SIXTH FRAMEWORK PROGRAMME
PRIORITY 2
Information Society Technology

LUNA

spoken Language UNderstanding in multilinguAl communication systems

Project n. 33549

D 6.3 – LUNA functionalities description

Due date of deliverable: July 3, 2008

Actual submission date: August 7, 2008

Start Date of Project September 4, 2006 **Duration** 36 months

Organization name of lead contractor for this deliverable LOQ

Version 1.0

Dissemination level [PU]

Authors and reviewers tables

Authors	Company Name
Silvia Mosso	LOQ
Renato De Mori	UAPV
Frederic Bechet	UAPV
Stefan Hahn	RWTH
Ryszard Gubrynowicz	PJIIT
Geraldine Damnati	FT
Giuseppe Riccardi	UT

Reviewers	Company Name
Paolo Coppo	LOQ

Abstract

Speech technologies are being used increasingly in a variety of over-the-phone applications and most people by now have experienced at least one automated phone service. The quality of the human-to-computer interaction is still far from being enjoyable and effective, as many call center interfaces maintain the rigid structure of the touch-tone menus from which they evolved.
To improve the usefulness and acceptability of automatic dialogue systems, a solution is to raise the level of intelligence of automatic systems up to Spoken Language Understanding (SLU). The deployment of SLU technology offers the customers a more flexible navigation, reducing the number of steps required for callers to achieve a desired objective.

The objective of LUNA (spoken Language UNderstanding in multilinguAl communication systems) project is to realize Spoken Language Understanding components for the fast deployment of robust automated telephone services. The ambitious goal of the project is to position itself at the forefront of the third generation of spoken language interfaces.

TABLE OF CONTENTS

1 Introduction

Speech recognition technology is being used increasingly in a variety of over-the-phone applications in telecommunications, transportation, financial, and other industries. Most people by now have experienced at least one automated phone service, in which questions are posed by a machine to callers seeking help or information, and the callers' spoken answers are understood and handled by that machine. The quality of the human-to-computer interaction is still far from being enjoyable and effective, as many call center interfaces still maintain the structure of the touch-tone menus from which they evolved. The main reason for menu-driven systems was the poor quality of the speech technologies, which have now made important advances.
To improve the usefulness and acceptability of automatic dialogue systems, a solution is to raise the level of intelligence of automatic systems up to Spoken Language Understanding (SLU). Instead of the rigid system initiative, the deployment of SLU technology can offer the customers a more flexible navigation, reducing the number of steps required for callers to achieve a desired objective.
The movement towards less structured interactions is likely to improve automation rates and shorten call times while maintaining high levels of customer satisfaction.

Telephone services perform actions dependent on user intentions expressed by spoken sentences. Partial automation can be achieved by having systems capable of extracting meaning from spoken sentences in an application domain. Interesting results can be obtained even if automatic speech recognition is not perfect by exploiting redundancies in the expression of concepts and opinions, by using constraints of semantic coherence in the interpretation process and by introducing suitable confidence indicators of automatic comprehension.

The objective of this document is to present the main results and functionalities developed in the LUNA (spoken Language UNderstanding in multilinguAl communication systems) project to a public of non-experts in the field of Spoken Language Understanding.

2 The LUNA project

2.1 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. The aim of the LUNA project is to realize Spoken Language Understanding components for the fast deployment of robust automated telephone services. The ambitious goal of the project is to position itself at the forefront of the third generation of spoken language interfaces.

The consortium is made up of the following organisations: Loquendo (coordinator, LOQ), RWTH Aachen (RWTH), the University of Trento (UT), the University of Avignon (UAPV), France Telecom (FT), CSI-Piemonte (CSI), the Polish-Japanese Institute of Information Technology (PJIIT) and the Institute of Computer Science of the Polish Academy of Sciences (PAS).

A set of challenging scientific problems are addressed in order to make SLU systems more robust and capable of dealing with the spontaneous enunciation of concepts.
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.

The basic functionalities of the LUNA project are:

- generation of semantic constituent hypotheses from the speech signal using different methods,
- composition of constituents into semantic structures in a dialogue turn,
- progressive composition of semantic constituents through dialogue turns to represent user intentions,
- use of redundant interpretation methods for establishing indices of reliability of semantic interpretation,
- progressive updating of interpretation knowledge with active and explanation based learning.

To reach its goals, the project needed a large number of examples of what callers usually say in the target applications. The project partners collected and transcribed conversations from users engaged in complex tasks (e.g. from call routing with utterance classification to dialogue systems with complex semantic domains) in French, Italian and Polish. The data collection was based both on human-human telephone conversations as well as human-machine interactions.

Currently LUNA's research results are being validated on different application scenarios, targeted on dialogue-based telephone services of different complexity. The SLU models are being trained and applied to different multilingual spoken dialog systems (SDS) in the three project’s languages. The language-independent components have been shared among the participants and then adapted to each particular language by means of language resources already available or collected within the project.

2.2 Scientific challenges

Natural language interpretation by computers is the process of obtaining semantic representations in a meaning representation language (MRL) from signs coded into texts or speech signals. This process can be based on theories (e.g. Jackendoff, 1990) that build components of a semantic structure which represents the interpretation of an entire sentence, a dialogue turn or a text paragraph.
Detecting components from signs coded into the speech signal is particularly difficult because spoken language is often ungrammatical and may contain repetitions. Furthermore, using an automatic speech recognition (ASR) system introduces additional difficulties due to the imperfect behaviour of available systems.

Various attempts have been proposed for obtaining predicate-argument semantic representations from speech using robust, possibly partial, parsers (Woods, 1976, Kasper 1990, Seneff 1992), finite-state representations of languages expressing concepts (Pieraccini, 1991), classifiers (Shapire, 2005) and their combination with parsers (Moschitti, 2007). Complex sentences, like the ones that can be found in negotiation dialogues may convey meaning which has to be represented by structures obtained with composition of predicate-argument representations. Such a composition uses context-sensitive knowledge for generating instances of computational objects called frames.

A frame is a model for representing semantic entities and their properties (Fillmore, 1968, Backer, 1998). Properties are represented by slots in the frame structure. Slots are useful for representing associations between names of aspects of an entity and its descriptions, also called slot fillers.
Finding values for properties corresponds in the model to generate a frame instance and to fill its slots. Fillers are constrained to respect certain given types.
These constraints are stored into facets associated with a slot. The slot filler of a frame instance can be the instance of another frame in a structure which can be quite complex.

In (Brachman, 1975) it is suggested that frame structures conceived for representing semantic knowledge should be integrated with structural descriptions to represent global constraints among the components of a structure. Frame structures and structural descriptions contain n-ary relations which are supported by relations between signs often represented by syntactic relations. In many cases, semantic n-ary relations can be hypothesized by composition of local syntactic relations supported by specific patterns containing words and other syntactic features. An attempt to build interpretations by composition of local relations is appealing because, in practice, better results are often obtained in learning from examples to perform local decisions as opposed to global decisions (Punyakanok et al., 2004).

Instances of semantic structures can be affected by a degree of imprecision taken into account by probabilistic relational models as described in (Friedman et al., 1999).

A large number of SLU systems and in particular the ones developed for Air Travel Information Systems (e.g. Pieraccini, 1991), generate frame hypotheses for representing user intentions to derive system actions like data-base queries. For this purpose, there is no need to compose structures of frame instances since the instance of a single frame is sufficient for obtaining a query. Complex tasks, like problem solving involving user-system negotiation and automatic logic inference, may require operations which are not limited to the preparation of a frame instance for interacting with an information system.

The first scientific challenge of LUNA is to compose and maintain semantic structures which are not limited to a single frame instance. Structures are obtained from constituents hypothesized by the analysis of a dialogue turn or by logic inference. It is performed on partially hypothesized structures with supports obtained by shallow parsing on sequences of words in a lattice of hypotheses generated by an ASR system.

The second challenge is to use confidence indices and semantic constraints for composing instances of frame structures with supports which are not limited to the best sequence of word hypotheses generated by an ASR system. Possible inferences are used for driving the search for compositions in a structured list of component hypotheses (Raymond, 2007). Uncertainty is evaluated using probabilistic relational models capturing dependences among components of logical systems (see Taskar at al., 2002 for a discussion on this issue).

When dialogue turns are long and complex, human annotation of a single dialogue may take a considerable amount of time. For this reason it is useful to develop semi-automatic annotation procedures with suitable strategies for selecting specific cases for human inspection and possible correction. Examples that are found similar to previously validated ones can be automatically validated.

Corrections made by human inspections can be generalized by humans and the results made persistent in the knowledge base used for this purpose. Further generalizations can be obtained by automatic learning when enough data of certain types become available. Correct annotation in terms of domain independent speech acts is useful for this purpose. An annotation scheme proposed in (Allen 1997) has been adopted.

Support patterns for semantic interpretation can be obtained by conceiving computational models of explicit prior knowledge or by automatic learning from annotated data. Conditional Random Fields (CRF) (Raymond and Riccardi, 2007), finite-state machines (FSM) (Raymond et al., 2007), and dynamic Bayesian networks (DBN) (Meurs et al., 2008) have been used for this purpose. There are practical limits on automatic learning methods when the available data for model training are insufficient (Mitchell, 1997, p. 308). When automatic learning cannot be performed, supports for semantic relations can be obtained by parsing spoken sentences and mapping features of parse trees into semantic hypotheses. Human experts parse difficult sentence types, explain their relations with meaning and conceive support patterns that generalize their findings.

This type of generalization is also performed using features obtained by automatic learning. A support pattern obtained in this way may span an entire turn and has a head word used for indexing it. These supports are cached into a memory and used as triggers that signal the presence of sentence or phrase types and their associated interpretation. This form of Explanation Based Learning (EBL) makes it possible to combine prior knowledge with what is obtained by automatic learning.

A third challenge of LUNA is semi-automatic progressive semantic annotation of transcribed speech utterances. A type of EBL can be performed if a corpus is annotated in terms of semantic constituents and supports for constituents are available. In this case, local syntactic relations supporting semantic relations can be obtained by humans who consider syntactic compositions of constituent supports into meaningful sentences. When these patterns do not apply, fragments of semantic structures are hypothesized in a dialogue turn without being composed. This fact becomes a precondition for submitting an example to human analysis for annotation or for performing on-line parsing during automatic interpretation.

Patterns of syntactic relations may have specific words like prepositions which are essential for supporting semantic relations. These words are often short and likely to be deleted in the one-best sentence hypothesis generated by an ASR system. In such a case a patter matching may fail just for one short word error. This can be avoided by looking for a possible matching in a lattice of word hypotheses in which an insertion of a pattern is expected to be less likely that the deletion of a short word in the most likely sequence.

An n-ary semantic relation exists in general between a frame instance and its slots. Such a relation can be seen as a collection of binary relations between the frame and each slot. Furthermore, many semantic relations are supported by patterns expressing local syntactic relations between the supports of the relation constituents. A discussion on the advantage of obtaining n-ary relations from binary ones can be found in (McDonald, 2005 and Punyakanok, 2004).

2.3 Key innovations

The LUNA project is set to be at forefront of the third generation of spoken language interfaces for the reasons explained below.
The project collected data from “real” users engaged in “complex” tasks, as opposed to previous research projects. This was achieved through an important contribution from industry and public-sector companies. Data were collected from call center users performing tasks of different complexity, ranging from tasks such as information seeking dialog (e.g. Warsaw bus transportation domain) to Help-Desk for SW and HW repair (e.g. CSI call center).

The research program is investigating algorithms for SLU to solve the tasks and adapt in the context of a dialog system. Furthermore, the research project is multilingual and SLU investigations are being validated in three different languages (French, Italian and Polish) and domains.

A large number of dialogue turns has already been semantically annotated for applications in three languages using a unified annotation protocol and semi-automatic procedures based on active learning.

Methods based on conditional random fields, finite state transducers, kernel methods in stochastic parsers, dynamic Bayesian networks and statistical machine translation have been systematically applied and compared in the generation of semantic hypotheses from continuous telephone speech.

Telephone prototypes for the evaluation of semantic interpretations have been developed.

Finally, a novel system for spoken opinion analysis has been developed and tested.

2.4 Impact

A number of current telephone services allow users to interact with automatic systems by pushing buttons or answering questions with specific words or short sentences proposed by the systems. The SLU approach makes possible highly flexible prompts, avoiding prompting with a long list of specific options.
From a business perspective, the need for Spoken Language Understanding arises particularly when customers can get confused about their answers to basic prompts in a vocal interface.

The key to voice-application success is to create a user interface which allows interactions between the system and the users to be both pleasant and effective, thus overcoming the distrust users have of communicating with an automatic service instead of a human operator.
The vision of the LUNA project is to improve current automated telephone systems, allowing easy human-machine interactions through spontaneous and unconstrained speech, replacing menu-driven voice recognition. The project aims to enhance the users' experience, helping callers to use vocal services quickly and accurately.

A well-designed and well-deployed automatic system can yield increased revenues, lower costs, customer satisfaction and retention, brand development.
The introduction of SLU solutions can also increase the performance and the automation rate of the services currently in use, as well as reduce the development costs for new services.

Results in SLU are, furthermore, useful in other application domains such as indexing of spoken documents, automatic translation, and automatic opinion analysis for performing surveys.

LUNA project is going to provide solutions for natural speech recognition and understanding to be usable for industrial applications, thus raising the competitiveness of the industrial partners and strengthening the position of academic partners.

The industrial partners of the LUNA consortium are going to integrate the most promising results in their future commercial products and deploy large-scale applications; in addition, since one member of the consortium is a provider of IT services to local government, this could be the ideal vehicle for proposing new and advanced services to the public.
It is worth mentioning that other EU projects are going to at least partially reuse some of the results of LUNA; this confirms the success of the project, while opening up new possible scenarios for cooperation and evolution of the research.

3 Our approach towards SLU

3.1 Organization of the project

SLU is considered to be a problem solving activity, performed with the combination of processes at different levels depending on the type of application.
The most general architecture consists of three levels; in the LUNA project these levels have been implemented by three workpackages respectively. This workpackage structure allows main technological objective to be reached, that is to provide a SLU toolkit with components that can be combined within various configurations depending on the type of application it is dedicated to.

In summary, the project is organized as follows:

Workpackage 0 (WP0) is dedicated to project management.

Workpackage 1 (WP1) sized at the very beginning of the project the complexity of the application tasks to be used to research, develop and evaluate the SLU systems. It also addressed the definition of the architecture of the multilingual SLU system along with the communication protocols between the different components. In WP1 criteria and protocols for data collection and annotation were also formulated.

Workpackages 2, 3 and 4 are devoted to do research on the three main components of the SLU system.

WP2 is about the generation of semantic concept lattices (and of single concept sequences) from speech recognition output (word lattices). The main objective of WP2 is the study of open problems in SLU research regarding flat concepts and the implementation of the most promising results into an SLU module. In particular, research work will focus on modeling and training of stochastic models, on robustness w.r.t. speech recognition errors, on the exploitation of language models (beyond the classical n-grams), and on the coupling of speech recognition and spoken language understanding.

The main objective of WP3 is the study of open problems in semantic composition and validation and the implementation of the most promising results into an SLU module. Semantic composition knowledge is used for performing composition by a search process. The project is exploring semantic composition using knowledge representation expressed by structure-building rules which can be constructed using an initial application dependent semantic structure. Confidence measures and strategies are being investigated in order to propose to the Dialogue Manager structured interpretations of an utterance with reliability state information attached to each semantic object detected. In order to reduce the corpus collection effort needed to train the models, an active learning paradigm is being implemented that consists in selecting automatically the utterances considered relevant for updating the models and which are going to be manually labelled. Finally, cross-application and cross-language portability is a major preoccupation in this workpackage.

The main objective of WP4 is to make use of contextual information from other components of the dialogue system (mainly the Dialogue Manager) in order to rescore semantic structure hypotheses (output by WP3 component) or concept tag hypotheses (output by WP2 component) and to study to what extent the const-sensitive approach is portable to a new language in order to provide efficient cross-language bootstrapping. A first issue was to explore to what level of precision is it relevant to extract the context in order to improve SLU performance and to express the context within the semantic framework defined in WP1. General information about the way the system is used has been modelled and specific information about the current dialogue has been taken into account through an on-line learning scheme. On-line information reflects both the current user's behaviour and an automatic diagnostic of the SLU performance.

Workpackage 5 deals with the problem of corpora acquisition and annotation following an annotation protocol based on Knowledge Representation at various levels. Prototypes for new services have been developed and tested integrating modules of the toolkit with existing ASR and dialogue systems; they will evolve during the project life including the output results of the other WPs.

Workpackage 6 involves dissemination, exploitation and standardization.

3.2 The LUNA SLU system architecture

The architecture of the LUNA SLU system can be seen in Figure 1. On the one side, the ASR exports a large number of recognition hypotheses through the Word Lattice that becomes part of the input for the SLU module. On the other side the contextual information provided by the Dialogue Manager (DM), like information about the current dialogue act, or information retrieved from the dialogue history, complements the input. The resulting enriched lattice is known as the LUNA Lattice.

For Lattice representation the HTK Standard Lattice Format, which is flexible enough to support additions to the contextual information, has been adopted. The format used for the contextual information is XML.

The SLU output is the interpretation lattice. This lattice is also compliant with the HTK Standard Lattice Format. The semantically structured interpretations, represented through an XML string, are linked to the Lattice's transitions.

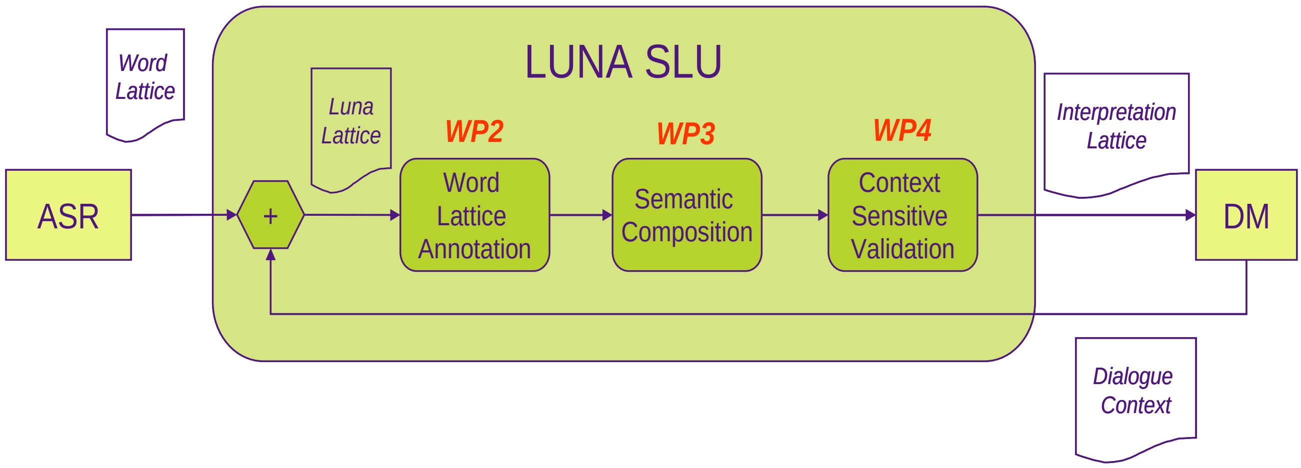

Figure 1 – architecture of the LUNA SLU system

From the architectural point of view, LUNA's SLU module can be broken down into three core sub-modules: the Word Lattice Annotation (WLA) module, the Semantic Composition (SC) module and the Context Sensitive Validation (CSV) module.

The target for the WLA module is the generation of semantic concept lattices using the speech recognition output generated by the ASR module (in the form of word lattices). Using the given word sequence, a stochastical model is used to segment this sequence into basic semantic concepts. These smallest units of meaning are usually associated with a certain concept value, which has to be derived from the speech input. Confidence scores provided by the ASR module may be incorporated to improve the robustness of the module. The input word lattice is augmented with the concepts and corresponding values and passed on the SC module.

The SC module is responsible for building rich semantic representations of utterances based on the simple/basic concepts produced by the WLA module and contextual information provided by the DM. These semantic interpretations can be scored according to a combination of several confidence scores, some of them related to the ASR models, and some specific to the semantic models. A decision strategy can then be applied, based on these confidence scores, in order to re-score and discard one or several interpretations.
An automatic selection process of problematic utterances, used as an Active Learning strategy, enriches the SC component

The main objective of the CSV module is, firstly, to characterize which contextual information, from the dialogue system, is relevant for SLU and, secondly, to exploit this contextual knowledge through a re-scoring process over semantic structure hypotheses.

Since the portability issue, both to a new domain and to a new language, is an important aspect of the LUNA project, this issue is addressed for each sub-module.

3.3 Key results

3.3.1 Corpora

Both human-human and human-machine spoken dialogues were collected and semantically annotated from the different application domains, in the three languages.

The application domains were the following: in Italian an information technology help desk, provided by CSI Piemonte, was used, in Polish data were collected from public transport services under the city of Warsaw Transportation Authority.
For the French language, France Telecom 1014 (customer care) and 3000 (customer self-care), as well as a stock exchange service, were used. Additionally, an existing dialogue corpus already collected (MEDIA) was annotated in terms of semantic structures made of previously annotated semantic constituents.

French corpora
The first corpus in French was collected from the '3000 application', which is a deployed self-care application for France Telecom customers. The 3000 service, the first deployed vocal service at France Telecom exploiting natural language technologies, was launched and made available to the general public in October 2005. 3000 is France Telecom's self care service that enables customers to obtain information about and purchase almost 30 different services, check their consumption, pay their bills and access the management of their services such as call forwarding or voice messaging.

The second corpus was collected from the '1014 application', which is a call-routing application for the France Telecom commercial services. It has been tried out on France Telecom residential customers.
1014 is the calling number for France Telecom commercial services concerning residential customers. A customer dials 1014 to get information, set up a telephone line, modify or cancel a France Telecom service.
This interactive service processes the residential customer requests and routes the calls according to a pre-established classification:

- service requests: directed towards different treatment groups of the 1014 agencies (subscription, service management, building of customer loyalty, …),
- payment problems: directed towards a billing service,
- Internet dysfunction or line connection problems: directed towards the France Telecom customer support calling center (1013),
- credit card payments: directed towards the France Telecom self-care calling number (3000).

Both the 3000 and 1014 corpora were collected from real users before the beginning of the project.

The third corpus was collected from the trial Stock Exchange application with users that were simulating automatic transactions.
This application has mainly four tasks which are processed within a single model with a mixed-initiative dialogue system.

- Retrieve financial and stock exchange information
  - First, second and new French market (~ 1000 shares)
  - ~ 15 European market indicators and ~30 sectors (telecom, automotive…)
  - ~ 15 infos per share
- Access to portfolio
  - 2 types of portfolios.
  - Max 4 portfolios per user.
- Buy and sell shares

  - 2 types of order
- Consult the one-month orders book

The MEDIA dialogue corpus had been previously collected and annotated through the French Technolangue/Evalda project (2002-2006) and is available through ELRA. Actions were carried out in order to adapt the existing manual annotations of the MEDIA corpus into the LUNA formalism.
The task defined in MEDIA is the reservation of hotel rooms with tourist information, using information obtained from a web based database. The MEDIA dialog corpus was recorded using a WOZ (Wizard-of-Oz) system simulating a vocal tourist information phone server. In this way, each user/caller believes she or he is talking to a machine whereas she or he is talking to a human being (a wizard) who simulates the behaviour of a tourist information server. Eight scenario categories were defined each with a different level of complexity. 1250 dialogs were recorded, from 250 different speakers where each caller was involved in 5 different hotel reservation scenarios. Dialogs may start with focus in different topics, i.e. choice of town, itinerary, tourist event, festival, price, date etc.

Italian Corpora
The Italian application domain for corpus data collection and annotation is the CSI's Help Desk. It provides customer care and technical support to institutional customers (Piedmont Region, Province of Turin, Municipality of Turin, Local Health Unit and the Consortium itself) through a unique telephone number. The usual problems are about infrastructure (for hardware problems in the workspace) and applications (for issues regarding software services provided by CSI) used by institutional customers' employees.
Both human-human and human-machine (WOZ) dialogues were collected.
About sixty volunteers were involved for gathering seven hundred Human-Machine dialogues.

Polish corpora
To collect Polish spoken dialogs an automatic recording system was installed at the Warsaw City Transportation call center. During three months about 15 000 dialogs were recorded. Such a huge database will be used of course not only within the framework of the LUNA project (only a very limited part was applied), but also in other research project referred to Polish spontaneous speech transmitted over telephone lines. For LUNA project 500 human-human dialogs were selected and divided into 5 domains, following the domains of conversations (info-requests): timetable information (when), transport lines routes (does it go to, through), transportation paths (how to get to), lines stops (where is it) and information on free and reduced fares. All topic groups are divided further into two subgroups – female and male callers and there are in the database almost equal in number. All dialogs are transcribed and annotated.

The second part the Polish database is composed of 500 human-human dialogs and from more than 500 human – machine gathered under Wizard-of-Oz procedure, the main part was recorded at the Warsaw Transportation call center. The other part of the database referred to the domain of fares was recorded using the same system simulating a vocal information server on fares reductions in city transportation. For this domain more than 100 dialogs were recorded, from 14 speakers.

The architecture of the WOZ system is presented on Figure 2

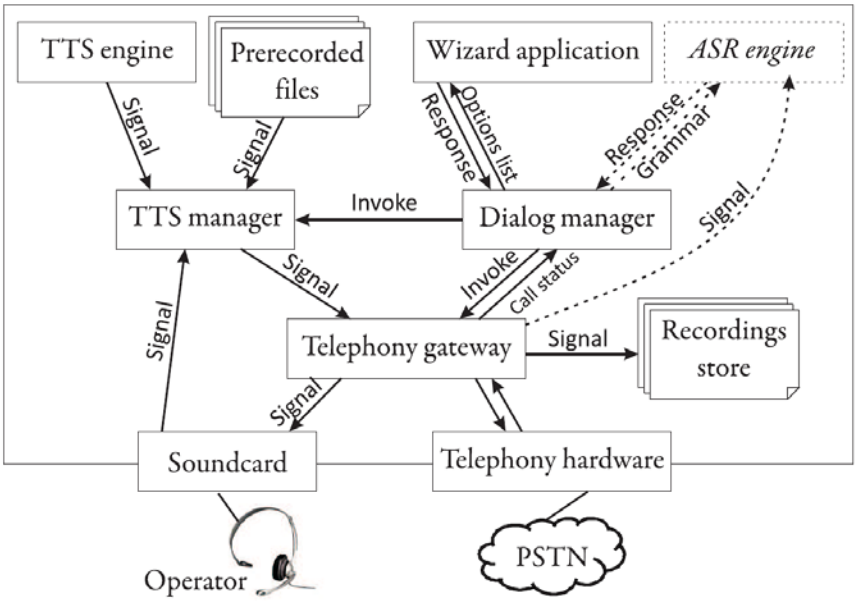

Figure 2 – architecture of the system for Polish WOZ data collection

The collected database is the first one made for Polish telephone spontaneous speech recorded at real information call center.

3.3.2 Telephone Spoken Dialog System Prototypes

The project has realized by now a first version of interactive prototypes of automatic telephone services using multilingual technology, as well as a set of tools which can help understanding the complex task of spoken language recognition and understanding.

The goal of the prototypes developed within the LUNA project is firstly to demonstrate the efficiency of the LUNA framework for deploying state-of-the-art Spoken Dialog Systems and secondly to illustrate by graphical interfaces and interactive demonstrations the scientific issues tackled by this project.
Below there is a description of the main functionalities of the prototypes available at the end of the second year of the project.

French language

The French prototype is about a negotiation dialogue by telephone. It is developed within the LUNA project by France Télécom and the University of Avignon, based on the publicly available French corpus called MEDIA. This prototype illustrates how a new SDS can be directly developed from a corpus annotated thanks to the LUNA annotation scheme.

The task defined in MEDIA is the negotiation for performing transactions and seeking information in a tourism application.

The goal of the French LUNA baseline prototype is to reproduce the behaviour of the MEDIA WOZ for the hotel booking task, including the negotiation phase of the dialogues (for example a hotel can be full or not all the facilities required by a user are available). This prototype includes most of the LUNA components and has been developed using a combination of rule-based and corpus-trained models. All the semantic annotations going beyond the constituent level already existing in the MEDIA corpus have been added with the LUNA active learning strategy.

Different methods have been applied and tested for generating meaning representations from ASR output. They include Finite State Transducers, Conditional Random Fields, shallow parsers, Dynamic Bayesian Networks, Statistical Machine Translation from natural language into a meaning representation language. Combination of these methods improves system robustness while comparison of the outputs obtained with the various methods provide confidence indicators to be used, among other things in deciding when data have to be proposed to a human expert for validation or annotation.

User-system negotiation also requires to maintain a knowledge base of hypotheses about user intentions. The content of this base is modified as a dialogue progresses by a semantic composition process which evaluates and modify hypotheses about semantic structures.

Italian language

The application context of the Italian prototype is a Help-Desk for Software/Hardware repair, relying on the human-human dialogs collected in the CSI call centre.

There are 10 possible dialog scenarios; the goal of the application is to understand in which scenario the user problem fits in and to route the call to an operator

Scenarios
1. Printer problem
2. PC Network problem
3. Slow computer
4. Screen/Video card problem
5. Keyboard problem
6. Mouse problem
7. Office network problem
8. Virus issues
9. CD-ROM reader problem
10. PC power problem

Figure 3 – scenarios of the Italian prototype

The first prototype of the SDS for Italian provides a call routing (problem identification) functionality in the context of such domain. The system reacts to user's spontaneous spoken requests belonging to one of ten possible dialog scenarios inspired from the services provided by CSI (see Figure 3). In addition to the human-human dialog data, we use such scenarios for the acquisition of Wizard-of-Oz Human-Machine dialogs. The goal of the first prototype is to understand the problem category the user request fits in and to route the call to an operator able to solve the user problem.

Polish language

The human-human dialogs collected from the Warsaw City Transportation call-center are related to problem of information distribution about mass transportation network

The first prototype of the SDS for Polish provides a call routing (requested information type identification) in the context of such domain.

The system reacts to user's spoken requests belonging to one of five possible dialog domains (lines routes (does it go through)), itinerary between given points of the city (how to get to), schedules, stops, reduced and free-fares.

The first version of the baseline prototype uses the acoustic and language models prepared by PJIIT.

It must be underlined that due to many spontaneous speech effects this is a particularly difficult task for the recognizer. This was the first experiment on ASR of Polish spontaneous speech.

3.3.3 Interactive Graphical Spoken Language Understanding prototypes

Two PC-based prototypes are available as well, presenting the different components of a Spoken Language Understanding system with interactive graphical interfaces. These graphical interactive demonstrations can present, to a non-specialist audience, the basic principles of Spoken Dialog Systems and, more specifically, the scientific challenges targeted by the LUNA project which could lead to the development of the next-generation SLU systems.

– The SLU "glass-box": this interface illustrates the process of understanding speech by displaying in a graphical interface the data-flow, from the speech signal to the meaning of a spoken utterance. The goal of this demonstration is to clearly present to a non-specialist audience the basic principles of SLU, and more specifically where are the current scientific challenges that need to be addressed in order to build the next-generation SDS (Figure 4).

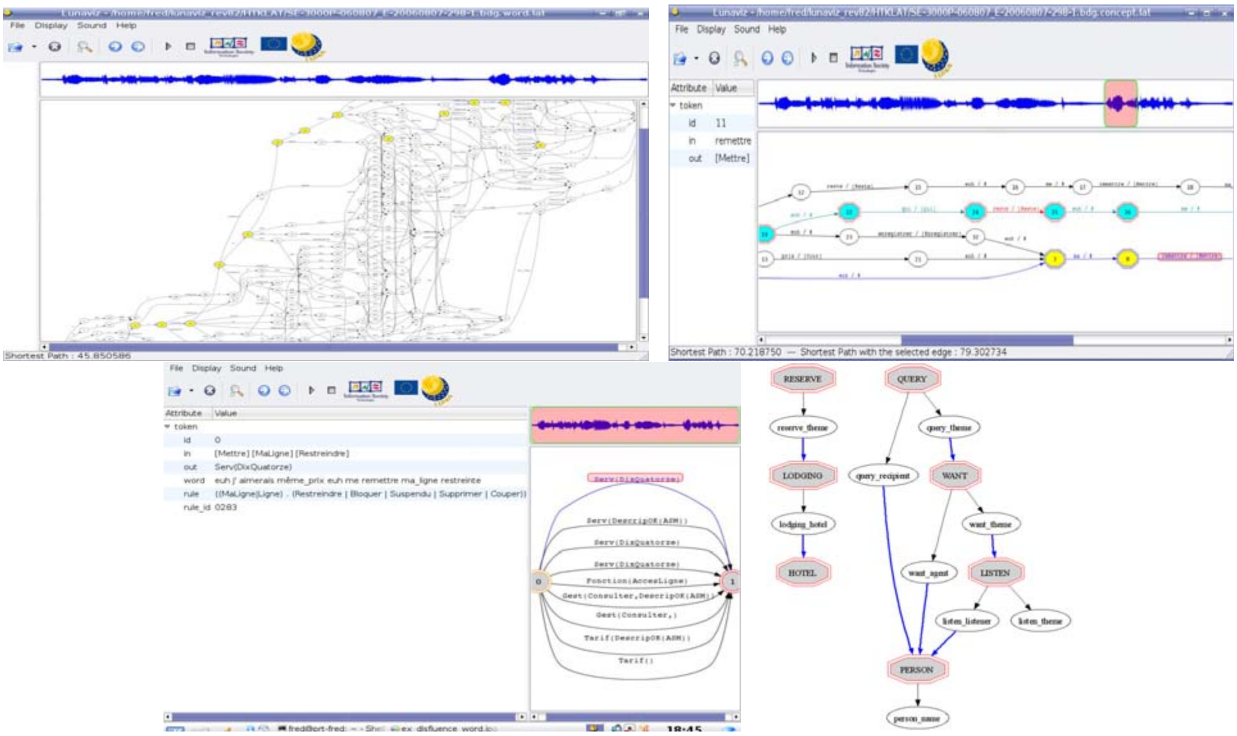

Figure 4 – SLU "glass box"

– LUNA Talk: this prototype proposes a talking-head interacting with the visitors and illustrating the current research in the LUNA project. More specifically the goal of this demo is to make accessible to non-specialist the current limits of state-of-the-art SDS, and how the strategies proposed in LUNA can push forward these limits.

Figure 5 – LUNA Talk

4 Glossary

ASR: Automatic Speech Recognition
CSV: Context Sensitive Validation
DM: Dialogue Manager
SC: Semantic Composition
SDS: Spoken Dialogue System
SLU: Spoken Language Understanding
WLA: Word Lattice Annotation
WOZ: Wizard-of-Oz

Annex – References

J. Allen and M. Core. Draft of DAMSL: Dialog act markup in several layers. Coding scheme developed by the MultiParty group, 1st Discourse Tagging Workshop, University of Pennsylvania, March 1996, 1997.

C. F. Baker (1998). Seeing clearly: Frame Semantic, Psycholinguistic, and Cross-linguistic Approaches to the Semantics of the English Verb See. PhD thesis, University of California at Berkeley

R. J. Brachman (1985). On the epistemological status of semantic networks, in R. J. Brachman and H. J. Levesque Eds, Readings in Knowledge Representation, pp. 191-216, Morgan Kaufmann, San Mateo, CA, 1985.

C. J. Fillmore (1968). The case for case. in E. Bach and R. Harms eds. Universals in linguistic theory, Holt, Rinehart and Winston, New York, 1968.

N. Friedman , L. Getoor , D. Koller and A. Pfeffer (1999). Learning Probabilistic Relational Models Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, p.1300-1309, July

R. Jackendoff (1990). Semantic Structures. The MIT Press, Cambridge Mass

Kasper R.T. and Hovy E.H. (1990). Performing integrated syntactic and semantic parsing using classification. Proc. Speech and Natural language Workshop, 54-59, Hidden Valley, PA, Morgan Kaufmann publ, Los Altos, CA.

R. McDonald, F. Pereira, S. Kulick, S. Winters, Y. Jin, P. White (2005). Simple Algorithms for Complex Relation Extraction with Applications to Biomedical IE. Proc. ACL, Ann Arbor, Mi, June 2005, pp. 491-498.

M.J. Meurs, F. Lefèvre, R. De Mori, A Bayesian approach to semantic composition for spoken language interpretation. Interspeech 2008, Brisbane, AUS.

T.M. Mitchell (1997). Machine learning. Mc Graw Hill, New York

A. Moschitti, G. Riccardi and C. Raymond, (2007). Spoken Language Understanding with Kernels for Syntactic/Semantic Structure. Proc. IEEE ASRU Workshop, Kyoto.

Pieraccini R., Levin E., and Lee C.-H. (1991). Stochastic Representation of Conceptual Structure in the ATIS Task. Proceedings of the, 1991 Speech and Natural Language Workshop, 121-124, Morgan Kaufmann publ, Los Altos, CA.

V. Punyakanok, D. Roth, W. Yih, and D. Zimak. 2004. Learning via inference over structurally constrained output. InWorkshop on Learning Structured with Output, NIPS

C. Raymond, F. Béchet, N. Camelin, R. De Mori and G. Damnati (2007). Sequential decision strategies for machine interpretation of speech. IEEE Trans. on Speech and Audio Processing, 15(1):162-171. 2007.

C. Raymond and G. Riccardi. Generative and Discriminative Algorithms for Spoken Language Understanding. Interspeech, Antwerp, Belgium, August 2007

R. E. Schapire, M. Rochery, M. Rahim, and N. Gupta (2005). Boosting With Prior Knowledge for Call Classification. IEEE Transactions on Speech and Audio Processing,SAP-13 (2) : 174-182

Seneff S. (1992). Robust Parsing for Spoken Language Systems. Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing, 1 : 189-192, San Francisco, CA.

K. Toutanova, A. Haghighi, C. D. Manning (2005). Joint Learning Improves Semantic Role Labeling. Proceedings of the meeting of the ACL, Ann Arbor, Mi, June 2005.

Woods, W. A. et al (1976). Speech Understanding Systems. Bolt, Beranek and Newman Inc., Cambridge, Ma., Final Report, Vol. IV, V