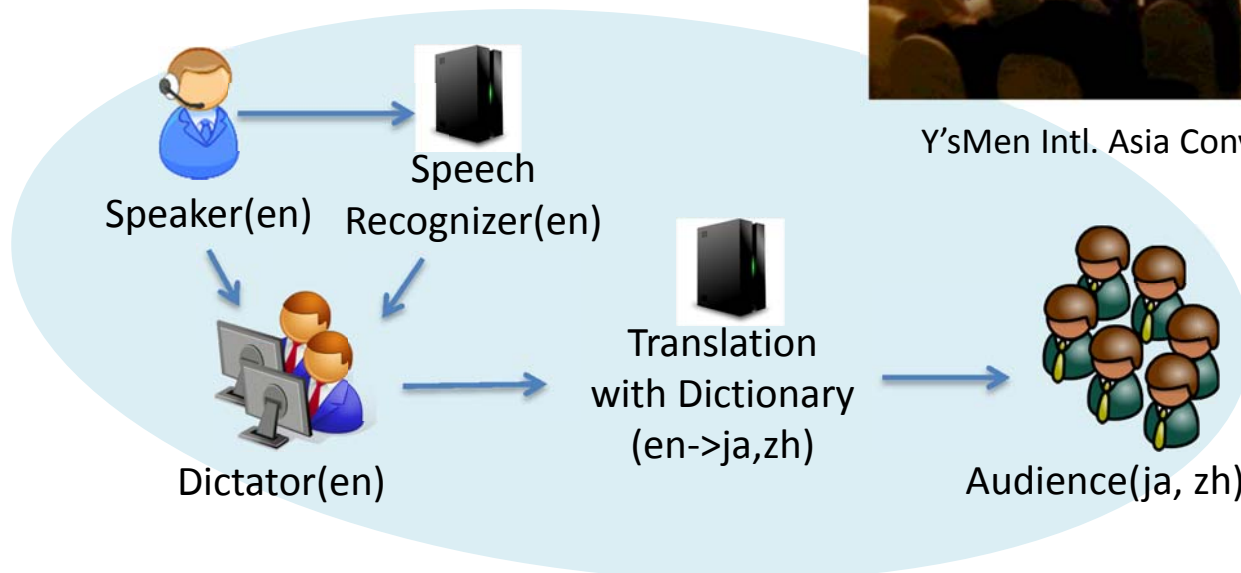
Conference Support System by Human-Machine Cooperation

Enhance translation quality by human-machine cooperation

- Automatic speech recognition helps human dictators.
- Dictators modify the text so that machine translators can create better results.
- Translation service cooperates domain specific dictionaries.



Y'sMen Intl. Asia Convention 2015 (1000 participants)



Software is Available!

- Service Grid Server Software:
 - <http://servicegrid.net/oss-project/>
 - Source code repository:
<http://sourceforge.net/projects/servicegrid/>
- Language Grid Multilingual Studio:
 - <http://langrid.org/developer/>
 - Source code repository:
<https://github.com/langrid/langrid-php-library>
- Language Mashup:
 - Will release from Google Play & App Store



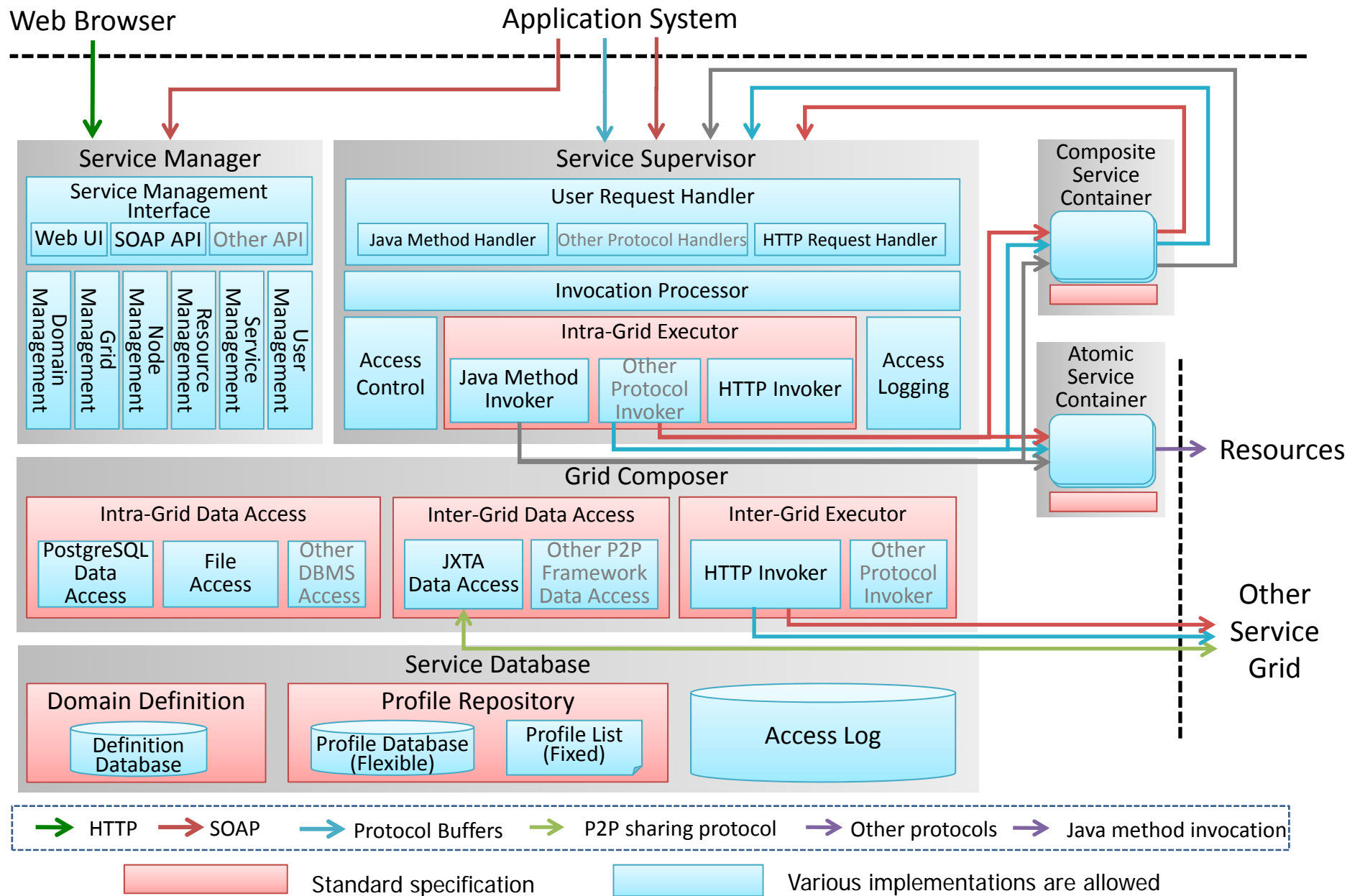
SERVICE GRID
Open Source Project Community Site

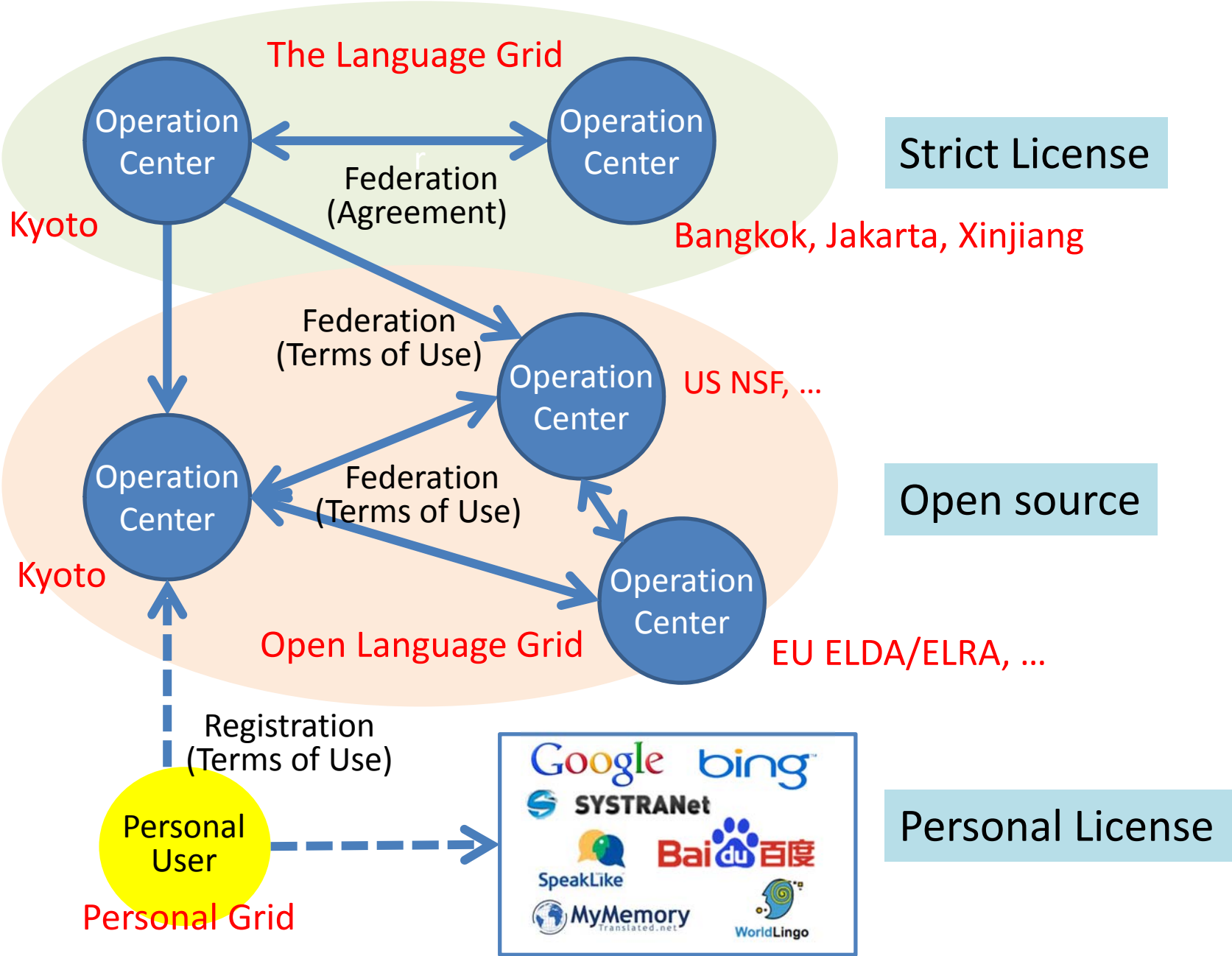


Language Grid
Multilingual Studio

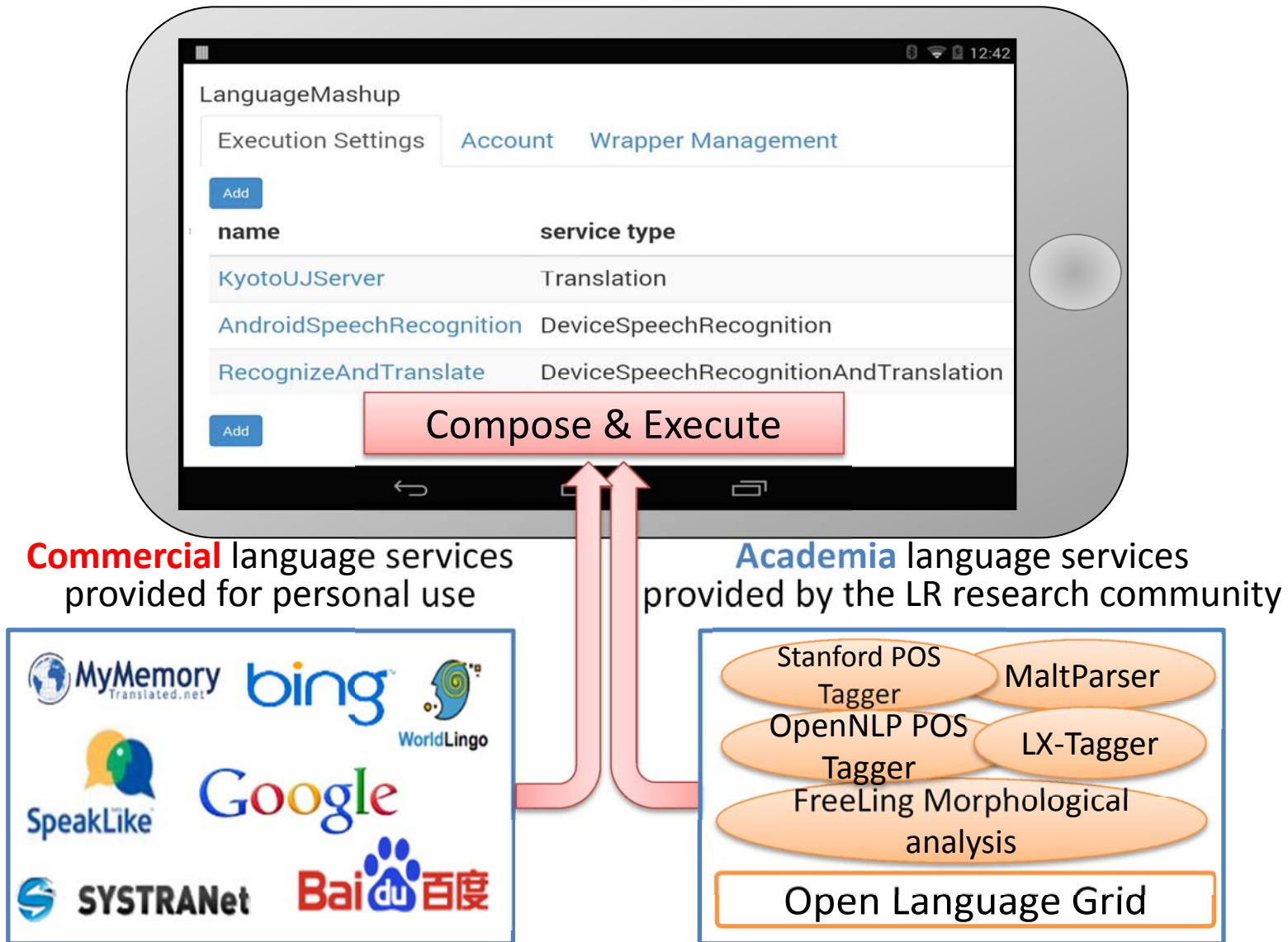


Language Grid Software (Available on Sourceforge)





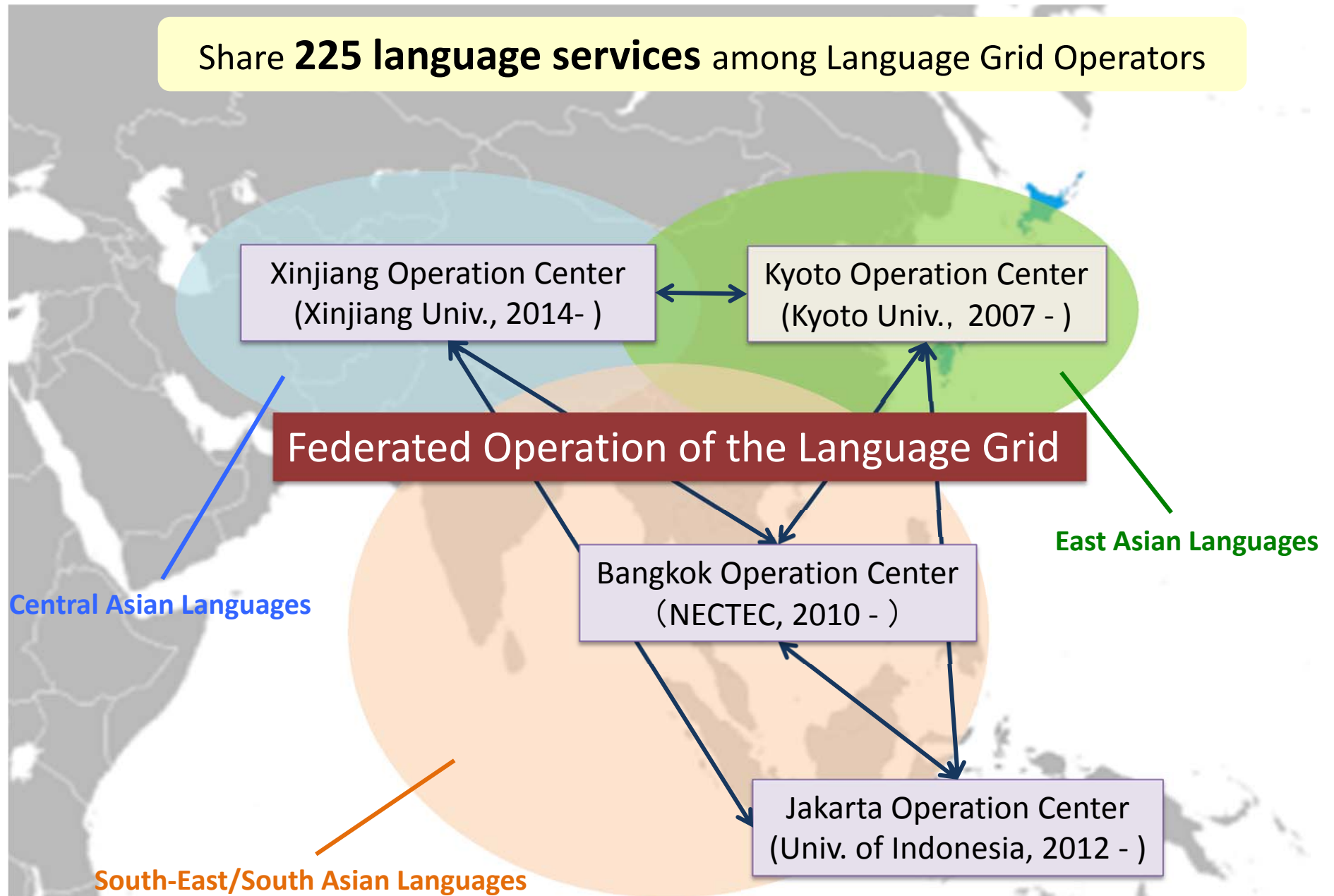
Language Mashup: Personal Grid for Language Services



■ Android version and iOS version will be released.

Language Grid Operation Centers

Share **225 language services** among Language Grid Operators



International Collaboration



Philadelphia
Univ. of Pennsylvania



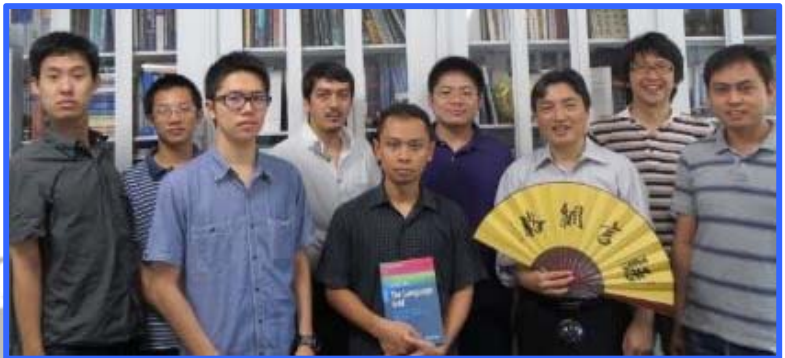
New York
Vassar College



Paris
ELDA



Xinjiang
Xinjiang Univ.



Barcelona
Universitat Pompeu Fabra



Boston
Brandeis Univ.



Pisa
CNR

Athens
ILSP



Nanjing
Southeast Univ.



Bangkok
SIIT



Ho Chi Minh
VNU-HCM



Jakarta
Univ. of Indonesia



Kyoto
Kyoto Univ.

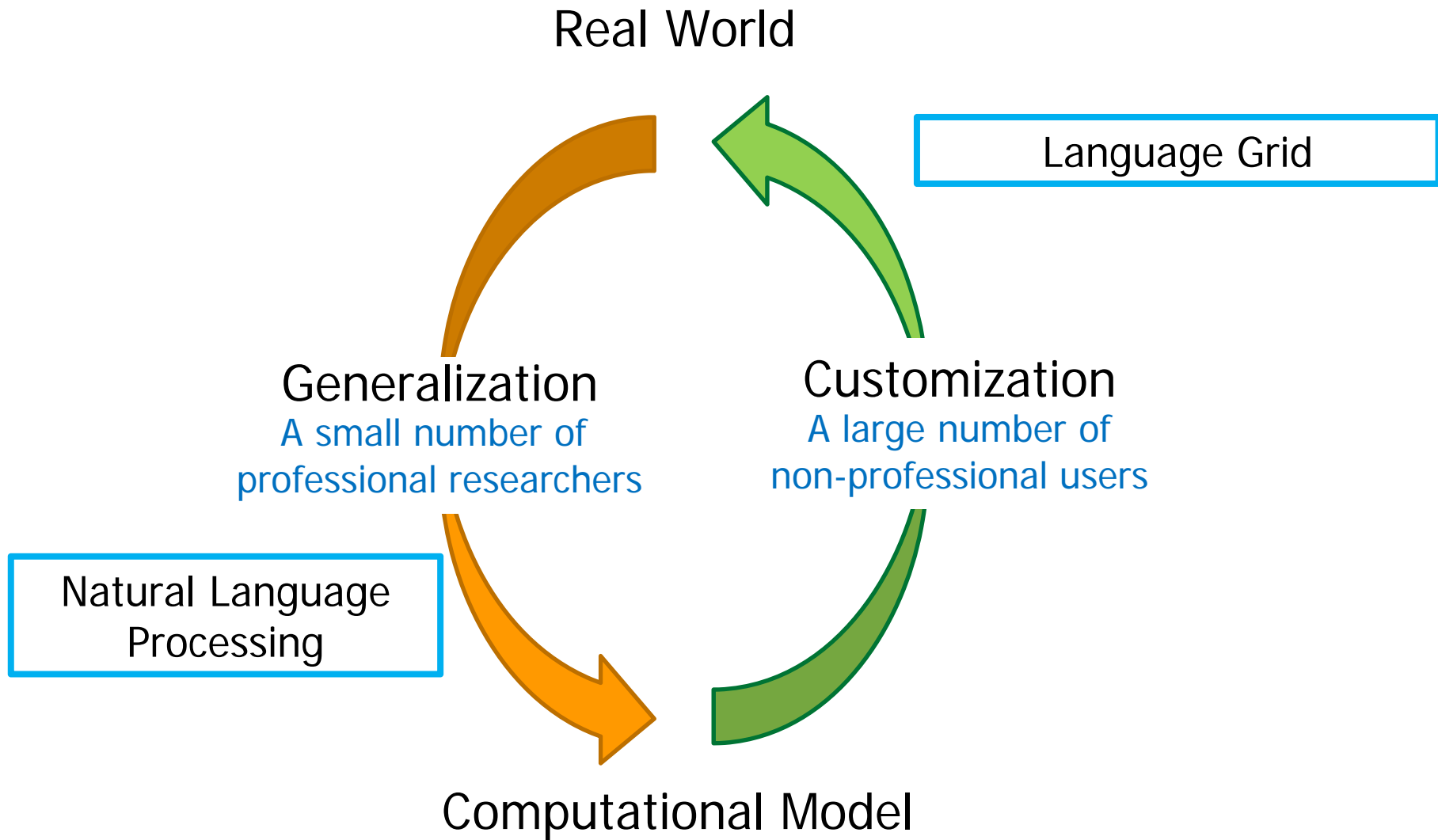
Toyohashi
Toyohashi Univ. of Tech.



Shizuoka
Shizuoka Univ.



Role of the Language Grid



Part 3

Intercultural Collaboration Project in Vietnam

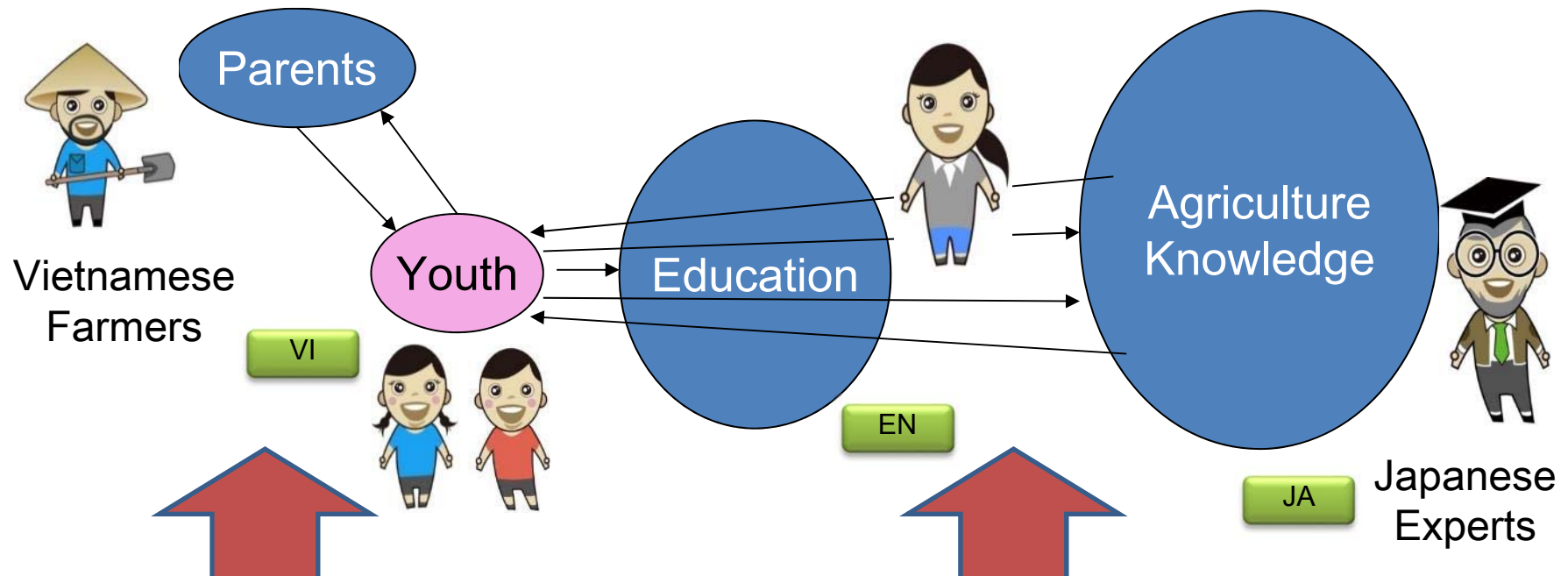


Tien My Commune, Tra On District, Vinh Long Province, Vietnam, Jan 5th, 2013

Farmers have difficulties in reading/writing messages.
A youth-mediated communication (YMC) model was invented,
where the children act as mediators between their parents and experts.

Agriculture Support Project in Vietnam

Our Idea is to Realize Agricultural Knowledge Communication between Japanese Experts and Vietnamese Farmers via Youth



Other Media Includes

- Mobile Phones
- Sensors
- Passports
- Recipe Cards

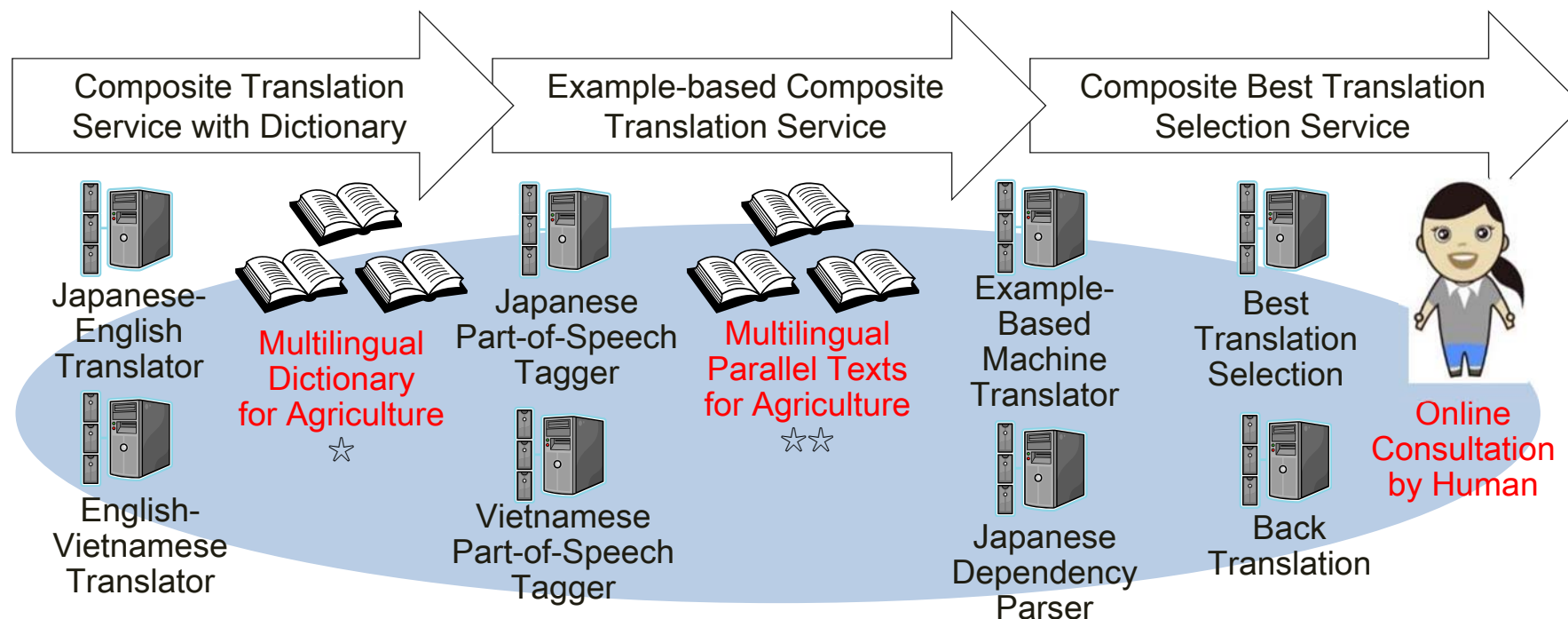
Available Language Services

- Japanese-English-Vietnamese Translation
- Agricultural Dictionaries
- Agricultural Parallel Texts
- Vietnamese Speech Synthesis

Combining Language Services



Japanese: たんぼの準備として、田起こしや代掻き、あぜぬりをして下さい。
 Vietnamese: Chuẩn bị đất là kết hợp giữa canh tác đất, cày bừa và đắp bờ.
 (English: Land preparation is a combination of tillage of the soil, puddling and levee painting.)



★ Multilingual Dictionary for Agriculture ([YMC Rice Dictionary](#), [Japan Agriculture Dictionary](#)) is provided by NPO Pangaea, Japan National Agriculture Research Center, Vietnam MARD. Entry Number: 3,099 (Sep. 2014)

★★ Multilingual Parallel Texts for Agriculture ([YMC Rice Parallel Text](#)) is provided by NPO Pangaea, Japan National Agriculture Research Center, Vietnam MARD. Entry Number: 2,485 (Sep. 2014)

Overview of YMC-Viet Project

YMC-Viet is a Project Organized by Different Stakeholders

- Organizations

- Kyoto University: Multilingual Communication
- NPO Pangaea: Education, Activities
- University of Tokyo, Mie University : Agricultural Knowledge
- Vietnam National University : Local Arrangements
- MARD, DARD: Planning and Controlling Experiments



- Schedule


- 2011/02 ~ 2011/03 1st Experiment (Thien My, Vinlong Province)
- 2012/10 ~ 2013/01 2nd Experiment (Thien My, Vinlong Province)
- 2013/09 ~ 2014/01 3rd Experiment (Thien My and Dong Thanh, Vinlong Province)
- 2014/02 ~ 4th Experiment (Dong Thanh, Vinlong Province)




Example of Experiment Schedule

Each Experiment Takes 4-5 Months

ID:131



これがたんぼにいます。




Yaz
 場所:
 専門分野:
 やりたいこと:

こんにちは、Chocon。この卵はスクミリンゴガイです。
 どうすれば良いかファシリテーターに聞いてください。

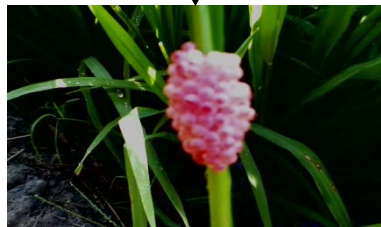
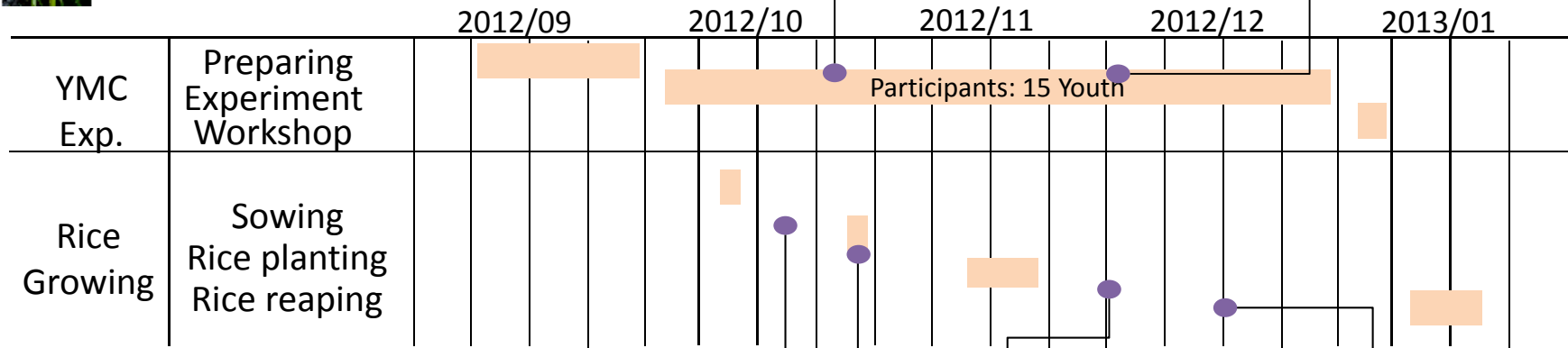
02/12/2012
 11:13

Status



Bình
 Tuổi: 11
 Việc em thích làm:
 Người em thích: Cha

Hạt giống lộ ra khỏi mặt đất. Tôi nên làm gì?



2012/10/19 Chocon



2012/10/29 Phuc



2012/12/1 Bang



2012/12/15 Phuc



“In Vietnam, more than 1 Million farmers need a support from the YMC project.”
Dr. Nguyen Viet Chien (Vietnam MARD)



アナログ

お問い合わせは NHKアナログ終了コールセンター 0570-07-2011 (毎日 午前9時～午後6時)

User Interface for Youth

The interface is designed for youth and features a central content area with two posts. The top post is by user 'Chuôt' (02/03/2011 19:57) and asks for advice on bird damage. The bottom post is by user 'Ikeda' (03/03/2011 18:05) and provides detailed advice on bird damage prevention, including a 'RecipeCard' for 'Orange-8'. Below the posts is a translation tool with a '翻訳' button and a '140' character limit. A central robot character is positioned between the translation tool and a language dropdown menu set to 'English'. At the bottom right, there is a 'コメント投稿' (Comment Post) button. The interface is decorated with a blue border featuring cartoon characters of children and a robot.

Chuôt 02/03/2011 19:57
鳥による被害をなくしたいです。どうしたらよいですか？
Create a voice sound.

Ikeda 03/03/2011 18:05
スズメ等の被害を完全になくすのは、難しいです。昔から、田んぼにかかしを立てたり、カラスに似せた黒い布やビニールをひもで吊るしたりしますが、どれでけ効果があるかは分かりません。
鳥を来なくするためには、鳥の餌となる生ごみの多い場所や、畜産施設、生息地域から離れたところで稲作を行いましょう。ゴミの管理が大切です。
鳥による被害を回避するには種もみを地中深さ1cm位に埋めることで隠す、灌水を行うことによって被害が回避できますが、落水後種もみが地表に露出していると被害にあう可能性があります。
RecipeCard: オレンジ-8
Create a voice sound.

翻訳 140 English

コメント投稿

User Interface for Youth



User Interface for Expert

ymc viet

メッセージ閲覧 フィールドデータ

日本語 ▾ Ikeda ログアウト

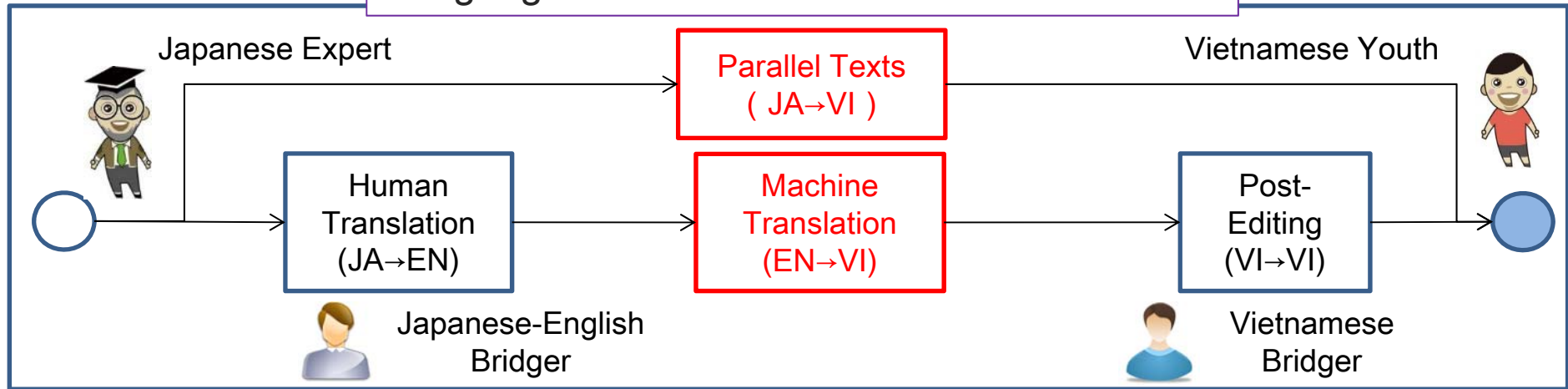
Username

Password

Login

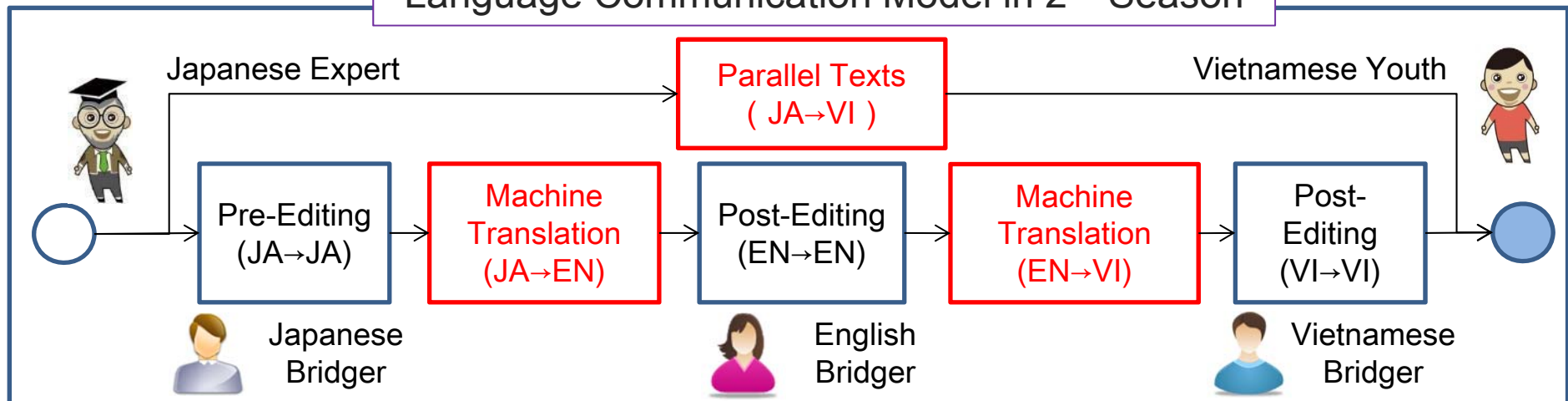
Language Communication Model

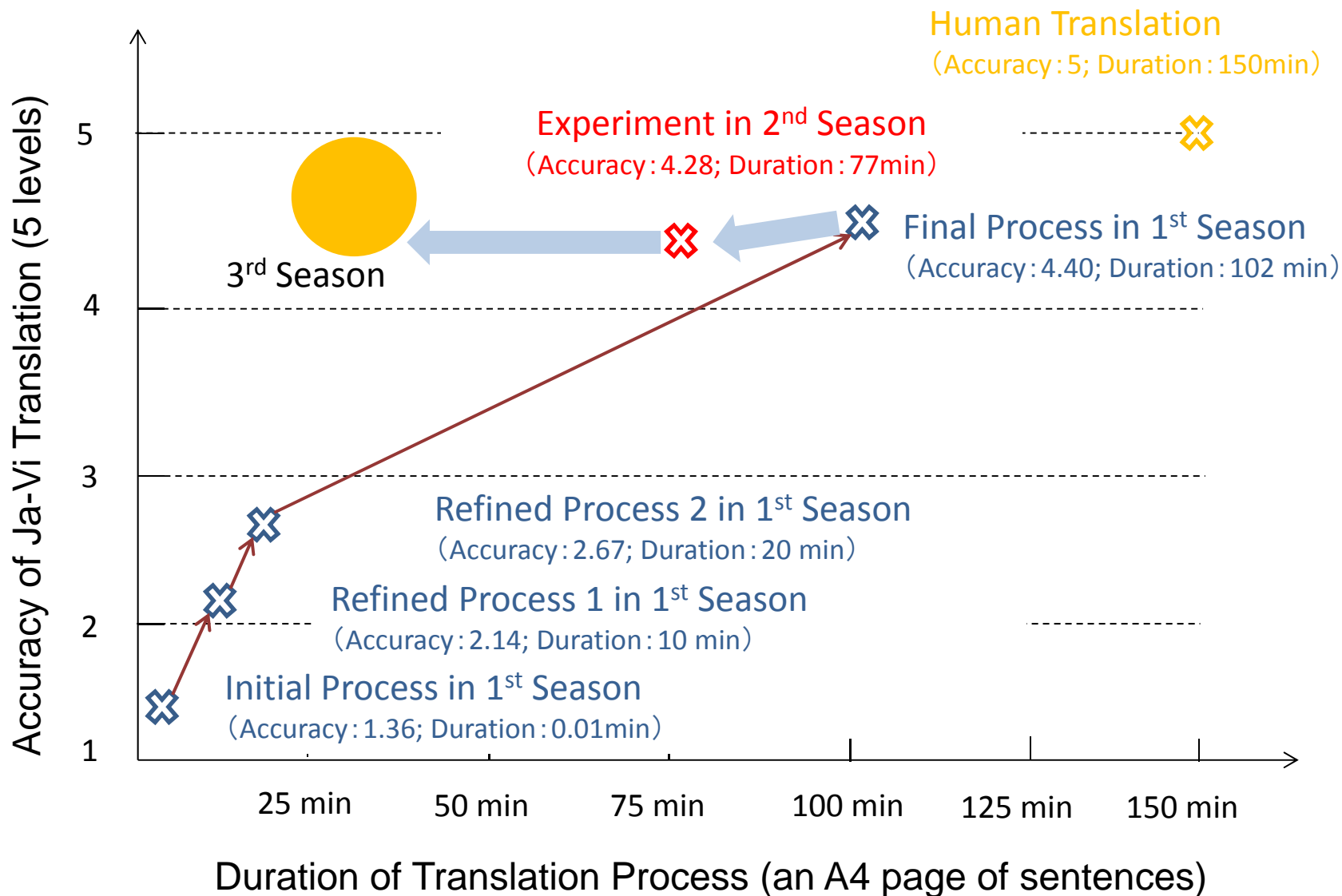
Language Communication Model in 1st Season



- Enabling monolingual bridger for all languages
- Improving agricultural dictionaries and parallel texts

Language Communication Model in 2nd Season





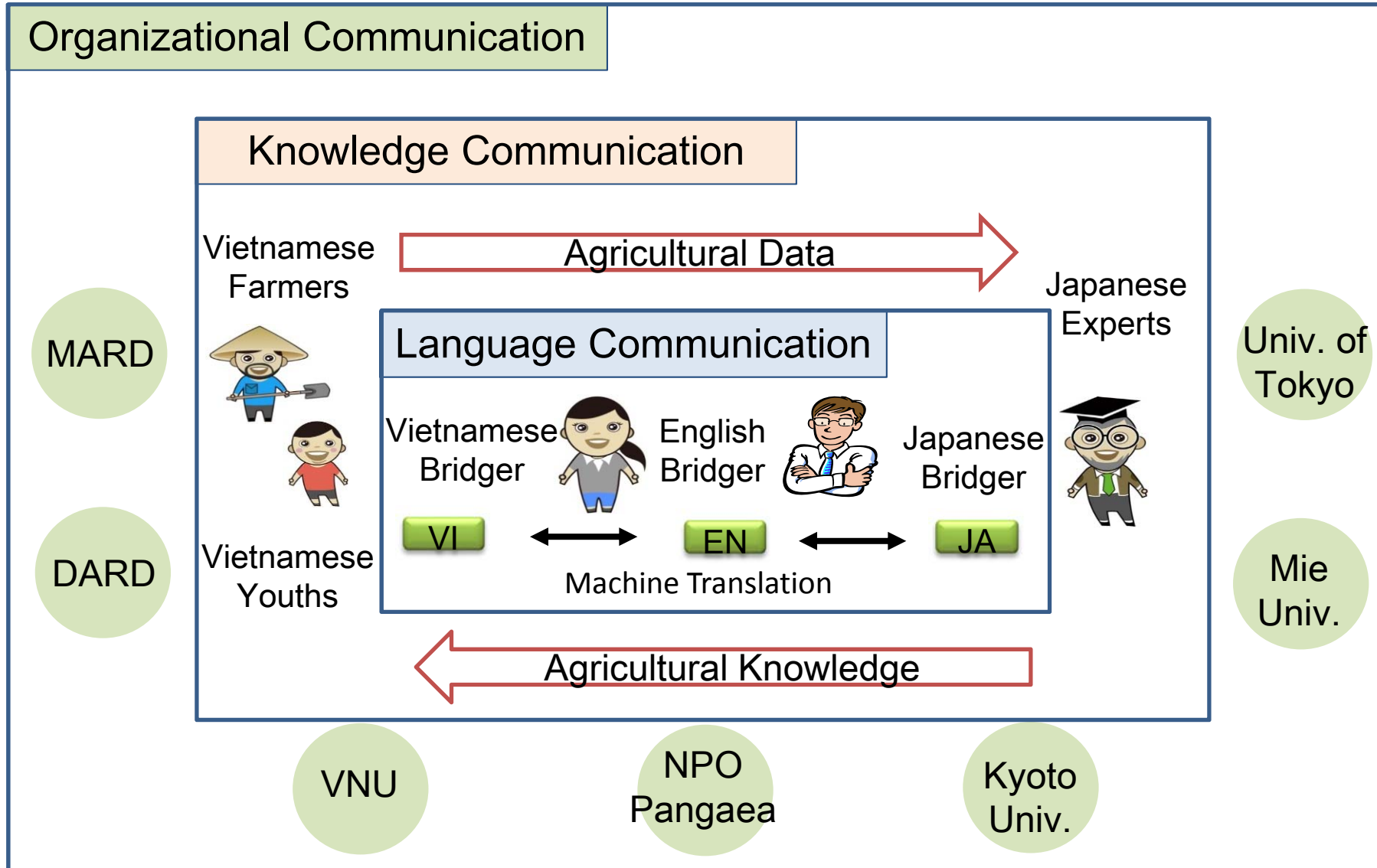


This project aims at increasing the yield of rice. One project member, an agricultural expert, recognized the lack of nitrogen in this particular rice field, and suggested farmers to use fertilizer a little more. The farmers did not follow this advice, since they believe bugs will gather from neighboring fields, if they increase the amount of fertilizer.



One of the goal of this project is to reduce the mount of agrichemical.
Employees of the local government heavily involved in the project
organization to help farmers' children to learn ICT.
Later, it appears they are selling agrichemical to farmers as their side business.

Three Layer Communication Model



Part 4

Ongoing PhD Research



Laboratory for Global Information Network
Department of Social Informatics, Kyoto University

Ongoing Research

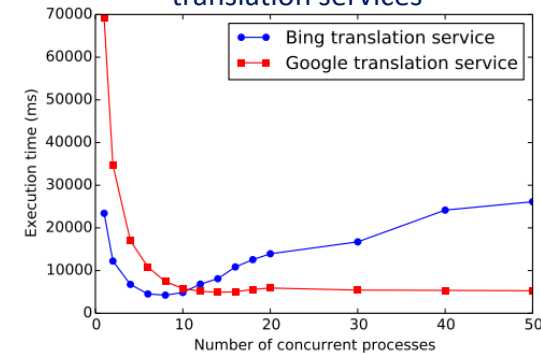
- **Human Interaction Analysis**
 - Creating conversational common ground machine translations (CSCW 2006, CHI 2009)
 - Collaborative translation by monolinguals with machine translators (IUI 2009)
 - Analysis on multilingual collaboration (CSCW 2013)
 - Agent metaphor for machine translation mediated communication (IUI 2013)
 - Open smart classroom (IEEE TKDE 2009)
- **Service Selection and Composition**
 - Horizontal service composition (ISWC 2006)
 - Predicting and learning executability of composite web services. (ICSOC 2008)
 - Reputation-based selection of language services. (SCC 2011)
 - User-centered QoS computation for web service selection (ICWS 2012)
 - Dynamic service selection based on context-aware QoS (SCC 2012)
 - Policy aware service composition (SCC 2015 best paper, IEEE TSC 2015)
 - Dynamic sliding window model for service reputation (SCC 2015)
- **Service Architecture**
 - A hybrid integrated architecture for language service composition (ICWS 2008)
 - Service supervision (ICWS 2009, SCC 2010)
 - Market-based QoS control for voluntary services (SCC 2010)
 - A service binding framework for open environment (SCC 2012)
 - Service Grid Federation Architecture for Heterogeneous Domains (SCC 2012)
 - Cascading failure tolerance in large-scale service networks (SCC 2015 , IEEE IoT 2016)
- **Language Resources and Services**
 - Emotions and pictograms in language services (ESWC 2008)
 - Context aware service composition for pivot translations (IJCAI 2009)
 - Constraint approach to context based word selection (IJCAI 2011)
 - Pivot-based dictionary creation for low-resource languages (ACM TALLIP 2015)

Policy-Aware Optimization of Parallel Execution of Composite Services

Research Problem

- Service providers employ policies to control parallel execution of their provided services based on arbitrary decisions
 - Different providers may have different parallel execution policies
- Service users, cannot change the service providers' policies
 - Need to adapt services' invocation to the services' policies
- A composite service is combination of different atomic services
 - Need to consider all the atomic services' policies to determine the optimal degree of parallelism (DOP) of the composite service.

Example: Different parallel execution policies of real translation services



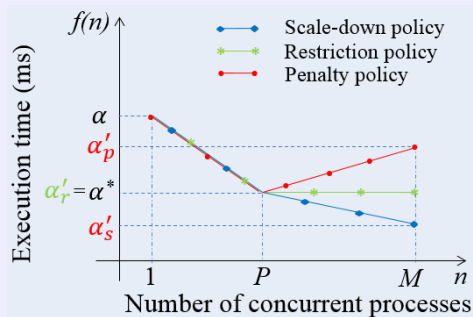
Existing Approaches

- Existing studies on parallel execution in HPC consider trade-offs among several criteria, e.g. execution time, energy consumption, and resource usage to optimize parallel execution of applications (Freeh *et al.*, 2007; Curtis-Maury *et al.*, 2008).
 - Existing approaches on composing composite service which can attain the optimal QoS (Zeng *et al.* 2004; Canfora *et al.*, 2008)
- Existing approaches fail to consider atomic services' policies in determining the optimal DOP of composite services.**

Our Solution

Parallel Execution Policy Model:

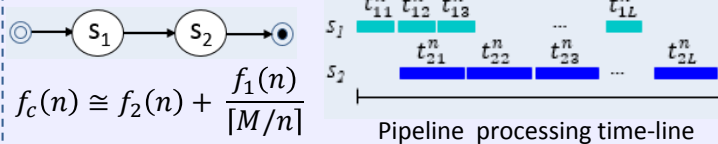
- Define three types of parallel execution policies: Scale-down, Restriction, and Penalty



- Estimate execution time of an atomic service with number of concurrent processes: $f(n)$

Prediction Model:

- Use data parallelism and pipeline execution to execute composite services
- Embeds the service policies into formulae to calculate execution time of composite services.
- E.g. Sequential structure:



Generally:

$$f_c(n) \cong \max_{i \in [1, k]} f_i(n) + \frac{\sum_1^k f_i(n) - \max_{i \in [1, k]} f_i(n)}{[M/n]}$$

- Create different formulae for different structures: Sequential, Parallel, Conditional, and Loop.

Finding the optimal DOP:

- The optimal DOP: Where the composite service attain the best performance improvement.
- From the formulae and calculation, the optimal DOP is estimated.

Experiments are conducted on real-world translation services:

Our proposed model has good prediction accuracy in identifying the optimal DOPs



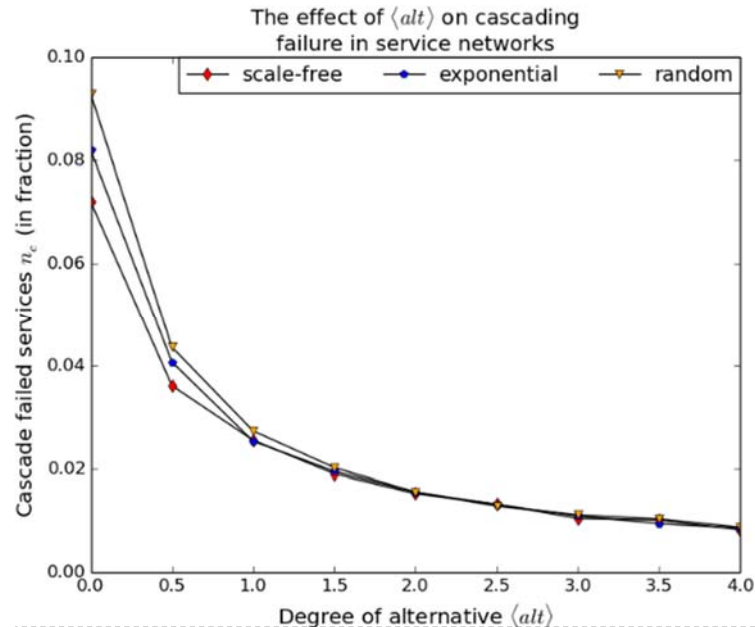
Cascading Failure Tolerance in Large-Scale Service Networks

Research Problem

- Cooperating **services in large-scale service networks**, e.g. services in the Internet of Things, **are interdependent** to each other and potentially experience cascading failure.
- **Cascading failure** in service networks occurs when failure of one service causes its dependent services to be also fail. For example, in service composition, failure of a component service may cause the composite service becomes unable to function properly.
- Understanding cascading failure is important for designing tolerant service networks.
- Service network tolerance to cascading failure is determined by the **network topology** and the **services interdependency**.
- There is a **lack of research** addressing cascading failure in service networks.

Existing Work

- **Load-based cascading failure** has been widely addressed in power network
- In power network, failure propagation is triggered by **load-redistribution**. This is contrast to cascading failure in service network, where the propagation is driven by the **dependency between services**.
- **Random** network topology provides **better tolerance** to load-based cascading failure than the other network topology (scale-free and exponential network topology)



Our Contribution

Our important findings are as follows:

- **Scale-free** topology provides **better tolerance**, followed by exponential and random topology.
- The effect of **network topology** on tolerance is more significant on **lower degree of alternative** (the average number of alternate services for each component service).
- The number of nodes experiencing **cascading failure** increases as the **inverse of the degree of alternative**.
 - Adding few alternate services for each component service could significantly improve the tolerance if *degree of alternative* is currently low (e.g. *degree of alternative* < 3).
- The number of nodes experiencing **cascading failure** is somewhat **linear** to the average number of component services (**degree of dependency**)

Kemas Muslim Lhaksmana, Yohei Murakami and Toru Ishida. Analysis of Large-Scale Service Network Tolerance to Cascading Failure. IEEE Internet of Things Journal, 2016.

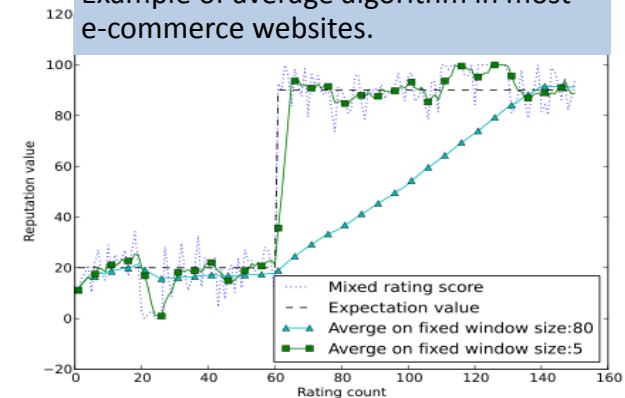


Dynamic Sliding Window Model for Service Reputation

Research Problem

- Service consumers have no or little direct experience to select a service provider among the functionally equivalent services.
 - Reputation system is proposed as an assistant tool.
- However, reputation system may vulnerable to unfair ratings.
 - Adulating ratings and defaming ratings can cause the reputation value increase or decrease deliberately.
- Reputation value should reflect the latest dynamic changes
 - Especially in e-commerce, the time lag problem is vital for transaction.

Example of average algorithm in most e-commerce websites.



Existing Approaches

- Bayesian reputation system: The reputation value is updated and learned from behavior history based on the beta distribution. (Jøsang *et al.*, 2002, Teacy *et al.*, 2012).
 - Vulnerable to unfair rating if malicious users are majority.
- Olfaction-based algorithm(OACR): Dynamically assign weight on unfair ratings. (Wu *et al.*, 2013)
 - Fixed unfair rating pattern is assume to detect and mitigate the unfair rating.

Our Solution

• Dynamic Sliding Window Model (DSW):

- Dynamically detect the appropriate window size using linear regression algorithm on the ratings.
- **Modelling the ratings:** Suppose we have the observed ratings r , the predicted reputation value $f(x) = kx + C$ can be evaluated as:

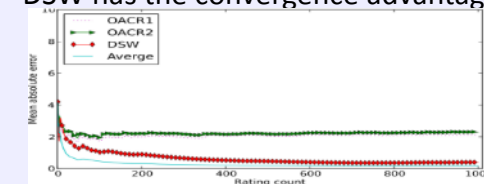
$$k = \frac{\partial \ln p(k, C | r)}{\partial k}; C = \frac{\partial \ln p(k, C | r)}{\partial C}$$
- **Filtering out unfair rating:** Base on $f(x)$, the deviation of next received rating is calculated and unfair ratings will be detected as small probability event.

• Rating-ratio based DSW Model:

- Nearly 50%~60% users on eBay did not leave their feedbacks[Cabral *et al.*, 2010]. While malicious users tend to give every rating after transaction.
- By calculating the rating ratio in a suspicious rating time span to detect the unfair rating.
- Can relax the limitation that fixed unfair rating sequences are required to detect unfair rating in OACR.
- **Evaluated the accuracy and robustness on various behavior patterns:**
 - **Constant:**(Zaki *et al.*, 2009, Vogiatzis *et al.*, 2010).
 - **Linear and pairwise:**[Wu *et al.*, 2013, Zaki *et al.*, 2009, Vogiatzis *et al.*, 2010, Sabater *et al.*, 2001).
 - **Sinusoidal:**(Vogiatzis *et al.*, 2010).

• Convergence:

- Compared with OACR algorithm, DSW has the convergence advantage.



Our proposed model outperforms the existing reputation system by 40% on average.

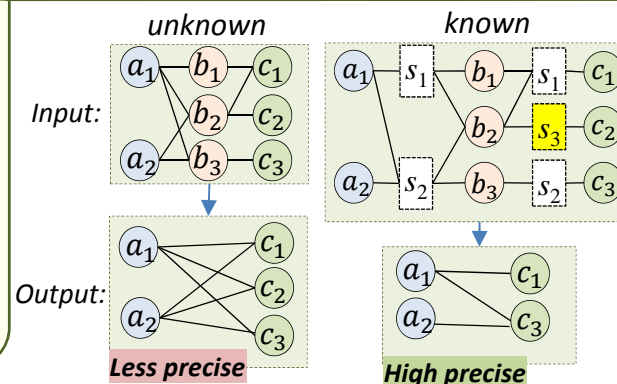


An Optimization Approach to Pivot-based Bilingual Dictionary Induction

Example: A Pivot-based induction when word senses are:

Research Problem

- Bilingual dictionaries are essential for many tasks in NLP and IR
 - but available only for well-resourced languages such as English and Chinese.
- Traditionally extracted from a large amount of parallel corpus (Wu, 1994).
 - however, parallel corpus remain scarce and are expensive to obtain.
- Pivot-based induction is an alternative for less-resourced languages:
 - Induces a new dictionary from existing two using a pivot language.
 - Precision is often low due to ambiguity problem caused by polysemy words.



Existing Approaches

- Word sense data are obtained from additional resources such as WordNet (Istvan and Shoichi 2009).
- Parallel corpus, comparable corpora, POS tagger or structures of input dictionaries are utilized to measure the semantic distance between a cross-lingual word pair (Kaji 2008; Shezaf 2010; Saralegi 2011; Bond 2008; Tanaka 1994).

Most approaches require an additional resource or have low performance because the data incompleteness is not considered.

Our Solution

Intra-family languages:

- Share large parts of their lexicons.
- 44% ~ 94% among Turkic languages.

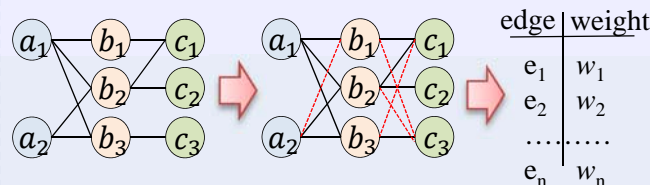
%	Kyrgyz	Kazakh	Uzbek	Uyghur
Kazakh	92			
Uzbek	82.9	82.8		
Uyghur	83.8	81.9	86.3	
Turkish	64.9	64.8	67.2	66.7

A partial Lexicostatistical Matrix of Turkic Languages
(<http://turkic-languages.scienceontheweb.net>)

Assumption:

Lexicon of intra-family languages are one-to-one (1-1) mapping.

- Realize the assumption as a set of constraints.
- Find missing meanings and measure their chance of being missing as weights.



A statistical calculation based on structure

Modeling as an optimization problem:

Objective

- Find as many as one-to-one mapping pairs.
- Minimize the addition of incorrect meanings.

Subject to

- A 1-1 pair is unique and share same meanings.

SAT-based formalization + free solver

Our method outperforms a baseline method by 10% more precision under similar recall.

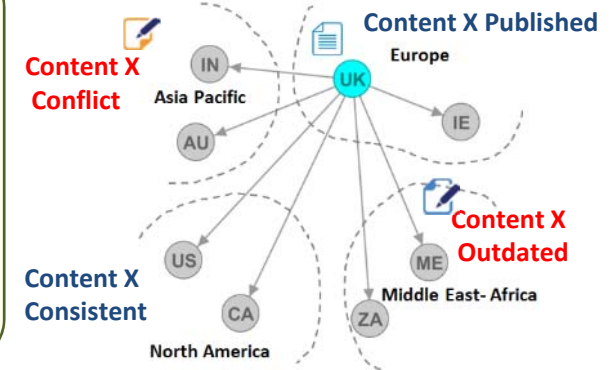


Information Sharing Among Countries A Perspective from Country-Specific Websites in Global Brands

Research Problem

- Global brands publish huge volumes of content collaboratively in several languages (both shared and unshared languages) in country-specific websites.
- Problems in Information Sharing among countries via country-specific websites arises from
 - Presence of outdated content, conflicting content.
 - Some countries have common language, some are multilingual so inconsistency occur even between same language or beyond it.

Example: Inconsistency in Cross Site Content



Existing Studies and Hypothesis

- Several Categories of content published in websites that vary with cultural groups. (Robbins *et al.*, 2003)
- Globalization, Glocalization, Localization are current trend in design (Singh *et al.*, 2014) *but what about content categories*

Hypothesis 1. Propagation among country-specific websites is constrained by content categories: Corporate Information, Product Information, and Customer Support Information.

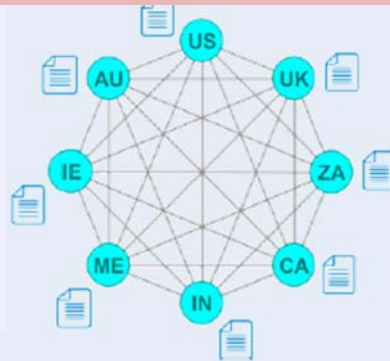
Method

Webpage Samples of content categories collected from Global Brand
- Corporate Information
- Product Information and Customer Support Information

Propagation among websites compared
- Complete Graph
- Website Pair

- Global Brands selected are ranked highly in globalization web report card.
- Countries from each geographic region is chosen

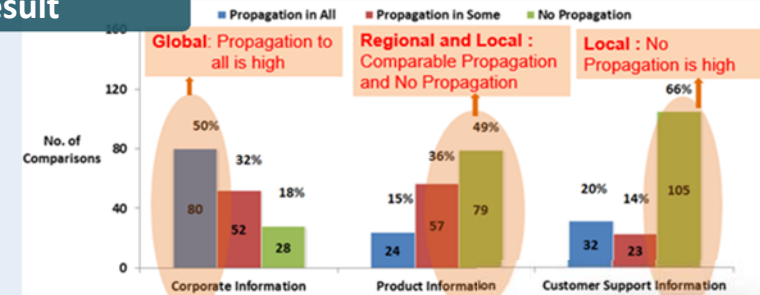
Complete Graph: Total Comparison: 480



Website Pair: Total Comparison: 1680



Result



- Corporate and Customer support related information tend to be shared globally and locally while product related information is both locally and regionally suitable for sharing
- Corporate related information also showed tendency for high coupling between websites suggesting the suitability for rigid consistency policy



Crowdsourcing for Evaluating Machine Translation Quality

Research Problem

- The recent popularity of machine translation (MT) has increased the demand for the evaluation of translations
- Manual checking by a bilingual professional and automatic evaluation like BLEU have problems in cost or quality
- **MT evaluation are used in broad objectives, therefore various comparison scores are needed**

Evaluate Translation Quality

Guidelines:

- Please read a Chinese sentence and English translations.
- Please evaluate all the translations by adequacy.
- **Adequacy:** refers to the degree to which information present in the original is also communicated in the translation.
- Finally, please answer the questionnaires.

Evaluation Tasks:

Chinese Sentence	English Translation	Adequacy
障碍物的一个可行的实施方案为在鼓内的螺旋状障碍物(图1中未示出)。	One possible embodiment is within the drum of an obstacle (not shown in Fig. 1).	<ul style="list-style-type: none"> ⊙ 5 (All meaning) ⊙ 4 (Most meaning) ⊙ 3 (Much meaning) ⊙ 2 (Little meaning) ⊙ 1 (none)

Existing Approaches

- Ranking tasks of five translations are conducted by crowdsourcing workers. Results of five workers showed the 60% agreement with expert ranking data[Callison-Burch, 2009]
- Multimedia evaluation based on pairwise comparison offered the less cost an good quality evaluation[Chen, 2010]

Existing approaches are relative evaluations, not absolute evaluations; they did not analyzed crowdsourcing evaluation by translation or machine translation systems

Our Solution

Evaluation Scores

- We applied three scores to consider the possibility of crowdsourcing evaluation

Translation-score: absolute score for each translation

$$S_{translation}(e) = \{s_{11}(e), \dots, s_{mn}(e)\}$$

Sentence-score: relative evaluation between translations of the same source sentence

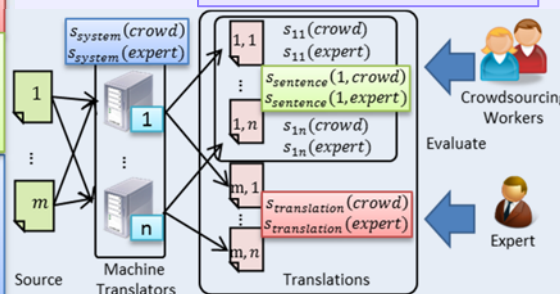
$$S_{sentence}(i, e) = \{s_{i1}(e), \dots, s_{in}(e)\}$$

System-score: average score of machine translation systems in the given dataset

$$S_{system}(e) = \left\{ \frac{\sum_{i=1}^m s_{i1}(e)}{m}, \dots, \frac{\sum_{i=1}^m s_{in}(e)}{m} \right\}$$

Experimental statistics

- 24 workers participated
- 23 translation systems
- 10 source sentences
- 4,347 evaluations in total



Experimental Result

Translation-score: Workers tend to assign higher scores to low-quality translations

Sentence-score: Large difference in source sentence and evaluation metrics

System-score: High correlation between crowd workers and expert

Experimental result showed 0.88 correlation in system score, which is promising to use crowdsourcing to select the best translation system



How Non-Native Speakers Perceive Listening Comprehension Problems

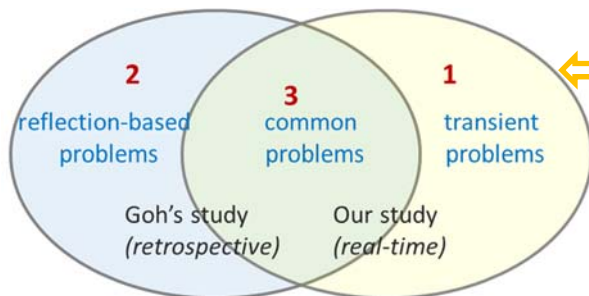
Research Problem

- Non-native speakers (NNSs) often face comprehension difficulties when listening to native speakers (NSs)' speech, particularly **in real-time** settings
- Support for NNS:
- providing speech translation (Wahlster2000)
- providing automated transcripts (Pan et al. 2009)
- adding artificial delays (Yamashita et al. 2013)

one technology is **not all-round**

providing multiple technologies may pose NNS **extra burden**

- Our goal: to design an **adaptive system** which automatically changes the type of support based on NNSs' changing needs.
- the **types** of comprehension problems faced by NNSs
- how these problems are **perceived**



Existing Approach

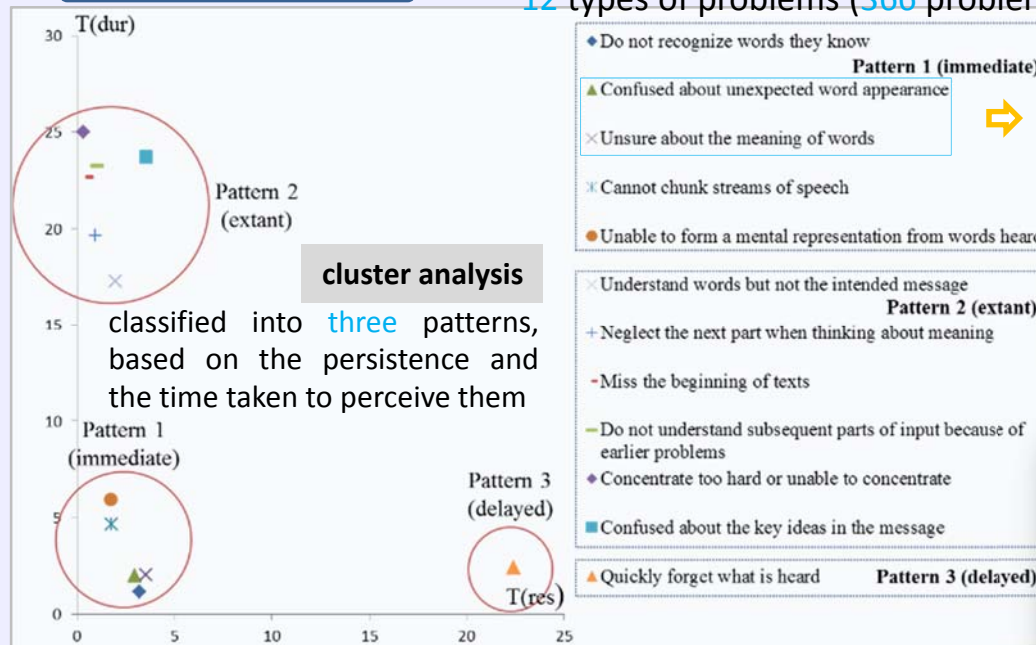
- identified listening comprehension problems faced by NNSs from a **cognitive perspective** through a **retrospective** approach.
- main data source: weekly diaries kept by 40 students (Goh 2000)

Method

- Participants
 - 40 NNSs (20 Japanese; 20 Chinese)
 - 20 NSs
- Procedure
 - During the task, participants **pressed a button** whenever they encountered a comprehension problem
 - Next, they explained each problem, **the point at which** they recognized the problem, and for **how long** it persisted

Findings

12 types of problems (366 problem occurrences)



two **newly** identified problems



- Although more than 700 languages are spoken
 - Numbers of speakers decreases in a half of these languages (341 languages)
 - Only 20 languages has 1 million speakers
(<http://www.ethnologue.com/>)
- UNESCO acknowledged 148 Indonesia languages as endangered languages
(<http://www.unesco.org/languages-atlas/>)
- Regional language speakers are rapidly decreased due to the education policy or migration to urban area

-> Protection and preservation of regional language is urgently needed

