The Language Grid for Intercultural Collaboration

Toru Ishida

Department of Social Informatics, Kyoto University
NICT Language Grid Project
Overview of This Talk

- This talk is on a new language service infrastructure on the Internet to combine existing language resources (machine translations, morphological analyzers, dictionaries etc.) to create customized language services, and to provide those language services for non-profit activities.
- We started the Language Grid project at NICT in April 2006.
- We are not working on language technologies. Our technologies are semantic Web services, multi-agent systems, and computer supported collaborative work (CSCW).
A Story of NPO Pangaeae


Universal Playground enables kids around the world to develop bonds despite differences in distance, languages and cultural backgrounds.

Picton: Pictograms to talk with kids in different countries.
How to support Pangaea facilitators using Korean, Japanese, English, and German translations with their own dictionary?
A Story of a Tunisian Researcher in Japan

Rescue Ahlem!

Help foreign researchers in Japanese meetings.

Ahlem is one of our project members, who is from Tunisia in North Africa. Although she can speak Arabic as her native language, and French and English, she cannot yet speak Japanese.

Three Japanese researchers input a summary of an ongoing discussion on the blackboard in Japanese. The input text is translated into English, and also in French, and displayed on Ahlem's laptop.
Rescue Ahlem!
All for One Collaboration Project

- All for One Collaboration Project at Kyoto University

- Help foreign students in Japanese seminars.

How to support Korean, Japanese, English, Chinese, French, and Indonesian people in this seminar in technological domains?
Language Grid as a solution


Role of the Language Grid

Use of language services is difficult

Language Resource
(Dictionary)

Language Processing
(Morphological Analysis)

Network Service
(Machine Translation)

Copyright

Price

Quality

Contract

Technology

Language Grid increases usability and accessibility

Language Services
Developed by Experts

Language Services
Developed by Communities

Rights Protection

Development Support
Architecture of the Language Grid

**Vertical Language Grid**
- Dictionary for interpreters of Nepali mountain climbing party
- Mountaineering Glossary
- Medical Dictionary
- Japanese WordNet
- EDR

**Horizontal Language Grid**
- Web Service Wrapper
- Atomic Component
- Composite Component

**Community Language Services**
(developed in activity fields)

**Standard Language Services**
(developed by professionals)

Web Service Technology
Stakeholders of the Language Grid

- Language Resource Provider
- Computing Resource Provider
- Language Service User
- Language Grid Operator

Kyoto university will be a Language Grid Operator from November 2007.

The *Letter of Agreement* for joining the Language Grid is available upon a request.
Current Status of the Language Grid


Service Layers of the Language Grid

1. Intercultural Collaboration Tools
   Multilingual communication is supported using various language services.

2. Language Services
   Multiple language resources are composed using Web service workflows.

3. Language Resources
   Language resources are made usable as Web services with standardized interfaces.

4. P2P Grid Infrastructure
   Allow users to connect to Language Grid servers on the Internet.
1. Intercultural Collaboration Tools

**Langgrid Input**

- Langgrid Input is a tool to support multilingual text input.
- It can input multilingual text into existing collaboration tools such as BBS.

![Diagram of Langgrid Input features]

**Community Dictionary**

- Send an English message by one click.
Langgrid Chat

- Langgrid Chat is a tool for multilingual chatting.
- Users can add pictograms to their messages to express emotions, which are often lost through machine translation.
Langgrid Blackboard

- Langgrid Blackboard is an electronic blackboard for multilingual information sharing.
- This tool helps users to summarize meetings. Users can input texts and read them in their first languages.
2. Language Services

- Any collaboration tool can use language services.
- To create a new language service, describe an abstract workflow in BPEL4WS using the existing BPEL editors.
- Then, assign a concrete Web service to each task in the abstract workflow.

<An Abstract Workflow for Back Translation>

Translation
ja->zh

Translation
zh->ja

Change Service

J Server  Amikai  Web Transer

Actual Translation Services
Workflow can be Complex!

Multilingual Backtranslation (ja->zh->ja, ja->de->ja, ja->en->ja)

3 translation results, 3 back translation results

Japanese Morphological Analysis

Technical Term Extraction

remaining terms? Yes

Intermediate Code Table

remaining terms? Yes

Technical Term Multilingual Dictionary

Japanese-German Domain Specific Translation

Service Entity

Chasen (by Kyoto Univ.)

Pangaea’s Community Dictionary (by NPO Pangaea)

J Server (by Kodensha)

Web Transer (by Cross Language)
3. Language Resources

Language resources are registered as Web services equipped with a standard interface based on the “language service ontology.”
Now Available!

- **Machine Translation**

- **Morphological Analysis**
  - Japanese, Chinese, Korean, English, German, Spanish, French, Italian, Bulgarian

- **Bilingual Dictionary**
  - **Medical field**: Japanese, Chinese, Korean, English, Portuguese
  - **Disaster field**: Japanese, Chinese, English, French, Korean, Spanish, Thai
  - **IT field**: Japanese, English, Chinese
4. P2P Grid Architecture

- **Core Node**: Search language resources. Control accesses to resources.
- **Service Node**: Provide language resources.

![Diagram of P2P Grid Architecture]

- **Japanese-Uygur Translation**
- **In Medial Domain**
- **Japanese-Chinese Translation**
- **Uygur-Chinese Translation**

**Language Service**

**User**

**Information sharing**

**Service invocation**
Research Issue 1

Language Service Ontology


What is Language Service Ontology?

- A set of formalized concepts necessary for describing elements of a variety of language services.
  - Static language resources (dictionaries, corpora, …)
  - Algorithmic processing resources (NLP systems/tools)
  - Abstract linguistic objects (expression, meaning, description, annotation, …).

- It will enable wrapper program generation (Semantic Wrapping).
  - *Standard APIs* can be defined based on the ontology.
  - *Skeleton of the wrapper program* for a type of language service can be configured in advance based on the ontology-based description.
Semantic Wrapping

Language Service

Ontology

Wrapper Generation

Tool

Language Service
to be wrapped

specifying a service type

giving constraints for guiding wrapper induction

a skeleton is selected

skeleton library

Giving a few hints for wrapper induction with GUI interface

deployed in the Language Grid

Wrapper Program
Sketch of the *Upper* Ontology

Future Plan:

2006~2008: Establish a strawman proposal through discussions by an international research community including DFKI (Germany), CNR-ILC (Italy), Osaka U (Japan).

2009~: Make a proposal to international standardization body such as ISO TC37/SC 4.
Research Issue 2

Constraint Based Web Service Composition

Why Horizontal Composition?

- To realize services where users can combine machine translation, dictionary, morphological analysis, etc.

  - Two composite processes
    - **Vertical** composition → “best” combination of abstract Web services while satisfying all existing interdependent restrictions.
    - **Horizontal** composition → “best” selection of concrete Web service, from among a set of available functionally equivalent ones.

- The number of language services included in a composite service is at most six or seven. On the other hand, there are more than 100 parallel dictionaries available in the Internet.
  - Horizontal composition is more significant than vertical composition.
Why Constraint Optimization Problem?

- Users’ preferences and constraints can be naturally represented.
- Existing algorithms on constraint optimization problems can be utilized.

- Create a constraint network from a Web service workflow.
- Control construct “Sequence”
  - Execute more than one atomic services in order.
  - Example: Invoke a Japanese-English translation service and then invoke an English-German translation service
  - Users have a variety of constraints and preferences.
- Control construct “Loop”
  - Specify iterative structure. The number of iterations is unknown at the beginning.
  - Example: REPEAT to replace technical terms to intermediate codes UNTIL the input set becomes
  - Apply a framework of dynamic constraint optimization problems.
Example Language Service Composition

- $X_1$: Morphological analysis
  $D_1=\{LX-Suite, Postage-K, FreeLing, TreeTagger, HAM\}$

- Constraint:
  - Soft constraint: total cost
    $C_I: Cost(X_2)+Cost(X_4) \leq 100$
    $\rho_{C_I}$ denotes the penalty associated to soft constraints violation. For example,
    $\rho_{C_I} = \min((Cost(X_2)+Cost(X_4) - 100)/100, 1)$
  - Hard constraint: control construct
    $C_2: X_2.morph = X_1.morph$

- Objective function:
  The difference between user’s preference and the penalty of violation against soft constraints.
Analysis of Machine Translation Mediated Communication


Research Question

- How do participants establish common ground using machine translation?

- **Lexical Entrainment:**
  When people refer repeatedly to the same object, they converge to use the same terms. (Susan E. Brennan)

  “the curved round fish with the green stripe down its back”
  “the curved round fish with the green stripe”
  “the curved round fish”

  During the process of convergence, speakers and addresses come to take the same perspective on a referent.

- If people use machine translation, can we observe the process of convergence?
Problems in Communication over MT

- Inconsistency
  Translations of same words in different sentences can be inconsistent.
  Machine translators translate each sentence separately.

- Asymmetry
  Translations may not be transitive.
  A machine translator from language A to B is developed independently of a machine translator from B to A.
Controlled Experiment

Try to match identical sets of figures

< Japan side >

Multilingual chat

< China side >
Translation Inconsistency

<table>
<thead>
<tr>
<th>Japanese Screen</th>
<th>Chinese Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>(translated in English)</td>
<td>(translated in English)</td>
</tr>
<tr>
<td><strong>First Trial</strong></td>
<td><strong>First Trial</strong></td>
</tr>
<tr>
<td>J: My second figure looks like an animal.</td>
<td>J: My second figure is like an animal.</td>
</tr>
<tr>
<td>J: It has four feet and a tail.</td>
<td>J: It has four feet and a tail.</td>
</tr>
<tr>
<td>C: That's my the 9th.</td>
<td>C: That's my 9th.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&lt;Second Trial&gt;</th>
<th>&lt;Second Trial&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>J: My second figure is an animal with a tail and four feet.</td>
<td>J: My role of a young handsome beau is a boy with a tail and 4 feet.</td>
</tr>
<tr>
<td>C: What kind of meaning and rice boy?</td>
<td>C: What do you mean? A handsome boy?</td>
</tr>
</tbody>
</table>

Shortening of referring expressions does not work!

Machine translation generates something quite different based on very small changes.
Translation Asymmetry

<table>
<thead>
<tr>
<th>Japanese Screen</th>
<th>Chinese Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>(translated in English)</td>
<td>(translated in English)</td>
</tr>
<tr>
<td>J: 1 is a dancing lady.</td>
<td>J: 1 is a dancing lady.</td>
</tr>
<tr>
<td>C: <em>It jumped.</em></td>
<td>C: Ok, a dancing one.</td>
</tr>
<tr>
<td>J: 3 is a person with his head</td>
<td>J: <em>A person looking down is 3.</em></td>
</tr>
<tr>
<td>down.</td>
<td></td>
</tr>
</tbody>
</table>

Echoing for ratification does not work!

J: “I couldn’t understand what my partner meant, so I decided to proceed with another figure, which looked easier to match.”
Summary of This Talk

- This talk is on a new language service infrastructure on the Internet to combine existing language resources (machine translation, morphological analyzers, dictionaries etc.) to create customized language services, and to provide those language services for non-profit activities.

- Kyoto University started operating the Language Grid from December 2007. The Letter of Agreement for the Language Grid is available upon a request.

- More than 30 organizations including CNR, DFKI, CAS, Kyoto U, and Osaka U will join the project. It is our honor, if you will join the Language Grid.