



IJCNLP 2011

Proceedings of
the Workshop on
Language Resources, Technology and
Services in the Sharing Paradigm

November 12, 2011
Shangri-La Hotel
Chiang Mai, Thailand



IJCNLP 2011

**Proceedings of
the Workshop on Language Resources, Technology and
Services in the Sharing Paradigm**

November 12, 2011
Chiang Mai, Thailand

We wish to thank our sponsors

Gold Sponsors



www.google.com



www.baidu.com



[The Office of Naval Research \(ONR\)](#)



[The Asian Office of Aerospace Research and Development \(AOARD\)](#)



[Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong](#)

Silver Sponsors



[Microsoft Corporation](#)

Bronze Sponsors



[Chinese and Oriental Languages Information Processing Society \(COLIPS\)](#)

Supporter



[Thailand Convention and Exhibition Bureau \(TCEB\)](#)

We wish to thank our sponsors

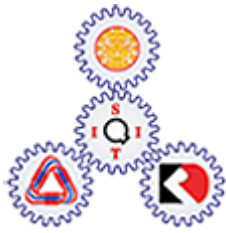
Organizers



[Asian Federation of Natural Language Processing \(AFNLP\)](#)



[National Electronics and Computer Technology Center \(NECTEC\), Thailand](#)



[Sirindhorn International Institute of Technology \(SIIT\), Thailand](#)



[Rajamangala University of Technology Lanna \(RMUTL\), Thailand](#)



[Maejo University, Thailand](#)



[Chiang Mai University \(CMU\), Thailand](#)

©2011 Asian Federation of Natural Language Processing

Introduction

The Context

Some of the current major initiatives in the area of language resources – FLaReNet (<http://www.flarenet.eu/>), Language Grid (<http://langrid.nict.go.jp/en/index.html>) and META-SHARE (www.meta-share.org, www.meta-net.eu) – have agreed to organise a joint workshop on infrastructural issues that are critical in the age of data sharing and open data, to discuss the state of the art, international cooperation, future strategies and priorities, as well as the road-map of the area.

It is an achievement, and an opportunity for our field, that recently a number of strategic-infrastructural initiatives have started all over the world. It is also a sign that funding agencies recognise the strategic value of our field and the importance of helping a coherent growth also through a number of coordinated actions. Some of these initiatives, two European and one Asian, have agreed to join forces to foster a debate that may lead to future coordinated actions all over the world.

- FLaReNet aims at providing recommendations for future initiatives in the field of Language Resources and Technologies: we are aware that it is important to discuss future policy and priorities not only on the European scene, but also in a worldwide context. This is true both when we try to highlight future directions of research, and – even more – when we analyse which infrastructural initiatives are needed. The growth of the field should be complemented by common efforts that try to look for synergies and to overcome fragmentation. FLaReNet is now ready to deliver its final recommendations and priorities deriving from the various events it organised, summarised in the FLaReNet Blueprint of actions and infrastructures.
- The Language Grid aims at constructing a multilingual service infrastructure on the Internet that allows users to share language services such as online dictionaries, bilingual corpora, and machine translations, and create new language services by combining existing services to support intercultural collaboration. The Language Grid service-oriented infrastructure shifts from language resources to language services. The Language Grid is operated by Kyoto and Bangkok operation centers in a federated fashion. More than 110 language services are registered and shared by 140 groups in 18 countries.
- META-SHARE aims to build an open, integrated, secure and interoperable exchange facility for language resources (data and tools) for the Language Technologies domain and other applicative domains (e.g., digital libraries, cognitive systems, robotics, etc) where language plays a critical role. It aims to act as an infrastructure that will enable language resources documentation, cataloguing, uploading and storage, downloading and exchange, aiming to support a resources economy at large. META-SHARE has defined the principles underlying its design and governance model, its proposed architecture for language resources sharing and collaborative building as well as the technical and legal instruments supporting its operation.

Topics and Aims

Cooperation is an issue that needs to be prepared. This joint strategic Workshop intends to continue a discussion, started on several occasions in the last years, on the usefulness and the interest of promoting international cooperation among various initiatives and communities around the world, within and around

the field of Language Resources and Technologies.

The Workshop aims at addressing (some of the) technological, market and policy challenges posed by the “sharing and openness paradigm”, the major role that language resources can play and the consequences of this paradigm on language resources themselves.

Examples of topics and issues addressed by the 14 papers accepted are:

- Sharing Language Resources and Language Technologies, as implemented in a number of international initiatives
- Need for global information on Language Resources and Language Technologies: relevant initiatives in the various regions/countries
- Interoperability and Reusability, both in infrastructures and in applied systems (MT)
- Linguistic web services and language applications development
- Metadata and Cataloguing
- Collaborative initiatives for annotated resource creation
- Infrastructures, policies, gaps and critical areas

We hope that the papers will contribute to boosting a constructive and fruitful debate.

The Workshop is also an occasion for the researchers interested in infrastructural initiatives to get together to discuss and promote collaboration actions. It should also lead to discussing the modalities of how to organise the cooperation among various initiatives. In a conclusive Panel these topics will be discussed, with invited panellists from all the continents and with all the Workshop participants, with the aim of delineating:

- Common roadmap and strategies: local, regional, international frameworks
- International cooperation, models for collaboration and agreements on joint initiatives

We hope that by organising this workshop in Thailand we can attract participation of many Asian initiatives and can lead to fruitful collaboration between Asian and other international initiatives.

Finally, we wish to thank very much Irene Russo who has significantly helped in the organisation of the Workshop and in the preparation of the Proceedings.

Nicoletta Calzolari

Toru Ishida

Stelios Piperidis

Virach Sornlertlamvanich

Workshop Chairs:

Nicoletta Calzolari (ILC-CNR, Pisa, Italy)
Toru Ishida (Kyoto University, Japan)
Stelios Piperidis (ILSP, Athens, Greece)
Virach Sornlertlamvanich (NECTEC, Thailand)

Organizational and Editorial Assistant:

Irene Russo (ILC-CNR, Pisa, Italy)

Program Committee:

Sophia Ananiadou
Nuria Bel
Emily Bender
Nicoletta Calzolari
Tommaso Caselli
Khalid Choukri
Yoshihiko Hayashi
Chu-Ren Huang
Nancy Ide
Toru Ishida
Steven Krauwer
Joseph Mariani
Monica Monachini
Yohei Murakami
Stelios Piperidis
Valeria Quochi
Claudia Soria
Virach Sornlertlamvanich
Thepchai Supnithi
Take Tokunaga
Dan Tufis
Kiyotaka Uchimoto

Table of Contents

<i>Prospects for an Ontology-Grounded Language Service Infrastructure</i> Yoshihiko Hayashi	1
<i>A Method Towards the Fully Automatic Merging of Lexical Resources</i> Núria Bel, Muntsa Padró and Silvia Neculescu	8
<i>Service Quality Improvement in Web Service Based Machine Translation</i> Sapa Chanyachatchawan, Virach Sornlertlamvanich and Thatsanee Charoenporn	16
<i>The Semantically-enriched Translation Interoperability Protocol</i> Sven Christian Andrä and Jörg Schütz	24
<i>Interoperability and Technology for a Language Resources Factory</i> Marc Poch and Núria Bel	32
<i>Interoperability Framework: The FLaReNet Action Plan Proposal</i> Nicoletta Calzolari, Monica Monachini and Valeria Quochi	41
<i>Promoting Interoperability of Resources in META-SHARE</i> Paul Thompson, Yoshinobu Kano, John McNaught, Steve Pettifer, Teresa Attwood, John Keane and Sophia Ananiadou	50
<i>Federated Operation Model for the Language Grid</i> Toru Ishida, Yohei Murakami, Yoko Kubota and Rieko Inaba	59
<i>Open-Source Platform for Language Service Sharing</i> Yohei Murakami, Masahiro Tanaka, Donghui Lin and Toru Ishida	67
<i>Proposal for the International Standard Language Resource Number</i> Khalid Choukri, Jungyeul Park, Olivier Hamon and Victoria Arranz	75
<i>A Metadata Schema for the Description of Language Resources (LRs)</i> Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis, Monica Monachini, Francesca Frontini, Gil Francopoulo, Victoria Arranz and Valérie Mapelli	84
<i>The Language Library: Many Layers, More Knowledge</i> Nicoletta Calzolari, Riccardo Del Gratta, Francesca Frontini and Irene Russo	93
<i>Sharing Resources in CLARIN-NL</i> Jan Odijk and Arjan van Hessen	98
<i>META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries</i> Inguna Skadina, Andrejs Vasiljevs, Lars Borin, Koenraad De Smedt, Krister Lindén and Eiríkur Rögnvaldsson	107

Conference Program

Saturday, November 12, 2011

8:30–8:40 Introduction

Language Services, Resources, Applications

8:40–9:00 *Prospects for an Ontology-Grounded Language Service Infrastructure*
Yoshihiko Hayashi

9:00–9:20 *A Method Towards the Fully Automatic Merging of Lexical Resources*
Núria Bel, Muntsa Padró and Silvia Necsulescu

9:20–9:40 *Service Quality Improvement in Web Service Based Machine Translation*
Sapa Chanyachatchawan, Virach Sornlertlamvanich and Thatsanee Charoenporn

9:40–10:00 *The Semantically-enriched Translation Interoperability Protocol*
Sven Christian Andrä and Jörg Schütz

10:00–10:30 Coffee/Tea Break

Interoperability

10:30–10:50 *Interoperability and Technology for a Language Resources Factory*
Marc Poch and Núria Bel

10:50–11:10 *Interoperability Framework: The FLaReNet Action Plan Proposal*
Nicoletta Calzolari, Monica Monachini and Valeria Quochi

11:10–11:30 *Promoting Interoperability of Resources in META-SHARE*
Paul Thompson, Yoshinobu Kano, John McNaught, Steve Pettifer, Teresa Attwood,
John Keane and Sophia Ananiadou

11:30–11:50 *Federated Operation Model for the Language Grid*
Toru Ishida, Yohei Murakami, Yoko Kubota and Rieko Inaba

11:50–12:00 Short Discussion

Saturday, November 12, 2011 (continued)

12:00–14:00 Lunch

Services and Functionalities for the Community

14:00–14:20 *Open-Source Platform for Language Service Sharing*

Yohei Murakami, Masahiro Tanaka, Donghui Lin and Toru Ishida

14:20–14:50 *Proposal for the International Standard Language Resource Number*

Khalid Choukri, Jungyeul Park, Olivier Hamon and Victoria Arranz

14:50–15:10 *A Metadata Schema for the Description of Language Resources (LRs)*

Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis, Monica Monachini, Francesca Frontini, Gil Francopoulo, Victoria Arranz and Valérie Mapelli

15:10–15:20 *The Language Library: Many Layers, More Knowledge*

Nicoletta Calzolari, Riccardo Del Gratta, Francesca Frontini and Irene Russo

15:20–15:30 Short Discussion

15:30–16:00 Coffee/Tea Break

Sharing and Infrastructures

16:00–16:20 *Sharing Resources in CLARIN-NL*

Jan Odijk and Arjan van Hessen

16:20–16:40 *META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries*

Inguna Skadina, Andrejs Vasiljevs, Lars Borin, Koenraad De Smedt, Krister Lindén and Eiríkur Rögnvaldsson

16:40–17:30 PANEL Language Resources and Technologies Sharing: Priorities and Concrete Steps