

Language Grid Association: Action Research on Supporting the Multicultural Society

Satoshi Sakai¹, Masaki Gotou¹, Masahiro Tanaka¹, Rieko Inaba², Yohei Murakami²,
Takashi Yoshino^{2,3}, Yoshihiko Hayashi^{2,4}, Yasuhiko Kitamura⁵,
Yumiko Mori⁶, Toshiyuki Takasaki^{1,6}, Yoshie Naya⁷, Aguri Shigeno⁸,
Shigeo Matsubara^{1,2} and Toru Ishida^{1,2}

¹Department of Social Informatics, Kyoto University,

²Language Grid Project, National Institute of Information and Communications Technology

³Department of Design and Information Sciences, Wakayama University

⁴Graduate School of Language and Culture, Osaka University, ⁵Department of Informatics, Kwansei Gakuin University

⁶NPO Pangaea, ⁷Natural Disaster Youth Summit Office, ⁸Kyoto Center for Multicultural Society

{s-sakai, m-goto, mtanaka}@ai.soc.i.kyoto-u.ac.jp, {rieko.inaba, yohei}@nict.go.jp,

yoshino@sys.wakayama-u.ac.jp, hayashi@lang.osaka-u.ac.jp, ykitamura@kwansei.ac.jp,

{yumi, toshi}@pangaeon.org, yoshie.naya@gmail.com, shigenoaguri@tabunka.jp,

{matsubara, ishida}@i.kyoto-u.ac.jp

Abstract

The Language Grid is a middleware with which people can connect and use language resources such as machine translations, morphological analyzers and others created in the fields of intercultural collaboration. The Language Grid cannot exist without the collaboration of Language Grid Users who provide language and computation resources, language services, and collaboration tools. This paper overviews Language Grid Association, a user group of the Language Grid and a body promoting action research to support the multicultural society.

1. Introduction

Although the Internet allows people to be linked together, language remains the biggest barrier for intercultural collaboration. Though online language services already exist, difficulties about accessibility and usability often arise while trying to use those language services in intercultural activities [5].

Several research groups including National Institute of Information and Communications Technology (NICT), universities, and NPOs started working a language infrastructure on the Internet called the Language Grid [4]. Language Grid is a middleware for registering and forming language services with standardized interfaces. Fundamental software for the Language Grid has been studied and developed at NICT from April 2006. The Language Grid has been operated by the Department of Social Informatics, Kyoto University from December 2007 [6]. Various types of application activities are ongoing at Language Grid

Association, a user group of the Language Grid. This paper reports organizations and action research of Language Grid Association with goal of guiding the development of the Language Grid.

2. Language Grid Association

Figure 1 shows the relation among three organizations related to the Language Grid. NICT is working on R&D and provides software to the Operation Center run by Kyoto University. Language Grid Association uses services provided by the Operation Center and provides feedback to NICT by evaluating services.

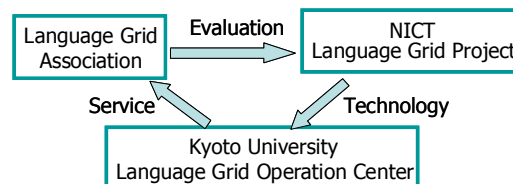


Figure 1. Language Grid organizations

The Language Grid Association is a loosely coupled organization formed by collaboration among industry, government, academia, and citizens. Its goal is to advance the technology and application of the Language Grid. The Association is located at COCON KARSUMA in the center of Kyoto. This place is mainly used for usability testing and for discussing business models for the Language Grid. Language Grid Association has been enlarged by creating focus groups and recruiting volunteer students and researchers.

The association aims to create and share technologies for using language resources registered with the Language Grid, and to overcome language barriers by connecting language resources worldwide. By the end of 2007, sixteen organizations including laboratories of universities, research institutes, NPOs and NGOs were participating in Language Grid Association. The number of participants is around fifty. The association consists of various SIGs (Special Interest Groups) such as research groups or projects, each of which aims to accumulate use cases and best practices. Each participant in the Language Grid Association joins one or more SIGs. SIGs are related each other through their activities classified as “Creating Language Services,” “Creating Collaboration Tools” or “Supporting Multicultural Activities.” The rest of paper provides a detailed description of each category.

3. Creating Language Services

A *language service* is a Web service for natural language processing. There could be a wide range of language services including specific ones, such as morphological analysis, and more general ones that could benefit users in accessing dictionaries and translating sentences.

3.1. Standardizing Language Service Ontology

Most existing language data resources and NLP tools/systems were created independently, resulting in a situation where data formats, annotation schemes, access methods, and other features are all idiosyncratic. This is obviously a barrier to establishing a language infrastructure like the Language Grid, which aims at increasing interoperability and reusability of language data resources and NLP tools/systems. To address this issue, standardization is inevitable: standardized APIs are necessary for NLP tools/systems, and standardized data semantics as well as data format are required for language data resources. In addition, and most importantly, these standards should be designed based on a comprehensive shared ontology, called *language service ontology*, which covers all possible elements of a language infrastructure.

We have already developed high-level sub-ontologies which are essential elements in the comprehensive language service ontology [1,2]. These sub-ontologies cover not only language data resources, processing resources (NLP tools/systems), but also a range of abstract linguistic objects, such as linguistic expressions and linguistic meaning. Figure 2 depicts the taxonomy of lexicon language resources, where the

entire taxonomy is defined by referring to the lexicon meta-model standard LMF (Lexical Markup Framework) developed by international standardization body ISO TC37/SC4.

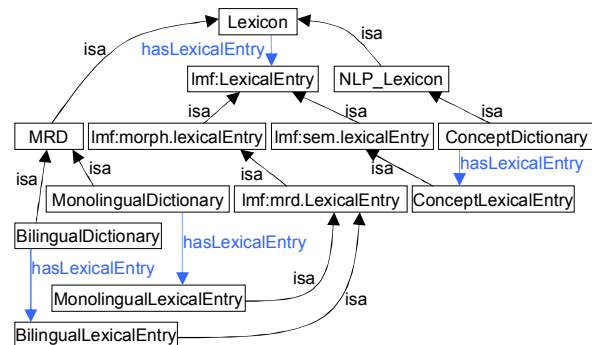


Figure 2. Taxonomy of language resources

The goal of developing a comprehensive language service ontology, on which all concerned can agree, calls for thorough discussions among experts from the relevant fields and the establishment of international standards. SIG on Language Service Ontology will serve as a vital body for facilitating such discussions among researchers specializing in related fields, such as NLP, Semantic Web, and AI planning.

3.2. Wrapping Language Resources

Language Resource Wrapping Project aims to develop wrappers that transform language data resources and language processing programs into Web services. Those services use the standard interfaces defined by the SIG on Language Service Ontology. This project contributes to users who want to provide their language resources to the Language Grid, and so increase the language resources available via the Language Grid. For example, Figure 3 illustrates how domain specific dictionaries created by multilingual communities are wrapped as bilingual dictionary services with a standard *search* operation.

Around 70 resources have been experimentally wrapped up to this point. Most of them are for the purpose of domain specific machine translation services for Japanese, Chinese, Korean, English, and other European languages. Resources like dictionaries and parallel texts in the medical domain have also been wrapped and used in services as described in Section 5.3. In addition, to accumulate wrapping techniques, our project members are developing the libraries and templates needed for wrapping. A wrapping manual is now available online. In the future, we will develop a

wrapping support environment to assist people using these libraries and templates. Furthermore, we are considering a language service wrapper repository that allows volunteers to publish their wrappers.

Interface name:	<i>BilingualDictionaryService</i>
Method name:	<i>search</i>
Parameters:	
<i>headLang</i>	the source language (ex. "en")
<i>targetLang</i>	the target language (ex. "ja")
<i>headword</i>	the word to search (ex. "culture")
<i>matchingMethod</i>	the method (ex. "complete")
Return value:	<i>Translation[]</i>
<i>headWord</i>	result in source language (ex. "culture")
<i>targetWord[]</i>	result in target language (ex. "文化")

Figure 3. Standard parallel text service

3.3. Composing Web Services

Web Service Workflow Lab provides workflows, which can be used to combine language services available on the Language Grid. For example, a workflow for translating technical documents is composed of translation engines, morphological analyzers, and domain specific dictionaries. An workflow example, translating technical terms, is shown in Figure 4. This workflow extracts technical terms from an input sentence, and then translates them using domain specific dictionaries.

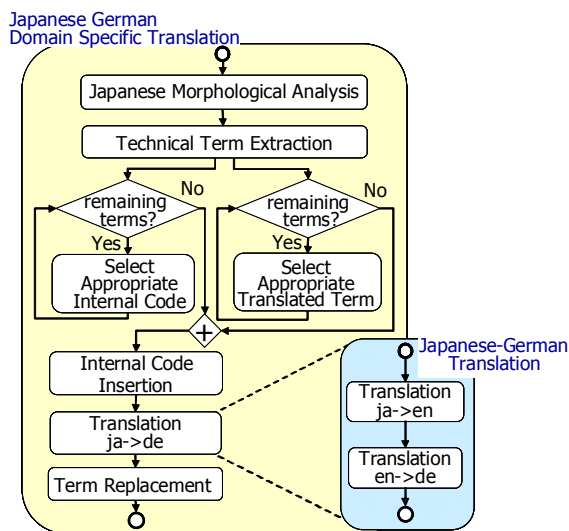


Figure 4. Domain specific translation workflow

In the Language Grid, workflows are described in WS-BPEL, a standard format of Web service composition. Workflows are interpreted and executed by the WS-

BPEL engine embedded in the Language Grid. We have produced a wide variety of workflows for applications developed by the Language Grid Association. To establish a better foundation for language service workflows, we are also working on research issues: applying semantic Web technologies to workflow composition, and discovering/reusing patterns of Web service workflow.

4. Creating Collaboration Tools

Language Grid Association has developed various collaboration tools by using Web services described in Section 3.3. Based on development activities, we formed the SIG on Multilingual Collaboration Tools to share techniques of using language resources and workflows.

Example tools include a multilingual chat tool called *AnnoChat*. This tool translates an input message into several languages and displays the translated sentences. Users can chat in their first languages, and the tool translates by using language services. When a user inputs a message in AnnoChat, a workflow deployed on the Language Grid is invoked. The workflow then invokes a web service to translate messages using domain specific dictionaries. AnnoChat server receives the translation results from the workflow and displays the messages in the different languages.

The second system is the collaboration tool called *Multilingual NOTA*. This tool adds translation services to the existing collaboration tool *NOTA*. *NOTA* allows participants to easily upload documents and draw pictures on a shared Web page. In *Multilingual NOTA*, documents can be translated by using the Language Grid. Users can confirm translation quality by using back translation. By using multilingual functions, multicultural communities can establish collaboration efficiently.

It is important to note that these tools from Language Grid Association were developed under user-centered design principles and are easy to use in practice. The Usability Lab is performing heuristic evaluations and user tests on these tools to improve their usability, and is providing feedback to developers to make them easier-to-use [3]. In particular, the Usability Lab evaluates how accurately the language services translate from one language to another. When multiple language services are connected to create a composite language service, the quality of translation often degrades as the number of connected services increases

[10]. The Lab determines whether a given language service, including composite services, is fit to use.

5. Supporting Multicultural Activities

Language Grid Association is supporting multicultural activities. There are several communities that are promoting international activities. The key to the success of these activities is that people from all over the world can actively participate in discussions. However, communication is not always smooth, because the official language of communities is not often the first language of the participants. Language Grid Association organizes support teams for communities working on intercultural collaboration.

5.1. Connecting Kids around the World

Pangaea Support Team combines language resources or linguistic services of Language Grid and existing tools, including bulletin board systems and chat systems, to support the multilingual communication that takes place under NPO Pangaea [8]. NPO Pangaea's mission is to be the peace engineering laboratory for the creation of the universal playground, where children across the world can develop personal bonds transcending the boundaries of language, time, and space. Pangaea has its children activity locations such as youth centers, schools, and community centers all over the world. Thus, language barriers exist among them, and communication between the staff of different locations is difficult, when organizing and holding multicultural events. In such circumstances, the *Langrid Chat* and *Langrid Input*, multilingual tools provided by NICT are used to support communication among the international staff as follows:

Langrid Chat, a kind of AnnoChat, translates the input messages and displays the translated multilingual messages. There are two use cases. First, it is used to realize remote communication among staff. Staff at each location need to communicate with each other a lot to design online activities involving multiple locations. However, many of the staff including volunteers know only their native language. Thus, communication is often slow and awkward. Using the multilingual chat system enables staff to communicate in their first language and reduces the stress involved in such discussions. Second, it is used to realize local communication in face-to-face meetings among international staff. In such meetings, members speak different languages, but NPO cannot always allocate human interpreters. Langrid Chat bridges the language gap by displaying multilingual chat via a projector to

enable fluent communication. The speaker can use his/her first language, and people who are good at typing enter the message into Langrid Chat. Other staff then read these messages displayed in their first language and understand what the speaker has said. Moreover, local staff can input their questions or opinions in their first language.



Figure 5. Facilitator discussion in Korea

Langrid Input is a multilingual input tool. Staff members write multilingual emails and documents using this tool. Domain specific dictionaries can be used. Adding words to Pangaea's dictionary produces a translation system that is specific to Pangaea, and community terms can be translated correctly. Collaboration tools using the Language Grid provide cost effective communication opportunities to NPO Pangaea.

5.2. Multilingual Discussion on Disaster Management

JEARN is the branch of NGO iEARN, the world's biggest international education network. Schools from various countries are participating in the project *Natural Disaster Youth Summit*; they worked on online collaborative learning with the theme of "Let's Create a Global Disaster Safety Map."

JEARN uses video conferencing and English BBS for communication. However, students and teachers who do not use English in daily life had difficulty in writing English messages on chat screens at video conference or the BBS. JEARNSupport Team used Langrid Input as a communication tool to help those students and teachers. JEARNSupport Team used Langrid Input to translate their messages into English [9]. Figure 6 shows an English message created by a Japanese student with the aid of Language Input.

The flow of information using Langrid Input is as follows: First, students think about their messages in Japanese, and the messages are entered via a word processor or directly to the input screen of Langrid Input. Next, the messages written in Japanese are translated into English. Finally, the English messages are used to transmit information or for communications by entering them on the BBS called iEARN Discussion Forum. The students can also use the translated messages as presentation material.

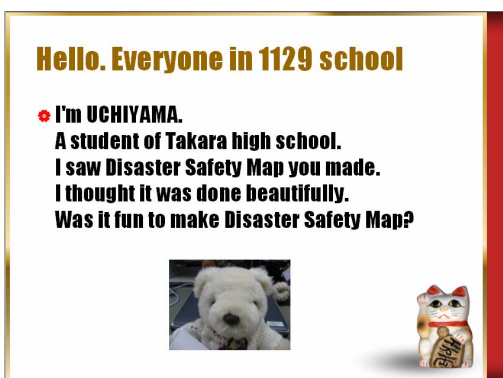


Figure 6. Example message created with the Language Grid

Through the communication in English using Langrid Input, the students' attitudes to communication in English has changed. To clarify this change, questionnaires were conducted before (September 2006) and after the project (February 2007). The first questionnaire showed the students' lack of confidence in communicating in English. However, the post-project questionnaire showed a strong improvement in the students' confidence to communicate in English. The result indicates that the students' lack of confidence in English was decreased through actual communication in English using Langrid Input.

Affirmative opinions of students in the free description part of the post-project questionnaire such as "I want to continue the communication with foreign friends" and "I enjoyed the communication, although I am not good at English" also provide positive evidence of the students' change. Although most of the students were satisfied with Langrid Input, some students wanted more accuracy and user friendliness. This feedback will contribute to enhancing the collaboration tools.

5.3. Foreign Patient Support at Hospitals

When foreigners, who are not fluent in Japanese, fall ill in Japan, they may be unable to receive adequate medical attention because of their inability to

communicate with Japanese medical doctors. The purpose of SIG on Medical Support Systems for Foreigners is to develop a support system for foreigners in such situations. The computer mediated multilingual medical communication support system called M³ was developed by Wakayama University in cooperation with Kyoto Center for Multicultural Society. This NPO started a medical interpretation program in 2003 to assist foreign patients. Volunteer interpreters were dispatched a total of 1700 times per year. Interpreters are also stationed in several affiliated hospitals. In the case of medical interpretation, machine translations are not useful due to their rough quality. Therefore, the interpreters refer to multilingual parallel texts of medical sentences.

M³ is a support system for communication between foreign outpatients and medical staff at hospital reception desks [7]. As a result, outpatients who cannot speak Japanese can receive information from M³ and communicate in their own language with hospital staff. The function "Hospital Navigation" provides guidance around each section in a hospital. The function "Q & A" provides FAQs for a hospital. The function "Consultation Help" describes the procedure used to seek consultation. After the outpatient answers some questions posed by the system, it replies with the appropriate consultation procedure. Outpatients who cannot speak Japanese can communicate with hospital staff by using M³. For example, hospital staff can ask outpatients about their symptoms. Figure 7 illustrates such communication using M³.



Figure 7. Screenshot of M³

6. Conclusion

Language Grid Association, a user group of the Language Grid, is conducting various support activities for multicultural societies. The first type of activity is to develop language services. The association forms

SIGs for defining the standard language service ontology, wrapping language resources according to the ontology, and creating composite language services using Web service workflows.

The second type of activity is to create collaboration tools. SIGs on tool development uses language services registered in Language Grid. The achievements include the multilingual chat tool Annochat, and the multilingual collaboration tool Multilingual NOTA.

Finally, the above language services and tools are being tested in actual instances of intercultural collaboration. NPO Kyoto Center for Multicultural Society is using parallel texts in the hospital domain to support foreign patients. The association also forms teams to support NPO Pangaea's worldwide activities, NGO iEARN/JEARN's multilingual discussion, and so on.

Acknowledgement

This report would not have been possible without the collaboration of the members of NICT Language Grid Project, Kyoto University Language Grid Operation Center, and Language Grid Association.

References

- [1] Yoshihiko Hayashi and Toru Ishida. A Dictionary Model for Unifying Machine Readable Dictionaries and Computational Concept Lexicons. *International Conference on Language Resources and Evaluation (LREC)*, pp.1-6, 2006.
- [2] Yoshihiko Hayashi, Thierry Declerck, Paul Buitelaar, and Monica Monachini. Ontologies for a Global Language Infrastructure. *International Conference on Global Interoperability for Language Resources (ICGL)*, pp. 105-112, 2008.
- [3] Rieko Inaba, Yohei Murakami, Akiyo Nadamoto, and Toru Ishida. Multilingual Communication Support Using the Language Grid. *LNCS 4568*, pp. 118-132, 2007.
- [4] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet (SAINT)*, pp. 96-100, 2006.
- [5] Toru Ishida, Susan R. Fussell and Piek TJM Vossen Eds. *Intercultural Collaboration*, Lecture Notes in Computer Science, 4568, Springer-Verlag, 2007.
- [6] Toru Ishida, Akiyo Nadamoto, Yohei Murakami, Rieko Inaba, Tomohiro Shigenobu, Shigeo Matsubara, Hiromitsu Hattori, Yoko Kubota, Takao Nakaguchi, and Eri Tsunokawa. A Non-Profit Operation Model for the Language Grid. *International Conference on Global Interoperability for Language Resources (ICGL)*, pp. 114-121, 2008.
- [7] Mai Miyabe, Kunikazu Fujii, Tomohiro Shigenobu, and Takashi Yoshino. Parallel-Text Based Support System for Intercultural Communication at Medical Receptions. *LNCS 4568*, pp. 182-192, 2007.
- [8] Yumiko Mori. Atoms of Bonding: Communication Components Bridging Children Worldwide. *LNCS 4568*, pp. 335-343, 2007.
- [9] Yoshie Naya. How Intercultural Disaster Reduction Education Change Students: A Case Study of an Evening Course Senior High School in Hyogo, Japan. *LNCS 4568*, pp. 368-381, 2007.
- [10] Naomi Yamashita and Toru Ishida. Effects of Machine Translation on Collaborative Work. *International Conference on Computer Supported Cooperative Work (CSCW)*, pp. 515-523, 2006.