



**SIXTH FRAMEWORK PROGRAMME
PRIORITY 2
Information Society Technology**

LUNA
**spoken Language Understanding in multilinguA communication
systems**

Project n. 33549

**D 6.2 – Dissemination, exploitation and
standardization plan**

Due date of deliverable: September 3, 2008

Actual submission date: October 15, 2008

Start Date of Project September 4, 2006

Duration 36 months

Organization name of lead contractor for this deliverable LOQ

Version 3.0

Dissemination level [PU]

Authors and reviewers tables

Authors	Company Name
Silvia Mosso	LOQ
Paolo Baggia	LOQ
Stefan Hahn	RWTH
Giuseppe Riccardi	UT
Frederic Bechet	UAPV
Geraldine Damnati	FT
Kepa Joseba Rodriguez	CSI
Krzysztof Marasek	PJIIT
Malgorzata Marciniak	PAS
Joanna Rabiega-Wiśniewska	PAS

Reviewers	Company Name
Paolo Coppo	LOQ
Patrizia Pautasso	LOQ
Renato De Mori	UAPV

Revision table

Version	Date	Modified sections
1.0	March 2007	
2.0	October 2007	2.4: Scientific dissemination 2.7: Link with other activities 3: Preliminary exploitation plan
3.0	October 2008	2.4: Scientific dissemination 2.5: Industrial dissemination 2.6: dissemination towards general public 2.7: Link with other activities 3: Preliminary exploitation plan 4:Standardization plan

Abstract

The objective of this document is to outline the LUNA dissemination, exploitation and standardization strategy and to point out the activities done and planned.

This is mainly addressed by work package 6, which aims to make the project known and to ensure communication of its results to the general public, the research community, the industry and the Public Administrations.

Activities will be of different kinds: on one hand, the academic partners will disseminate the project outcomes through workshops, conferences and journals; on the other hand, the industrial component of the project, with a significant interest in exploitation, will play a significant role presenting the results in press releases, industrial fairs and conferences as well as in standardization bodies.

This document is intended as a living document and is being updated at the end of each year to reflect the status of the activities.

TABLE OF CONTENTS

1	Project presentation.....	5
1.1	Project description	5
1.2	Partners	5
1.3	Public deliverables and outcomes	6
2	Dissemination plan	7
2.1	Project image	7
2.2	Website	7
2.3	Brochure	9
2.4	Scientific dissemination	9
2.4.1	Workshops and conferences.....	9
2.4.2	Scientific journals	12
2.4.3	Publications	12
2.5	Industrial dissemination	18
2.6	Dissemination towards general public	18
2.7	Link with other activities.....	18
3	Preliminary exploitation plan	20
4	Standardization plan.....	23
5	Annex: Contractual and Internal Requirements	24

1 Project presentation

1.1 Project description

LUNA is a three-year project focused on the problem of real-time understanding of spontaneous speech in the context of advanced telecom services.

LUNA is headed by Loquendo and involves eight partners from Italy, France, Germany and Poland.

The main objective of LUNA is the creation of a robust natural spoken language understanding toolkit for multilingual dialogue services, able to carry on human-computer communication with a good degree of user satisfaction.

From a technological point of view, the objectives of LUNA are to propose new methods, algorithms and tools for the fast development of robust SLU components for multilingual telephone services. To this aim, LUNA will address a set of challenging scientific problems, by focusing on five scientific objectives:

- Language Modelling for Speech Understanding;
- Semantic Modelling for Speech Understanding;
- Automatic Learning (including Active and On-Line Learning);
- Robustness issues for SLU;
- Multilingual portability of SLU components.

In particular, three steps are being considered for the SLU interpretation process: generation of semantic concept tags, semantic composition and context-sensitive validation.

LUNA's research results will be validated on different application scenarios, targeted to dialogue-based telephone services of different complexity (e.g. from call routing with utterance classification to dialogue systems with complex semantic domains). The SLU models will be trained and applied to different multilingual spoken dialog systems in French, Italian and Polish. The language-independent components will be shared among the participants and then adapted to each particular language by means of language resources already available or collected within the project.

The highly qualified academic presence in the consortium ensures scientific excellence and credibility in carrying out this leading-edge research activity, while the project results will immediately become a competitive advantage for industrial partners who will be able to exploit them directly, introducing them to the speech technologies market.

1.2 Partners

The project combines three major components: a highly qualified research team, two industrial partners and the technological arm of the Public Administration. All partners will be active in the dissemination activities. The industrial companies (LOQ and FT) and the PA organization will also be active in the exploitation activities.

The consortium is structured as follows:

Partic. no.	Participant name	Participant short name	Country
1	Loquendo S.p.A.	LOQ	IT

2	RWTH Aachen	RWTH	DE
3	University of Trento	UT	IT
4	University of Avignon	UAPV	FR
5	France Telecom R&D Division S.A.	FT	FR
6	CSI-Piemonte, Consorzio per il Sistema Informativo	CSI	IT
7	Polish-Japanese Institute of Information Technology	PJIT	PL
8	Institute of Computer Science Polish Academy of Sciences	PAS	PL

1.3 Public deliverables and outcomes

A number of deliverables, including the “Dissemination, exploitation and standardization plan” will be made available to a wide audience after the EC approval. Public deliverables are listed below:

Del. No	Deliverable name	WP No	Lead participant	Delivery date	Nature	Dissemination level
D1.1	Specification of tasks for corpus data acquisition	1	UT	2	R	PU
D1.3	Specifications of the annotation protocol for the data	1	UT	6	R	PU
D2.3	Report on Generation of Semantic concept tags functionality	2	RWTH	24	R	PU
D2.4	Report on integrated ASR and SLU search	2	RWTH	33	R	PU
D3.3	Report on semantic composition functionality	3	UAPV	33	R	PU
D4.3	Report on context-sensitive validation and portability	4	FT	33	R	PU
D5.2	Evaluation criteria and methodology	5	CSI	28	R	PU
D5.5	Evaluation of components and their combination	5	CSI	23, 36	R	PU
D6.1	Website	6	LOQ	3	P	PU
D6.2	Dissemination, exploitation and standardization plan	6	LOQ	6, 12, 24, 36	R	PU
D6.3	LUNA functionalities description	6	LOQ	21, 36	R	PU

In addition to the above, data corpus acquisition will be performed in Italian, French and Polish. A significant subset of the corpora will be made accessible for research purposes at very favourable conditions, and at fair market value for the commercial usage. Data will be distributed through specialized organisms (e.g. ELRA, LDC, or academic partners' websites), avoiding those which don't fulfil the above condition. Public availability will comply with Privacy protection laws of each Country.

2 Dissemination plan

The objective of the dissemination plan is to define a strategy for the communication of the results to the target audience and ensure that dissemination activities made by each partner are fully integrated in the overall strategy.

The LUNA dissemination will address the scientific community, the industry and the Public Administrations. Dissemination and exploitation activities are also planned towards Loquendo's shareholder, Telecom Italia, and France Telecom group.

Dissemination activities have started since the very beginning of the project and will also continue after the end of LUNA.

The research and technology partners will present the results of the project through different and complementary channels:

- Scientific dissemination will use the traditional channels of scientific journals, international conferences and workshops, and possibly – if a suitable arrangement with an Editor will have been found - a more organic publication such as a book after the end of the project
- Industrial dissemination will be made in most relevant industrial conferences and exhibition and through partners' publications such as newsletters and marketing announcements
- Dissemination towards the general public will be managed by partners which are in closer contact with PAs and end users and will also include, if feasible, information through media such as the TV

Other ways to disseminate the project's outcomes include the continuous update of the web site and possibly, before the end of the project, the organization of a special event in conjunction with a relevant speech conference.

All publications are also subject to terms related to publicity of the project, to confidentiality, and to protection of knowledge and will undergo the policy and rules the partners agreed on in the Consortium Agreement (see the extracts from the CA in the annex), as well as the EC rules as described in Articles II.12 and II.33 of the contract.

2.1 Project image

A preliminary activity has been the definition of the project image with the aim of making recognizable and consistent the documents, products and services.

A logo has been defined as well as the documents (deliverables, minutes, etc.) templates. The common presentation template reflecting the project identity will be used in presentations conducted in conferences and workshops. All the materials can be downloaded from the member restricted area of the LUNA website.

2.2 Website

The LUNA website (www.ist-luna.eu) is the project main dissemination tool and will be continuously updated.

The website was officially published on November 30, 2006 and is continuously evolving with the project.

It is composed by a series of static pages aimed at presenting the project, and some other interactive pages which collect project news, events and main results both from the project and the scientific community.

The website is structured in sections as follows:

- Home section (see Figure 1) providing a project overview.
- Project section with the description of the objectives, of the scientific and technical approach and of the workplan.
- Consortium section including links to partners' websites and email addresses for both general and technical contacts.
- Publications section providing project's documents such as leaflets, papers, public deliverables, ...
- News and events section providing project updates and event information.
- Log in section to provide access to the restricted area for project partners.

LUNA - SPOKEN LANGUAGE UNDERSTANDING IN MULTILINGUAL COMMUNICATION SYSTEMS

WELCOME TO LUNA

SPOKEN LANGUAGE UNDERSTANDING IN MULTILINGUAL COMMUNICATION SYSTEMS

PROJECT DESCRIPTION
 The LUNA project addresses the problem of real-time understanding of spontaneous speech in the context of advanced telecom services, and aims at the creation of a robust natural Spoken Language Understanding toolkit for multilingual dialogue services, able to carry on human-computer communication with a good degree of user satisfaction.
[continue..](#)

For Info

JOB OPPORTUNITIES IN LUNA

PostDoc in Spoken Language and Dialog Research
 University of Trento is looking for a candidate who will investigate machine learning techniques for adaptive spoken dialog prototyping.
[More info](#)

[Home](#) | [Project](#) | [Consortium](#) | [Publications](#) | [News and Events](#) | [Log in](#)

© 2007 LUNA. All Rights Reserved.

Figure 1- LUNA website home page

2.3 Brochure

The consortium has realized a brochure for the LUNA project general presentation. The design of the flyer is in line with the project's identity.

Since the goal of the brochure is to promote the project to general public, the language is simple and the communication is focused on the benefits for end users.

2.4 Scientific dissemination

Dissemination of scientific and technological results is using the traditional channels (conferences, workshops and scientific journals). Moreover, the participation of the LUNA partners to other national and international projects, as well as the links with international organizations, also increase the dissemination potential.

All published materials will be uploaded to the website and made available to the target audience.

It also worth mentioning that Frédéric Béchet and Giuseppe Riccardi are members of the IEEE Speech Technical Committee and Renato De Mori has been nominated IEEE distinguished lecturer for Spoken Language Understanding and Human-machine Dialogue.

2.4.1 Workshops and conferences

Papers describing significant results have been accepted and others will be tentatively submitted at appropriate international conferences and workshops. Below is the list of potential events which LUNA researchers attended or plan to attend. Within the project timeframe, the project partners plan to participate in different potential events which have been listed below. The list aims to give only a first idea of what the partners are targeting at and will be updated during the course of the project.

Planned/ actual dates and place	Event	Type of audience	Countries addressed	Partner(s) involved
September 17-21, 2006, Pittsburgh, USA	Interspeech 2006	research	international	UAPV, FT, UT
April 15-20, 2007, Honolulu, Hawaii	ICASSP 2007	research	international	UAPV, FT
April 23, 2007, Warsaw, Poland	Natural Language Processing Seminar	research	Poland	PAS
April 22-27, 2007, Rochester, USA	Human Language Technology – HLT 2007	research	international	UAPV, FT
May 28-June 1, 2007, Warsaw, Poland	Corpus linguistics session at The East-Central European School in the Humanities	research, students	East-Central Europe	PAS
May 30-June 1, 2007, Rovereto, Italy	DECALOG	research	international	UT, all

Planned/ actual dates and place	Event	Type of audience	Countries addressed	Partner(s) involved
June 7, 2007, Potsdam, Germany	project D1 "Linguistic Database for Information Structure: Annotation and Retrieval" - Collaborative Research Center "Information Structure"	research	international	CSI
June 23-30, 2007, Prague, Czech Republic	LAW (Linguistic Annotation Workshop) - ACL 2007	research	international	CSI, UT
July 12, 2007, Blois, France	Seminar at Tour University	research	France	PAS
August 27-31, 2007, Antwerp, Belgium	Interspeech 2007 (Eurospeech)	research	international	UT, UAPV, FT
September 3-7, 2007, Antwerp, Belgium	SIGDIAL	research	international	UT
September 3-7, 2007, Pilsen, Czech Republic	TSD 2007	research	international	UAPV
September 13-16, 2007, Gniezno, Poland	38th Poznan Linguistic Meeting (PLM 2007)	research	international	PAS, PJIIT
October 5-7, 2007, Poznan, Poland	LTC 2007 - 3rd Language & Technology Conference: Human Language Technologies	research	international	PAS, PJIIT
November 22-23 2007, Basel, Switzerland	Nomina2007 Proper Names in Spoken Language	research	international	PAS
December 9-13, 2007, Kyoto, Japan	Automatic Speech Recognition and Understanding IEEE workshop (ASRU)	research	international	UAPV, UT
March 12-19, 2008, Warsaw, Poland	Seminar on Semantics at Warsaw University	research	Poland	PAS
April 23, 2008, Warsaw, Poland	Institute seminar at PAS	research	Poland	PAS
March 28 - April 09, 2008, Las Vegas, USA	IEEE ICASSP 2008	research	international	UT, FT
May 26-June 1, 2008, Marrakech, Morocco	LREC 2008 LAW II	research	international	RWTH, UT, PAS, PJIIT
June 16-18, 2008, Zakopane, Poland	New Trends in Intelligent Information Processing and Web Mining, IIS 2008 SLUDS at IIS 2008	research	international	PAS, PJIIT all
June 2008, Avignon, France	JEP/TALN	research	France	UAPV, FT
August 18-23, 2008, Manchester, UK	CCL COLING 2008	research	international	UT

Planned/ actual dates and place	Event	Type of audience	Countries addressed	Partner(s) involved
August 24, 2008, Manchester, UK	TextGraphs-3 at COLING 2008	research	International	UT
August 18-20, 2008, New York, New York	SpeechTek 2008	technology	International	UT, LOQ
August 28-31, 2008, Hamburg, Germany	3rd International Conference "Perspectives on Slavistics"	research	International	PAS
September 8-12, 2008, Brno, Czech Republic	TSD 2008	research	international	PAS, PJIIT
September 22-26, 2008, Brisbane, Australia	Interspeech 2008 (ICSLP)	research	international	RWTH, FT
September 25-27, 2008, Bydgoszcz, Poland	Polyslav XII	research	International	PAS
June 19-20, 2008, Columbus, Ohio	Spoken Language Technology IEEE/ACL workshop SigDial	research	international	UT
October 26-30, 2008, Napa Valley, California	CIKM	research	International	UT
November 12, 2008, Warsaw, Poland	Conference (Tutorial) at PAS	research	Poland	PJIIT
September 6-10, 2009, Brighton, UK	INTERSPEECH 2009 – EUROSPEECH:	research	international	
2009, Taiwan	IEEE ICASSP 2009	research	international	
2009	HLT 2009	research	international	

A few events attended during the first year were particularly remarkable:

- **DECALOG:** LUNA project co-organized the DECALOG workshop, the 2007 edition of the SemDial workshop series, held at the University of Trento in Rovereto (Italy) on May 30-June 1, 2007;
- **Natural Language Processing Seminar** organized by the Institute of Computer Science – Polish Academy of Sciences in Warsaw: one day was dedicated to LUNA project (April 23, 2007);
- **TSD 2007:** Prof. Renato De Mori, LUNA Technical Manager, presented LUNA project as invited speaker;
- A representative of the project (Kepa Joseba Rodriguez, CSI) was invited to present the LUNA annotation scheme in Potsdam, Germany, at the Collaborative Research Center SFB 632 "Information Structure" for the project "**Linguistic Database for Information Structure: Annotation and Retrieval**";
- Three talks were given by Prof. Giuseppe Riccardi (UT):
 - **W3C Meeting** (expert Invited Talk) Roncade, June 2007 "Next Generation Conversational Interfaces: Evolution or Revolution"

- Tutorial at **Interspeech 2007** (August 2007): “Talking to Computers: from Speech Sounds to Human Computer Interaction”;
- **AT&T Labs Research** (September 2007) The LUNA Project: spoken Language Understanding in multilingual communication systems

Significant events of the 2nd year are the following:

- **Automatic Speech Recognition and Understanding IEEE workshop** (ASRU), December 2007: Renato De Mori (UAPV) was invited as key-note speaker, Giuseppe Riccardi (UT) was an invited panelist: “Research Challenges for Spoken Language Understanding”; in addition, three LUNA papers were presented at the event.
- **SLUDS at IIS 2008**, Zakopane, Poland: Polish partners of LUNA project organized the one-day workshop on Spoken Language Understanding and Dialogue Systems in Zakopane, Poland. The workshop was held together with 16th International Conference Intelligent Information Systems. There have been 6 talks delivered. It also involved the DICIT project.
- G. Riccardi was also a keynote speaker at **Czech Academy of Science** (April 2008): “Research and Innovation: Case Studies in Speech Technology”

In the third year LUNA project will possibly organize a special session during a major speech conference, and is submitting a proposal for INTERSPEECH 2011 with R. De Mori as general co-chairman.

2.4.2 Scientific journals

This section gives a preliminary list of targeted scientific journals; the list will be updated during the project lifetime in order to reflect all the different kinds of dissemination activities:

- IEEE Transactions on Audio, Speech and Language Processing
- IEEE Signal Processing Magazine
- Speech Communication, Elsevier
- Computer Speech and Language, Elsevier
- Computational Linguistics, MIT Press

A journal paper on spoken opinion analysis has been submitted to the IEEE Transactions on Audio, Speech and Language Processing.

A review paper on Spoken language Understanding (“Spoken Language Understanding for Conversational Systems”) has been published in the **IEEE Signal Processing Magazine Special Issue on Spoken Language Technologies**, Vol. 25, No. 3, pp. 50-58, May 2008.

The complete list of papers published so far can be found in paragraph 2.4.3.

2.4.3 Publications

Scientific papers

Christophe Servan, Christian Raymond; Frédéric Béchet, Pascal Nocéra. “Conceptual decoding from word lattices: application to the spoken dialogue corpus MEDIA”. Proceedings of the International Conference on Spoken Language Processing, 2006

Géraldine Damnati, Frédéric Béchet, Renato De Mori. "Experiments on the France Telecom 3000 Voice Agency corpus: academic research on an industrial spoken dialog system". Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, HLT 2007, Rochester, NY, 2007

Géraldine Damnati, Frédéric Béchet, Renato De Mori. "Spoken Language Understanding strategies on the France Telecom 3000 voice agency corpus". ICASSP 2007, Honolulu, Hawaii, USA, 2007

Christian Raymond, Giuseppe Riccardi, Kepa Joseba Rodríguez, Joanna Wisniewska. "The LUNA Corpus: an Annotation Scheme for a Multi-domain Multi-lingual Dialogue Corpus". Abstract for poster session. Proc. 11th Workshop on the Semantics and Pragmatics of Dialogue (DECALOG'07), Rovereto (TN), Italy, May 2007

Alessandro Moschitti, Silvia Quarteroni, Roberto Basili and Suresh Manandhar. "Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classification". Proceedings of the 45th Conference of the Association for Computational Linguistics: Human Language Technology (ACL), Prague, Czech Republic, June 2007.

Mona Diab, Alessandro Moschitti, Daniele Pighin, Semantic Role Labeling Systems for Arabic Language using Kernel Methods. In proceedings of the 46th Conference of the Association for Computational Linguistics: Human Language Technology (ACL). Main Conference. Prague, Czech Republic, June 2007.

Kepa Joseba Rodríguez, Stefanie Dipper, Michael Götze, Massimo Poesio, Giuseppe Riccardi, Christian Raymond, Joanna Wisniewska. "Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus". Proceedings of the Linguistic Annotation Workshop at the ACL'07 (LAW-07), Prague, Czech Republic, June 2007

Christian Raymond, Giuseppe Riccardi. "Generative and Discriminative Algorithms for Spoken Language Understanding". Proc. European Conference on Speech Communication and Technology, INTERSPEECH 07, Antwerp, Belgium, August 2007

Nathalie Camelin, Frédéric Béchet, Géraldine Damnati, Renato De Mori. "Speech Mining in Noisy Audio Message Corpus". Proc. European Conference on Speech Communication and Technology, INTERSPEECH 07, Antwerp, Belgium, August 2007

Bogdan Minescu, Géraldine Damnati, Frédéric Béchet, Renato De Mori. "Conditional use of Word Lattices, Confusion Networks and 1-best string hypotheses in a Sequential Interpretation Strategy". Proc. European Conference on Speech Communication and Technology, INTERSPEECH 07, Antwerp, Belgium, August 2007

Agnieszka Mykowiecka, Krzysztof Marasek, Małgorzata Marciniak, Joanna Rabięga-Wisniewska, Ryszard Gubrynowicz. "Annotation of Polish spoken dialogs in LUNA project". 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznan, Poland, October 2007

Agnieszka Mykowiecka, Krzysztof Marasek, Małgorzata Marciniak, Joanna Rabięga-Wisniewska, Ryszard Gubrynowicz, Annotation of Polish spoken dialogs in LUNA project, SP1-1, 3rd Language&Technology Conference LTC'07,Poznań, October 5-7,2007.

Sebastian Vargas and giuseppe Riccardi. "A Data-Centric Architecture for Data-Driven Spoken Dialog Systems". Proc. IEEE ASRU Workshop, Kyoto, 2007.

Alessandro Moschitti, Giuseppe Riccardi and Christian Raymond. "Spoken Language Understanding with Kernels for Syntactic/Semantic Structures", Proc. IEEE ASRU, Automatic Speech Recognition and Understanding Workshop, Kyoto, Japan, December 2007

Fabrice Lefèvre and Renato de Mori. "Unsupervised State Clustering For Stochastic Dialog Management". Proc. IEEE ASRU, Automatic Speech Recognition and Understanding Workshop, Kyoto, Japan, December 2007

Alessandro Moschitti and Fabio Massimo Zanzotto. "Fast and Effective Kernels for Relational Learning from Texts". Proceedings of The 24th Annual International Conference on Machine Learning (ICML 2007), Corvallis, OR, USA.

Christian Raymond, Frédéric Béchet, Renato De Mori, Géraldine Damnati. "On the use of finite state transducers for semantic interpretation". Speech Communication, 2006, 48, 288-304

Christian Raymond, Frédéric Béchet, Nathalie Camelin, Renato De Mori, Géraldine Damnati. "Sequential decision strategies for machine interpretation of speech". IEEE Transactions on Speech and Audio Processing, 2007, 15, 162

Stefan Hahn, Patrick Lehnen, Christian Raymond, and Hermann Ney, "A comparison of various methods for concept tagging for spoken language understanding," in Proc. of the Sixth Int. Conf. on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 26-28, 2008.

Rodriguez K., Raymond C. and Riccardi G., "Active Annotation in the LUNA Italian Corpus of Spontaneous Dialogues". Proc. Language Resources and Evaluation (LREC), Marrakech, Morocco, May 26-28, 2008.

Krzysztof Marasek, Ryszard Gubrynowicz, "Design and data collection for spoken Polish dialogs database", Proc. 6th Conf. on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 26-28, 2008.

Marie-Jean MEURS, Frédéric DUVERT, Frédéric BECHET, Fabrice LEFEVRE and Renato DE MORI, "Semantic Frame Annotation on the French MEDIA corpus", Proceedings of the Language Resources and Evaluation Conference, 2008, Marrakech, Morocco, May 26-28, 2008.

Mykowiecka A., Marasek K., Marciniak M., Rabięga-Wiśniewska J., Gubrynowicz R., "On Construction of Polish Spoken Dialogs Corpus", The 2nd Linguistic Annotation Workshop (The LAW II) held in conjunction with LREC 2008, Marrakech, Morocco, May 26-28, 2008.

Sebastian V, Riccardi G. and Quarteroni S., "Persistent Information State in a Data-Centric Architecture", SIGdial Workshop on Discourse and Dialogue, Columbus, 2008.

Raymond C. and Riccardi G., "Learning with Noisy Supervision for Spoken Language Understanding" Proc. IEEE ICASSP 2008, Las Vegas, 2008.

Isabelle Jars, Franck Panaget, "Improving Spoken Language Understanding with Information Retrieval and Active Learning methods", ICASSP 2008, Las Vegas, USA, 2008

Frederic Duvert, Marie-Jean Meurs, Christophe Servan, Frederic Bechet, Fabrice Lefevre, Renato De Mori, "Semantic composition process in a speech understanding system", Proceedings of the International Conference on Acoustic Speech and Signal Processing, ICASSP 2008, Las Vegas, USA, 2008.

L. Barrault, C. Servan, D. Matrouf, G. Linarès and R. De Mori, "Frame-Based Acoustic Feature Integration for Speech Understanding", 2008 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'08), Las Vegas, USA, 2008

Alessandro Moschitti, Silvia Quarteroni, "Kernels on Linguistic Structures for Answer Extraction". In proceedings of the 46th Conference of the Association for Computational Linguistics (ACL'08). Short Paper Section. Columbus, OH, USA, June 2008.

Alessandro Moschitti, "Kernel Methods, Syntax and Semantics for Relational Text Categorization". In proceeding of ACM 17th Conference on Information and Knowledge Management (CIKM). Napa Valley, California, 2008.

Yannick Versley, Alessandro Moschitti, Massimo Poesio, Xiaofeng Yang, "Coreference Systems based on Kernels Methods". In Proceedings of The 22nd International Conference on Computational Linguistics (Coling '08), Manchester, England, 2008.

Alessandro Moschitti and Fabio Massimo Zanzotto, "Encoding Tree Pair-based Graphs in Learning Algorithms: the Textual Entailment Recognition Case", Proceedings of TextGraphs-3: Graph-based Algorithms for Natural Language Processing – Workshop held in Coling '08 Conference, Manchester, England 2008.

Marciniak M., Rabięga-Wiśniewska J., Mykowiecka A., "Proper Names in Dialogs from the Warsaw Transportation Call Center". Intelligent Information Systems XVI Proceedings of the International IIS'08 Conference held in Zakopane, Poland, June 16-18, 2008. 517-525.

Géraldine Damnati, Frédéric Béchet, Renato De Mori, "First implementation of the LUNA Spoken Language Understanding strategy on a telephone service application", SLUDS'08, IIS Workshop on Spoken Language Understanding and Dialogue Systems, Zakopane, Poland, 2008.

Danijel Koržinek, Łukasz Brocki, Ryszard Gubrynowicz, Krzysztof Marasek, "Wizard of Oz Experiment for a Telephony-Based City Transport Dialog System", Proc. Intelligent Information Systems Conference, SLUDS Workshop, Zakopane, 16-18 June, 2008

Frédéric Duvert, Marie-Jean Meurs, Christophe Servan, Frédéric Béchet, Fabrice Lefèvre and Renato De Mori, "Semantic composition process in a spoken language understanding system", Workshop on Spoken Language Understanding and Dialogue Systems, Intelligent Information System Conference, Zakopane, June 2008

Marie-Jean Meurs, Frédéric Duvert, Fabrice Lefèvre and Renato De Mori, "Markov Logic Networks for Spoken Language Interpretation", Workshop on Spoken Language Understanding and Dialogue Systems, Intelligent Information System Conference, Zakopane, June 2008

Agnieszka Mykowiecka, Małgorzata Marciniak and Katarzyna Głowińska, "Automatic Semantic Annotation of Polish Dialogue Corpus", Lecture Notes in Computer Science, Springer, Berlin / Heidelberg, Volume 5246/2008.

Geraldine Damnati, Frederic Béchet, Renato De Mori, "LUNA : Compréhension en contexte pour le dialogue oral", JEP'08, Avignon, France, 2008.

Bogdan Minescu & Geraldine Damnati, "Construction et exploitation des réseaux de confusion dans le contexte d'une application de dialogue en langage naturel", JEP'08, Avignon, France, 2008.

Frédéric Duvert, Marie-Jean Meurs, Christophe Servan, Frédéric Béchet, Fabrice Lefèvre and Renato De Mori, "Composition sémantique pour la compréhension de la parole dans le cadre de dialogue", 2008 Journées d'Etude sur la Parole, Avignon June 2008

Marie-Jean Meurs, Frédéric Duvert, Frédéric Béchet, Fabrice Lefèvre and Renato De Mori, "Annotation en Frames Sémantiques du corpus de dialogue MEDIA", Conférence Traitement Automatique de la Langue Naturelle, TALN'08, Avignon June 2008

Krzysztof Marasek, Ryszard Gubrynowicz, "Construction of telephony Polish speech database within the framework of the LUNA project", tutorial given 12th November, 2008, at the Institute of Computer Sciences

Łukasz Brocki, Danijel Koržinek, Krzysztof Marasek, Telephony Based Voice Portal for a University, Speech Analysis, Synthesis and Recognition, Applications in Systems for Homeland Security 2008, Piechowice, September 8-12

Stefan Hahn, Patrick Lehnen, Hermann Ney, "System Combination for Spoken Language Understanding," in Proc. of Interspeech, Brisbane, Australia, September 2008

Georg Heigold, Patrick Lehnen, Ralf Schlueter, Hermann Ney, "On the Equivalence of Gaussian and Log-Linear HMMS," in Proc. of Interspeech, Brisbane, Australia, September 2008

Christophe Servan, Frederic Bechet, "Fast Call-Classification System Development Without In-Domain Training Data", Interspeech 2008, Brisbane, Australia, September 2008

Marie-Jean Meurs, Fabrice Lefèvre, Renato De Mori, "A Bayesian Approach to Semantic Composition for Spoken Language Interpretation", Interspeech 2008, Brisbane, Australia, September 2008

Presentations

Géraldine Damnati, Frédéric Béchet, Renato De Mori. "Spoken language understanding strategies on the France Telecom 3000 voice agency corpus". ICASSP 2007, Honolulu, Hawaii, USA, 2007

Joanna Rabięga-Wiśniewska, Małgorzata Marciniak, Agnieszka Mykowiecka. "Anotacja dialogów w projekcie LUNA" [Annotation of dialogues in LUNA project], presentation at ICS PAS seminar, 23th April 2007, Warsaw.

Joanna Rabięga-Wiśniewska. "Przygotowanie korpusu języka mówionego. O projekcie LUNA" [Collecting a corpus of spoken language. Project LUNA], Corpus linguistics session at The East-Central European School in the Humanities, 30th May 2007, Warsaw University, Poland.

Christian Raymond, Giuseppe Riccardi, Kepa Joseba Rodríguez, Joanna Wisniewska. "The LUNA Corpus: an Annotation Scheme for a Multi-domain Multi-lingual Dialogue Corpus". Poster presentation. The 11th Workshop on the Semantic and Pragmatics of Dialogue (DECALOG'07), Rovereto (TN), Italy, 2007

Kepa Joseba Rodriguez. "The LUNA Corpus: an annotation scheme for a multilingual multidomain dialogue corpus". Potsdam, Germany, 2007

Kepa Joseba Rodriguez, Stefanie Dipper, Michael Götze, Massimo Poesio, Giuseppe Riccardi, Christian Raymond, Joanna Wisniewska. "Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus". Poster presentation. The Linguistic Annotation Workshop at the ACL'07 (LAW-07), Prague, Czech Republic, 2007

Christian Raymond, Giuseppe Riccardi. "Generative and Discriminative Algorithms for Spoken Language Understanding". INTERSPEECH 07, Antwerp, Belgium, 2007

Joanna Rabiega-Wisniewska, Małgorzata Marciniak, Agnieszka Mykowiecka. "Overview of Annotation Tools for Polish Dialogs". 38th Poznan Linguistic Meeting, Gniezno, Poland

Krzysztof Marasek, Ryszard Gubrynowicz, "Polish human-human spoken dialog transcriptions - experience from LUNA project", 38th Poznań Linguistic Meeting, Gniezno, Poland.

Agnieszka Mykowiecka, Krzysztof Marasek, Małgorzata Marciniak, Joanna Rabiega-Wisniewska, Ryszard Gubrynowicz. "Annotation of Polish spoken dialogs in LUNA project". 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznan, Poland, 2007

Stefan Hahn, Patrick Lehnen, Christian Raymond, and Hermann Ney, "A comparison of various methods for concept tagging for spoken language understanding," in Proc. of the Sixth Int. Conf. on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 2008

Stefan Hahn, Patrick Lehnen, Hermann Ney, "System Combination for Spoken Language Understanding," in Proc. of Interspeech, Brisbane, Australia, September 2008

Joanna Rabiega-Wisniewska, Towards a description of Polish proper names, University of Tours, France, December 2007.

Joanna Rabiega-Wisniewska, On a Description of Polish Proper Names in a City Transportation System, nomina2007 Proper Names in Spoken Language, November 22-23 2007, Basel, Switzerland.

Joanna Rabiega-Wisniewska, IST LUNA – Cele projektu VI Programu Ramowego UE (Goals of the European Project IST LUNA), Warsaw University, seminar on semantics, March 2008.

Joanna Rabiega-Wisniewska, Polski korpus dialogów w projekcie LUNA (Polish Corpus of dialogs in LUNA project), Warsaw University, seminar on semantics, March 2008.

Małgorzata Marciniak, Agnieszka Mykowiecka, Joanna Rabiega-Wisniewska, Konstrukcja anotowanego korpusu dialogów (The structure of an annotated corpus of dialogs), PAS seminar, April 2008.

Agnieszka Mykowiecka, Krzysztof Marasek, Małgorzata Marciniak, Joanna Rabiega-Wisniewska, Ryszard Gubrynowicz, On Construction of Polish Spoken Dialogs Corpus, The 2nd Linguistic Annotation Workshop (The LAW II) held in conjunction with LREC 2008, Marrakech, Morocco, May 26-28, 2008.

Marciniak M., Rabiega-Wisniewska J., Mykowiecka A., Proper Names in Dialogs from the Warsaw Transportation Call Center. Intelligent Information Systems XVI Proceedings of the International IIS'08 Conference held in Zakopane, Poland, June 16-18, 2008.

Małgorzata Marciniak, Annotation of Polish spoken dialogs in LUNA project, University of Tours, France, July 2007.

Joanna Rabiega-Wisniewska, The analysis of speakers' errors in the Polish dialog corpus, 3rd International Conference "Perspectives on Slavistics", Hamburg, Germany, August 30, 2008.

Agnieszka Mykowiecka, Małgorzata Marciniak and Katarzyna Głowińska, Automatic Semantic Annotation of Polish Dialogue Corpus, TSD 2008, September 2008.

2.5 Industrial dissemination

Dissemination activities towards the industrial community is being made in the following ways:

1. Participations in events such as exhibitions, fairs, trade shows, conferences. This activity on one hand meets the strong interest of industrial partners in the exploitation and, on the other hand, helps in exploiting their network of contacts and potential customers. Main target conferences are the following:
 - AVIOS/SpeechTEK
 - VoiceWorld Europe
 - Vocal Expo
 - VoiceDays

In August 2008 LOQ and UT attended the **SpeechTEK conference**. UT took part to an invited Multimodal Demo Session: "Voice Multimodal Application Framework"

2. Internal dissemination in Telecom Italia and France Telecom groups:
 - On January 9, 2007, France Telecom R&D presented the LUNA project in a seminar targeted at marketing, innovation and development managers within the FT Group and the France Telecom scientific community. The title of the presentation was "Spoken Language understanding in the context of dialog applications".
3. Press releases to be published occasionally, distributed through the Loquendo newsletter, which reaches about 8.000 subscribers. These press releases are available on project's and partners' websites as well.
4. Publication of articles on specialized press.

2.6 Dissemination towards general public

In addition to the publications for the academic and industrial community, the LUNA consortium also addresses mass media.

An article was published at the very beginning of the project on an Italian magazine mainly targeted to Public Administrations and local companies.

During 2008, an Italian magazine targeting Telecommunication industry (Corriere delle Comunicazioni) published an article about speech technologies, including a description of the research activities made in LUNA as well.

The LUNA project has been selected to attend the **ICT 2008** exhibition in Lyon.

Finally, LOQ and CSI are planning to present the project at a local event in Italy during the third year.

2.7 Link with other activities

Most of the partners are engaged in other national and international R&D activities concerning speech technologies; this represents a further channel for sharing information about LUNA research activities and outcomes:

- French National Technolanguage project: MEDIA (Evaluation of spoken language understanding).

- FP6 N0Es: PASCAL that associates the European research in pattern analysis, statistical modelling and computational learning.
- FP6 IST projects: DIVINES (LOQ, FT, UAPV) – ended January 2007, HIWIRE (LOQ) – ended June 2007, TC-STAR (RWTH) – ended March 2007, COMPANIONS (LOQ, FT)
- Marie Curie Excellence Research grant (UT): Adaptive and Meaning Machines (ADAMACH)
- Marie Curie International Reintegration Grant (UT): teChnology and Architecture for Spoken dialog Systems (CASA)

In particular, COMPANIONS consortium is going to reuse the annotation protocol defined in LUNA project in order to annotate data that have been collected in a Wizard of Oz approach. The LUNA annotation scheme that is defined in deliverable D1.3 is going to be one of the basis for the annotation protocol that has to be defined in the COMPANIONS project.

FT is involved in the CLASSIC project, a STREP that started in March 2008 for 3 years. The objective of this project is to develop an end-to-end statistical dialogue architecture with a complete chain of components with explicit management of uncertainty. It is foreseen for the last year of the CLASSIC project that the LUNA SLU components will be integrated and evaluated in this end-to-end architecture once the LUNA project is completed.

Moreover, UAPV is involved in a new research project called EPAC, funded by the French agency ANR, which started in January 2007. This project is focused on the study of conversational speech in large databases of broadcast news speech. The LUNA annotation schemes that are being defined in the deliverable D1.3 for the annotation of the Human-Human and Human-Machine dialogues collected by the LUNA partners are going to be used in EPAC, for the annotation of the conversational speech in interviews and debates extracted from the EPAC broadcast news speech database.

LUNA project is also collaborating with the DICIT project; the two projects held a panel on Evaluation of Spoken Language Understanding and Dialogue Systems under the SLUDS Workshop at the Intelligent Information Systems Conference – IIS 2008 (June 18, 2008, Zakopane, Poland).

In addition, universities have strong academic links on a global scale, in particular with Canadian and US academies and institutions (Mc Gill University, Stanford Research Institute, International Computer Science Institute, University of California, Berkeley...); some partners also have an active role in outstanding organizations in the speech communities (IEEE Speech and Language Technical Committee, Editorial board: ACM Transactions on Speech and Natural Language, Computer Speech and Language, Speech Communication).

The LUNA project attended in December 2006 the ALL PROJECT workshop organized by the European Commission DG Info Unit E1 “Interfaces”; the objectives and the expected results of the project were presented to the community of institutions involved in projects funded under the same strategic objective of LUNA.

3 Preliminary exploitation plan

The exploitation task aims to aggregate in an organic view all the results and to understand how they can be mapped to actual market necessities and existing realities.

There are two important paths of the exploitation of the results of the project: one through the academia and research users, usually the simplest one, and the other one through the industry and commercial users.

Although the project has a strong component for research activities, the consortium will also focus on commercial exploitation, as there is a clear willingness of the industrial and service provider partners to improve their solutions integrating the project results and exploiting the acquired experience.

On the industrial side, the two private sector technology providers, namely FT and LOQ, have a strong interest in exploiting the competitive advantage and technical know-how offered to them through LUNA project. In addition, their significant experience in IPR management will ensure that new research results will be carefully protected by patents, where appropriate.

Both the aforementioned industrial partners will exploit robust speech understanding solutions in their commercial products and services.

LOQ is continuously improving its ASR technology also by integrating remarkable intermediate results of the project in the product. A first example is the availability of the Polish language pack in the company commercial offering. As the project evolves, all outcomes are being evaluated in order to identify mature results suitable for commercial exploitation, which can become part of the LOQ product portfolio. Currently, LOQ is exploiting the work carried out in the second year, by enabling the commercial ASR product to generate extended results based on the word lattice format.

Technical results will also be integrated, after a complete test on several languages and a software optimization in the FT system. FT is developing and deploying its own spoken dialogue solution. Vocal services are already deployed, on the basis of this solution, towards the general public. For instance the 3000 FT vocal agency was the first deployed vocal service at France Telecom exploiting natural language technologies. It was launched in 2005 and is widely used by residential customers. Other services are regularly launched on the basis of the FT dialogue solution (for instance the 1013 call routing application for residential customers' after-sales service deployed in 2008). The algorithms and methods explored in the LUNA project will be tested and validated on real data collected in real conditions, insuring that the results are directly transposable to the deployed systems. Prototypes on the basis of the FT3000 application and using some LUNA results have already been developed. Robustness and performance are naturally essential for such applications to be deployed. But from an industrial point of view, the cost of an application development and maintenance is crucial. Data collection and annotation has a significant impact in the overall costs and important results from LUNA such as active learning (appropriate selection of data to be annotated and integrated in the training processes) represent a strong expectation from FT. In applications deployed so far, semantic knowledge is implemented through hand crafted rules, necessitating high human expertise in their definition. The statistical models explored within the LUNA project (e.g. the large variety of statistical approaches evaluated in WP2 for word-to-concept annotation) can represent an alternative towards the compromise between human expertise and data annotation cost. Moreover, the semantic representation adopted through the LUNA project allows representing available data at a higher level which facilitates their reusability for another application. Cross-application portability and bootstrapping are crucial issues for reducing the cost of a new application development, which constitutes strong attempts from business units. Finally, FT has strong international implications, through FTR&D laboratory implantations abroad and cross-language portability constitutes a strong demand from the various

labs. For instance, ongoing collaborations between FT and TPSA in Poland will benefit from the work achieved within the LUNA project on Polish language.

On the other hand, CSI, being in charge of the provision of all ICT-based services for the local public administration will exploit the gained expertise and the results of the project by developing new services for its customers.

The academic and research partners in LUNA will exploit the project results through dissemination activities such as publications in international journals, conferences and workshops. Furthermore, the results will be exploited in graduation and post graduation studies, and other research projects.

A major result to underline is that the PJIT team has agreed with the direction of the Warsaw Transport Authority to follow-up the common work on the automated call center giving information related to city transportation in Warsaw. The results of experiments with WOZ application will be used in the design and developing the automatic information system based on advance dialog service supported by specific semantic knowledge.

These results will be included to lectures on dialog systems given to PJIT students as a practical application of them in real-life conditions. A more limited application in the domain of call routing system is applied in PJIT exchange where the phone number may be requested by name

The list of possible exploitable results is given in the following table:

Overview table

Exploitable Knowledge (description)	Exploitable product(s) or measure(s)	Sector(s) of application	Timetable for commercial use	Patents or other IPR protection	Owner & Other Partner(s) involved
Polish Language Pack	Loquendo ASR in Polish language	All sectors	Already available (2007)	-	LOQ
Wizard of Oz system	Loquendo VoxNauta and CSI's web application / Simulated Help Desk on infrastructural issues	Research	2009-2010	-	LOQ, CSI
Annotation protocol	Annotation manual/scheme	Research	2007	Copyrights	All
Italian annotated corpus	500 spoken human-human dialogues 500 WoZ dialogues	Research	2009	Database rights (complying with Privacy laws)	CSI
Polish annotated corpus	500 spoken dialogs human-human	Research	2008-2009	Database rights	PJIT, PAS
Semantic concept tags component	Core component for natural language processing	Concept tagging	2010	-	RWTH, UT

Polish Language pack

The language pack developed by LOQ has already been integrated in Loquendo ASR commercial product.

Wizard of Oz system

The Wizard of Oz system has been jointly developed by CSI and LOQ. The architecture is composed of Loquendo VoxNauta platform and CSI's web application. The web application is composed by two modules: the first interacts with VoxNauta platform, the second is a web-application that permits to the human "Wizard" to choose the subsequent vocal prompt. The audio signal is split in order to have the first signal as input of VoxNauta and the second as input of the human Wizard. The WoZ utility could be used in order to gather Human-Machine dialogues in different languages for future corpora collections.

Annotation protocol

The annotation protocol, released as Deliverable D3.1, is available to the research community for further research projects. Currently the protocol is going to be used at least partially by the IST project COMPANIONS.

Italian corpus

Italian human-human dialogues have been gathered in the Help Desk of the CSI-Piedmont and have been transcribed and annotated following the annotation protocol presented in the deliverable D1.3. This human-human dialogue corpus together with the WoZ dialogues may be used with different purposes like re-training ASR systems or evaluation of conversational solutions.

A further application is the academic research in corpus linguistics and stand-off corpus formats for speech corpora. In this area CSI and UT have taken effort together with the collaborative research center SFB 632 "Information Structure" (Germany).

At the same time we have detected interest in the use of the corpus for linguistic research, but we have not taken special efforts in this direction because the issues concerning to privacy laws should be clarified previously.

Polish annotated corpus

Polish human-human dialogs have been recorded at the call center of Warsaw Transportation Authority (WTA) and cover 500 information seeking dialogs (spontaneous speech) between call center operators and various callers using fixed and mobile phone networks. One channel speech recordings are stored in non-compressed form (16 kHz, 16 bits) and are rich annotated. The annotation includes time-aligned turns, detailed transliteration with acoustic events marks and multi-level language annotations (words, turns, POS, chunks, dialog acts, etc.) as standardized in LUNA project. This first database of Polish spontaneous dialogs can be used for training of speech recognizers, studies on Polish language, etc.

The database collected within LUNA framework and the gained insight into the work of the city transportation information structure will be exploited in design and developing the automated system in the WTA call center.

Semantic concept tags component

The semantic concept tags component extracts information from a spoken utterance in terms of semantic concepts. A semantic concept (in the following briefly termed as concept) is defined as the smallest unit of meaning that is relevant to a specific task (Levin and Pieraccini, 1995). In LUNA a concept is a composition of an attribute name and an attribute value, e.g. attribute name is "location" and the value is "Paris". A spoken utterance represented by a word lattice is processed to a concept lattice defined by the LUNA lattice format. The semantic concept tags component can be trained on corpora in the LUNA corpora format containing the spoken utterances, the attribute names and its values. This semantic concept tags component is one of the core parts of LUNA. It can be used as a part of a dialog system as well as for information extraction from written and

spoken text. For both, commercial and scientific use, it can be used as a reference implementation. A first demonstration of the concept tags component is realized within the prototype of deliverable 2.1.

4 Standardization plan

The project is going to give emphasis to the standardization activity. One of the objectives of LUNA project is to produce significant results which can be proposed to the standardization bodies. One aim of the project is, in fact, to follow and contribute to existing and coming standardization processes in the area of speech recognition and multimodal interfaces, with a particular focus on mark-up languages to annotate the output of speech or multimodal device output.

In particular, LOQ and FT are directly involved in national and international standardization committees.

W3C in its Voice Browser Working Group is responsible of numerous efforts devoted to standardization of speech and Spoken Language Understanding (SLU) specifications. Worth mentioning are the efforts for standardizing semantic interpretation representation languages. The W3C's Candidate Recommendation for "Semantic Interpretation for Speech Recognition" (SISR) has been published in February 2007 as Proposed Recommendation. The language, based on ECMA-327, an efficient version of ECMA-262 (also known as JavaScript), allows including in a speech recognition grammar a semantic interpreter in order to obtain powerful semantic results, very useful for a SLU application. SISR is the standard Semantic Interpretation of Speech Recognition Grammar Specification (SRGS) Version 1.0, which is the W3C Recommendation (March 2004) about speech recognition grammars.

Another effort is the creation of standard language models for SLU. A "Stochastic Language Models (N-Gram) Specification" Working Draft is available since January 2001, but currently there isn't any activity on this specification.

In this working group Loquendo followed the development of the Pronunciation Lexicon Specification (PLS) version 1.0, which recently demonstrated the implementability and it is close to become a W3C Recommendation. PLS might be of interest inside the project

W3C in its Multimodal Interaction Working Group is standardizing the Extensible MultiModal Annotation markup language (EMMA) which is targeted to represent and annotate semantic results from speech recognition or other modalities (e.g. gesture, pen). EMMA allows the composition of several input modalities. The ASR output, as word lattice, can be immediately mapped into this framework. The specification EMMA, version 1.0, is very relevant to represent semantic input for speech in LUNA dialog manager. The work on the EMMA specification was followed to the point to publicly demonstrate the implementability of EMMA inside Loquendo speech products, as well as in the integration of EMMA in the LUNA demonstrator done by UT.

LUNA project will further evaluate, use and validate EMMA language, and if needed to extend it to better suite the goals of the LUNA project. This will enable a more immediate availability of results to the research and industrial community.

Another activity in W3C is the new Incubator about W3C Emotion Incubator Group (LOQ is Member), whose goal is to define guidelines for a language to represent emotions to be used in many different contexts such as production of more expressive speech and recognition of emotions. Even if this activity is at an early stage, it might be important to monitor it.

Other targets are represented by the various bodies dealing with multimodality. Relevant to the project partners are the:

- VoiceXML Forum, which supports XHTML+Voice and where LOQ is present as Member of the Board of Directors.

- SALT Forum, of which LOQ is a Member.
- IETF, which has recently started an activity on Distributed Multimodal Synchronization Protocol (DMSP) and Media Server Control Protocol (MSCP).

Another important aspect of standardization is the annotation of corpora for SLU, which is a critical process to enable SLU research. At the current time, there are no initiatives of this kind and the consortium would encourage the standard bodies to engage in this initiative.

5 Annex: Contractual and Internal Requirements

From chapter IV.6 Publications, Press Releases and Reports to the Commission

Publications

Each *Party* shall have the right to publish or allow the publishing of data which constitute such *Party's Knowledge, Pre-existing Know-how* or confidential information it owns subject to and in accordance with *EU Contract Article II.33.3*.

For the avoidance of doubt no *Party* shall have the right to publish or allow the publishing of data, which constitute *Knowledge, Pre-existing know-how* or confidential information of another *Party* even where such data is amalgamated with such first *Party's Knowledge, Pre-existing know-how* or other information, document or material. Any such publication without such other *Party's* written agreement justifies, save for further remedies, objection to the publication by the *Party* concerned in accordance with *EU Contract Article II.33.3*.

Where a *Party* wishes to publish or allow the publication of data that includes *Knowledge, Pre-existing Know-how* or confidential information of another *Party*, approval for such inclusion shall be obtained from the appropriate *Party* owning such information or affected by the inclusion. The approval shall be sought by a written notice prior to notifying the *Commission* and the other contractors of planned publication according to *EU Contract Article II.33.3*.

For the avoidance of doubt the provisions of this Section IV.6 shall be subject and without prejudice to the confidentiality provisions of Section IV. 5, which shall prevail over the provisions of this Section IV.6.

Disclaimer & marking of confidential information provided to the Commission

In addition to the *EU Contract Article II.12*,

- (i) all information provided to the *Commission*, publications and press releases shall have a disclaimer saying "The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability";
- (ii) confidential information provided to the *Commission* will be marked stating the information is confidential and may be used only for information purposes by Community Institutions to whom the *Commission* has supplied it.