# Design and Data Collection for Spoken Polish Dialogs Database

**Krzysztof Marasek, Ryszard Gubrynowicz**
Department of Multimedia
Polish-Japanese Institute of Information Technology
Koszykowa st., 86, 02-008 Warsaw, Poland
Krzysztof.Marasek@pjwstk.edu.pl, Ryszard.Gubrynowicz@pjwstk.edu.pl

## Abstract

Spoken corpora provide a critical resource for research, development and evaluation of spoken dialog systems. This paper describes the telephone spoken dialog corpus for Polish created by Polish-Japanese Institute of Information Technology team within the LUNA project (IST 033549). The main goal of this project is to create a robust natural spoken language understanding (SLU) toolkit, which can be used to improve the speech-enabled telecom services in multilingual context (Italian, French and Polish). The corpus has been collected at the call center of Warsaw Transport Authority, manually transcribed and richly annotated on acoustic, syntactic and semantic levels. The most frequent users' requests concern city traffic information (public transportation stops, routes, schedules, trip planning etc.). The collected database consists of two parts: 500 human-human dialogs of approx. 670 minutes long with a vocabulary of ca. 8000 words and 500 human-machine dialogs recorded via the use of Wizard-of-Oz paradigm. The syntactic and semantic annotation is carried out by another team (Mykowiecka et al.,2007). This database is the first one collected for spontaneous Polish speech recorded through telecommunication lines and will be used for development and evaluation of automatic speech recognition (ASR) and robust natural spoken language understanding (SLU) components.

## 1. Introduction

Speech and language understanding forms an essential component in the design of spoken dialog systems. The integration of automatic speech recognition and spoken language understanding technologies is crucial for development of user friendly dialog systems. The Polish-Japanese Institute of Information Technology (PJIIT) is involved in preparation of a database for advancing and evaluating research and development of such systems using multi-language approach. The work presented is conducted within the LUNA (IST 033549) project (*www.ist-luna.eu*), which started in September 2006 as a STREP project within the 6th framework of the EU. The main goal of this project is to create a robust natural spoken language understanding (SLU) toolkit which can be used to improve the speech-enabled telecom services in multilingual context (Italian, French and Polish). The project consortium involves academic (RWTH Aachen, University of Trento, University of Avignon, Institute of Computer Science Polish Academy of Science and PJIIT) as well industrial partners (CSI-Piemonte, France Telecom R&D and Loquendo acting as the project coordinator). One of the central goals of the LUNA is to develop language-independent SLU tools. The multilingual portability of SLU components will be designed along with the communication protocols.

## 2. Dialog domains

The main goal of corpus collection was to provide a valuable resource for research in speech recognition, spoken language understanding, dialog management, machine learning, and language generation, especially for preparing and testing of LUNA project outcomes. The whole database will be composed of two parts – human-human dialogs and human-machine. The "Wizard of Oz" (WoZ) approach (Fraser&Gilbert, 1991) for gathering simulated human-machine dialog data is now being applied. We decided to collect the data at real call center to provide realistic data and insight into spontaneous dialogs.

Following this, the corpus of dialogs was collected at the Warsaw Transport Authority information center known as ZTM Service Center. Its telephone number is posted on all city transportation line stops in the city and within the vehicles.

The domain of dialogs recorded at the city transport call center is defined by the quite complex structure of the Warsaw public transport (numerous lines of trams, buses, city trains and one underground line). The surface lines are named by numbers (trams 1-46, busses 100-809 or letter-number combinations for night and special lines, together 328 lines) and their stops are named according to the places, street crosses, local buildings and public facilities, usually supplemented by a number to distinguish several stops located close to a given point (e.g. Dworzec Centralny 05, *Central Railway Station 05*).

The user's requests are balanced around five main groups:

- information requests on the itinerary between given points in the city;
- timetable for a given stop and given line and the travel time from given stop to destination;
- information on line routes, type of bus, tram (e.g. wheelchairs allowed);
- information on stops (the nearest from a given point in the city, stops for a given line, transfer stops, etc.);
- information on fares, their reductions and fare-free transportation for specified groups of citizens (children, youth, seniors, disabled persons, etc.).

It should be stressed however, that many calls cannot be easy classified into one of the presented groups. It happens quite often, that a caller change the topic during a talk, asking for more and more information.

This center is staffed daily by two to four operators who typically handle  200-300 calls per day with an average call duration of ca. 2 minutes. During data collection period one of the project staff members conducted on-site observations of call center operators activities to obtain a better insight into the center work-flow, group dynamics and interactions between operators and persons requesting information related to city transport services. An important part of callers' requests for information or assistance are addressed from users being outdoor and contacting the call center using their mobile phones.  Then, the call center staff assists directly users while they are on streets looking for nearby public transportation stops, routes, schedules, planning trips, etc. Those people are expected for rapid information.

Operators answering the users requests access different information sources – some data is available in a structured form and is accessible via internet, while some remain in an unstructured form used only by operators (text memos, detailed city maps with transportation lines, etc.) and generally there is no automated support for operators. Information is given solely in Polish.

## 3. Collection of human-human spoken dialogs

During 3 month of data collection a corpus with over 15000 dialogs has been recorded.  The quality of the recorded signals vary a lot, especially for the calls from mobile phones. The signal is noisy, most of calls are made in adverse acoustic conditions (many calls are carried on public places, streets, in a municipal transport vehicles, etc). Out of noise, low quality microphones and sudden transmission breaks there are another negative effects observed for signals transmitted over GSM.

From the set of recorded conversations 500 dialogs covering  five main domains were chosen for further research (of 223 male and 277 female callers). Main criteria for data selection were: adequate signal quality, moderate dialog duration and quite clear topic. The statistics for 500 human-human dialogs is given in the table below:

| Type of interaction | Number of dialogs | Time min. | Number of turns | Number of words | Number of different words (vocabulary) |
|---|---|---|---|---|---|
| Human-human | 500 | 670 | 12788 | 78082 | 7768 |

Table 1: The statistics of the recorded human-human dialogs

Distribution of calls over their subjects and callers' gender can be summarized in following manner:

| Type of information request | Total number of recordings | Female users | Male users |
|---|---|---|---|
| Transportation routes | 93 | 53 | 40 |
| Itinerary | 140 | 78 | 62 |
| Schedule | 112 | 60 | 52 |
| Stops | 55 | 24 | 31 |
| Reduced and free-fares | 101 | 61 | 40 |

Table 2: Distribution of calls in human-human dialogs

A  technical problem we had to solve was that the call center does not accept any interference into their LAN and PBX structures. For these reasons we decided to apply an external device that allows for automatic and independent recording of phone calls on multiple analogue phone lines. Device applied (FonTel phone call recorder) supports 4 analog lines and was connected to a PC through audio card and RS232 control ports. The software detected the beginning and the end of each call, its direction and saved the recorded contents of a call to an uncompressed audio file (wav). The exact date, time and line name of the dialog recording was used as a file name.  One dialog was stored per audio file and additional information (call duration, which line, etc. was stored in a database).

Prior to the connection to the operator the caller was notified that the conversation will be recorded. If this was not accepted by the caller he could  hang up. Following the local privacy laws (no personal data can be stored), operators do not ask for personal data and the calling number identification (CNID/CLIP) is not stored .

## 4. Dialogs transcription

The next step of the database preparation concerns transcription of all chosen dialogs. The recorded data has been transcribed  on word level and all transcripts are based on a fundamental unit of recorded speech behavior, which is referred to as a "speaker turn", or simply "turn". For all turns time borders are set and their content has been manually transcribed. In every dialog only two speakers were present (caller and operator), so turns were labeled to identify each speaker uniquely (user and operator).  All the corpus was adapted and annotated to the LUNA annotation scheme (Rodriguez et al., 2007).

The Transcriber software (Barras et al., 2000) was used to transcribe the recorded speech signal including annotation of paralinguistic and noise events. It must be underlined that transcription of spontaneous speech is not an easy task, especially in this case. The speech of people requesting information by phone is very emotional, especially, when they are seeking help in circumstances in which they cannot find their route or they are waiting on stops for the bus/tram that has not arrived on time. In such cases their articulation is often distorted, their speech is ungrammatical with many syntactical and inflection errors. It should be stressed that Polish is highly inflective language and such errors can, in particular cases, modify or made the utterance devoid of sense. However, linguistic errors (especially grammatical ones) are not corrected in dialogs transliteration (in such instances, very often the operator asked to repeat the sentence or the word he had not understand) and only evident errors in word pronunciation are annotated.

Other problems are related to the transcription of words of foreign origins very often pronounced following Polish pronunciation rules. Some of them has the "Polonized" orthographic form and are included in Polish dictionary. The other foreign words preserved, however, their original orthography (proper names, especially, often used in names of buildings or streets) but are pronounced following completely or partly Polish rules of pronunciation (for example, foreign words pronounced with Polish inflective ending). In the first case they are transcribed without any indication, in the second one – their annotations is composed of two parts: one referred to this portion of the word pronounced following its original rules of pronunciation, the other – to the segment pronounced as the part of a Polish word. Similar problems are with acronyms which are pronounced as distinct words. The speakers can pronounce them with Polish inflection (more often) or not. But in the first case, the inflection can modify also the phonetic structure of the acronym. However, both form of acronyms are accepted and converted into text using capital letters for their cores and a sequence of small letters for inflective ending.

The annotation of the recorded acoustic speech signals includes filler sounds, which could not be converted into text in unambiguous manner as well other human articulatory noises like breath, laugh, cough etc., while all non human noises are annotated with the common tag [noise]. Here a distinction was made when a noise does not overlap speech, or when the noise overlaps the beginning or the end of a word and when the noise overlaps a larger segment. This annotation can be useful in the stage of automatic speech recognition training. A babbling noise, yet another type of noise annotated, was quite often present in the recordings. This kind of noise is particularly difficult to manage in speech recognition, especially, when SNR is relatively low.

The information coded by Transcriber is stored in a XML format that is usable as input for next steps of semantic analyses. The papers (Mykowiecka et al., 2007; Rodriguez et al., 2007), give details on semantic annotation of the presented database. This part of work for Polish data is done by the team at the Institute of Computer Sciences of the Polish Academy Sciences.

## 5. Collection of human-machine dialogs

The Wizard-of-Oz tool was applied to collect dialogs taking an advantage of existing services, domain specific knowledge and corpora. For obvious reasons, at this step of data collection the domain of recorded dialogs will not cover all possible fields of the city transportation information. For this aim we have chosen a well defined and strictly limited domain referred to fares reductions regulations. It must be underlined that we have observed different speech phenomena in case of human-machine from those in human – human communication. In the first case, although people requesting information use simpler language when talking to a machine (if they are conscious of that), and therefore we avoid in the future the need for automatically understanding unconstrained language. This also filters out the turns irrelevant to this domain, which occur occasionally in human-human interactions.

The Figure 1 presents the system architecture applied to data collection (Koržinek et al., 2008). A proprietary system has been designed. The flow of signals and information is indicated by corresponding arrows. The main part of the system are the dialog manager and TTS manager which are coordinating the signal flows through telephony gateway. The system was installed at the city transport call center and it was crucial not to disrupt its normal operation when the experiment was carried out. The system allows simulation of fully automatic speech dialog system with strict approach which takes into account the limitation of a target automatic system.

The applied prompt and response units have a finite state dialog including of rather small set of input phrases. The strict WOZ simulation was fully done for the information service referred to the domain of city transport fares reductions. For other domains, after few steps of preliminary information collection from the user by the WOZ system the operator switch it off and continues the dialog with the caller.

The system uses modified computer telephony platform planned for automated dialog system. The computer was connected to the telephone line at the call center operator' desk. If there was a call from outside the computer hook up the phone and a TTS voice informed the user that the automatic system can answer questions regarding fares, schedules, routes, lost items and complaints. An operator was listening to the users wishes and depending on his answers switched the dialog to the next step at which a new request is synthesized, simulating an action of an automated dialog system. Instead of ASR part (Fig.1) a menu driven application was used. Through mouse click on a simple wizard prompt and response panels the operator could choose the most appropriate answer or ask for additional information from a user. An operator can also ask a user for repetition.  At each dialog step the user can ask for human operator's help and operator can break the WOZ mode of operation switching to human-human conversation. All speech activities are recorded in a separate file for each call.

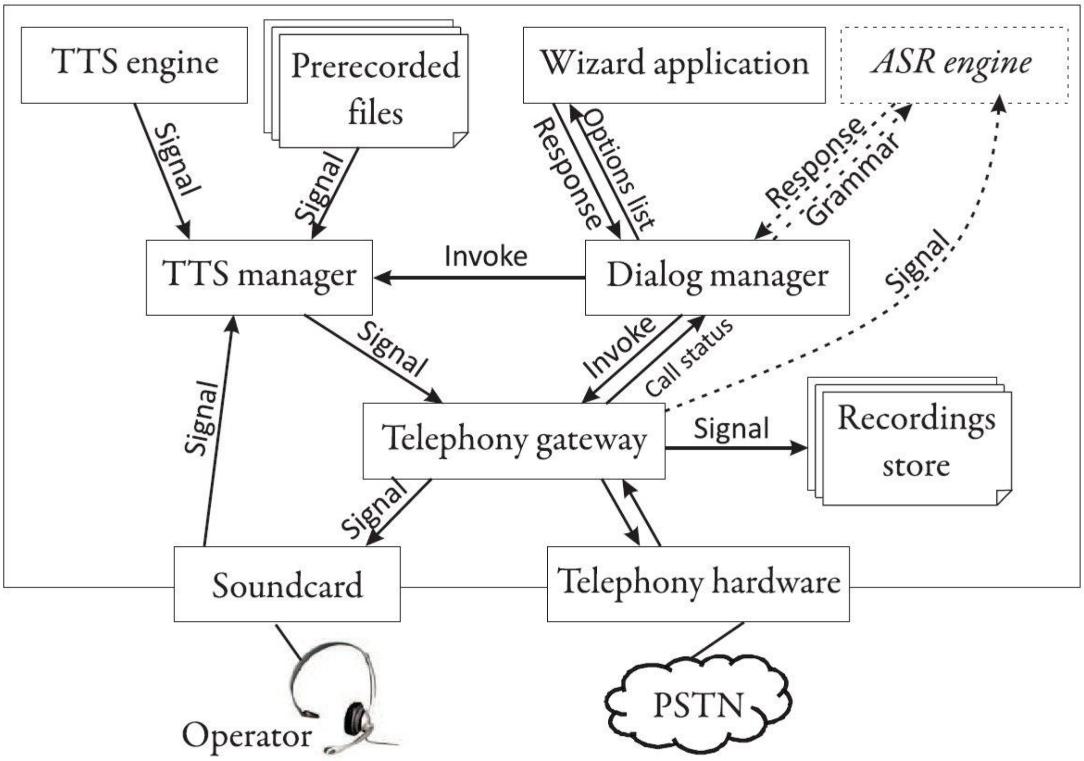

Fig. 1 The WOZ system architecture

## 6. Discussion

An important question is how the WOZ system should simulate accurately a real dialog system. The target system is a real SLU system (with an automatic speech engine) referred to city transport information service. If the WOZ should simulate an advanced SLU system, which has a flexible dialog management system, the problem of accurate simulation is a complex task, especially for Polish language. However, there are some benefits of adopted WOZ approach, between them it gives information how users would like to interact with a system, both in terms of dialog strategy and language. Although we used a partially strict approach, which is very constraining this strategy, the user is not aware of the system limitations initially and only gradually tends to restrict his/her input commands to get the system to work.

We observed  that the users, although they were advised at the very beginning they had to do with an automatic system, they often do not took it into account because they dialed the normal phone number of the call center. For this reason, the dialogs started hardly with many pauses, hesitation marks or WOZ repetitions. The mean durations of recorded dialogs is about 40-50% greater than in case of human-human communication, even though limited vocabulary and dialog language.

Yet another observation is a quite low acceptance of automated service. From 844 calls recorded, only 459 actually used the WOZ system. The rest wanted to either immediately speak to the operator, or waited silently for the operator to respond, ignoring the system altogether (Koržinek et al., 2008).

## 7. Conclusions

In current spoken dialog systems, the design of user interfaces strongly depends on tasks and domains. In LUNA project the main objective is to develop a general purpose spoken dialog system being language independent and of multi-domain application. Up to now, we have constructed for Polish an example of database by first collecting human-to-human and human-machine dialogs and converting speech to text.

This data are being applied by another team for morphological annotation of the transliterated dialogs, morphological tags disambiguation, chunks identification, concept definition and annotation (Mykowiecka et al., 2007; Rodriguez et al., 2007).

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.

## 8. Acknowledgements

Authors would like to thank the Warsaw Transportation Authority (ZTM Warszawa) and its call center employees for their support and cooperation.

This project is supported by EC LUNA project (IST033549).

## 9. References

Fraser, N.,Gilbert, N.S. (1991). Simulating speech systems. *Computer Speech and Language*, 5:81-99.

Koržinek D., Brocki Ł., Gubrynowicz R. and Marasek K. (2008). Wizard of Oz Experiment for a Telephony-Based City Transport Dialog System. In *Proceedings of the 16th Int. Conference Intelligent Information Systems,* Zakopane, Poland (in print)

Rodriguez K.J., Dipper S., Götze M., Poesio M, Riccardi G., Raymond C. and Wiśniewska J. (2007). Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus. In *Proc. of the Linguistic Annotation Workshop at the ACL'07 (LAW-07),* Prague, Czech Republic, 2007

C. Barras, E. Geoffrois, Z. Wu, and M. Liberman, Transcriber: development and use of a tool for assisting speech corpora production**,** *Speech Communication special issue on Speech Annotation and Corpus Tools,* Vol. 33, No 1-2, January 2000

Mykowiecka A., Marasek K., Marciniak M., Rabiega-Wiśniewska J. and Gubrynowicz R., Annotation of Polish spoken dialogs in LUNA project. *Proc. 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics,* Poznan, Poland, October 2007