

Dialogue Strategies for Improving the Usability of Telephone Human-Machine Communication

Morena Danieli, Elisabetta Gerbino, and Loreta M. Moisa

CSELT – Centro Studi e Laboratori Telecomunicazioni

Via G. Reiss-Romoli 274

I-10148, Torino, Italia

danieli,gerbino,loreta@cselt.it

Abstract

Interactions with spoken language systems may present breakdowns that are due to errors in the acoustic decoding of user utterances. Some of these errors have important consequences in reducing the naturalness of human-machine dialogues. In this paper we identify some typologies of recognition errors that cannot be recovered during the syntactico-semantic analysis, but that may be effectively approached at the dialogue level. We will describe how non-understanding and the effects of misrecognition are dealt with by Dialogos, a real-time spoken dialogue system that allows users to access a database of railway information by telephone. We will discuss the importance of supporting confirmation turns, and clarification and correction sub-dialogues. We will show the positive effects of robust dialogue management and dialogue state dependent language modeling, by taking into account both the recognition and understanding performance, and the success rate of dialogue transactions.

1 Introduction

During the last few years the recognition of spontaneous speech in telephone applications has greatly improved; nevertheless spoken dialogue between computers and inexperienced users still presents some problematic issues that reduce the user satisfaction in interacting with spoken language systems.

The occurrence of errors in the acoustic decoding of users' utterance is the potential cause of miscommunication in oral interaction with spoken language systems. Some of these errors have important consequences in reducing the naturalness of human-machine dialogues. Sometimes a robust approach in parsing and the use of language models during recognition are not sufficient to avoid recognition breakdowns. The recognition performance has a direct

impact on the requirements that the dialogue modules of spoken language systems have to meet. In order to increase the usability of the applications, dialogue management modules have to deal with partial or total breakdowns of the lower levels of analysis by preventing and detecting miscommunication sources.

In this paper we identify some typologies of recognition errors that cannot be recovered during the syntactico-semantic analysis, but that may be effectively approached at the dialogue level. Our analysis and the methodologies we describe have been tested in a task-oriented telephone application, but we deem that some considerations may also be useful for other display-less human-machine communication applications. We will describe how non-understanding and the effects of misrecognition are dealt with in Dialogos, a real-time spoken language system that allows users to access a database of railway information by using the telephone. A detailed description of the different modules of Dialogos may be found in (Albesano et al., 1997). In this paper we will discuss the importance of supporting confirmation turns, and clarification and correction sub-dialogues. The dialogue module of Dialogos makes an extensive use of context knowledge: contextual information is used not only for validating or rejecting semantic interpretations, but it is also sent to the lower levels of input analysis for helping the recognizer. We will show that the positive effects of robust dialogue management and dialogue state dependent language modeling may be evaluated by taking into account both the recognition and understanding performance, and the success rate of dialogue transactions.

From our experience we may conclude that if we provide robust behaviour in our dialogue systems, speech is a viable interface even with relatively low word accuracy rates. Nevertheless we believe that some important issues are still unexplored, and one of these is related to the weight that recognition errors have with respect to the degree of co-operativeness of the users. These open issues and some experimental data that emphasize their

urgency will be discussed in the section on experimental data.

2 Recognition errors and naturalness of dialogue

State-of-the-art systems that receive their input by high-quality microphone have word accuracy scores above 90%. However, the recognition of spontaneous speech in telephone environment is below that rate. Actually, the telephone input of the recognizer may greatly differ from the uttered acoustic signal, due to the noisy environment of the call, and to the quality of the telephone microphone and propagation network.

Most current task-oriented applications of telephone human-machine dialogue are developed for being used by a large population of potential users. These applications require speaker independent real-time systems, and the opportunity of having training sessions with the system cannot be provided. The speaker independent recognizers designed for such applications assure the coverage of a great number of speakers, but some aspects of the speech modality of some users can induce the recognizer to make mistakes, especially in recognizing long sentences.

The adverse recognition environment and the variability in user dependent features are the most frequent reasons of three kinds of recognition errors. The speech community usually classifies these errors into *deletions*, *insertions*, and *substitutions*. Some of them may be prevented by using language models during the recognition.

At present, some approaches to language modeling take advantage of contextual information sent by the dialogue model. However, the task of the dialogue state dependent language modeling is more difficult in some application domains. For example, some of the task-oriented systems that give information about railway timetable, or flight scheduling, have large vocabularies that contain an huge number of words belonging to the same class: for example, Dialogos vocabulary is 3,500 words, including 2,983 proper names of places; another example is the LIMSI Arise Railway Information system (Lamel et al., 1996) that has 1,500 words, including more than 680 station names. This peculiarity directly impacts on the performance of the language models, that is in these applications, language modeling predictions are weaker: when the dialogue prediction says that next user's utterance is likely to be about a departure place, this does not exclude that the recognizer substitutes the actually uttered name with a phonetically similar one. Only the user is able to detect such kinds of errors. In this situation the dialogue system should identify the user's detection of miscommunication and provide appropriate repairs.

All the problems described above lead to the decrease of the recognition performance and of the us-

ability of spoken language systems. More specifically, they identify some severe requirements that spoken dialogue modules have to meet. In particular, dialogue systems for telephone applications have to rely not only on an adequate model of the human user, but they should also implement particular techniques for preventing and recovering communication breakdowns.

3 Prevention and Repair of Miscommunication

3.1 Prevention of miscommunication

The user modeling of the dialogue module of Dialogos is based on the assumption that both the system and the user are active agents that cooperate in order to fulfill the goal of the speech interaction. In the application domain of Dialogos the goal of the interaction is to collect all the parameters that are necessary to access the railway database for retrieving the information that satisfy user's needs.

- T1-S: Hello, This is Train Enquiry Service.
Please speak after the tone.
Please state your departure and your destination.
- T2-U: Voglio partire da Mila(no)- Milano di sera.
Vado da Milano a Roma.
I would like to leave from Milano ... Milano in the evening.
I am going from Milano to Rome.
VOGLIO PARTIRE DA NO MILANO SERA
VADO DA MILANO ROMA
WANT LEAVE FROM NO MILANO EVENING
GO FROM MILANO TO ROMA
<confirm=NO, dep-city=MILANO,
<arr-city=ROMA, part-day=EVENING>
- T3-S: Do you want to go from Milano to Roma leaving in the evening?
- T4-U: Si ... da Milano a Roma di sera.
Yes ... from Milano to Roma in the evening.
SI BLOW DA MILANO ROMA SERA
YES BLOW FROM MILANO ROMA EVENING
<confirm=YES, dep-city=MILANO,
arr-city=ROMA, dep-time=EVENING>
- T5-S: There are many trains in the evening.
Which hour do you want to leave?
- T6-U: Mah ... mi dica se c'e' qualcosa alle otto.
Who knows ... Tell me if there is something at eight.
MATTINO GINOSA ALLE OTTO
MORNING GINOSA AT EIGHT
<part-day=MORNING, city=GINOSA,
hour=EIGHT>
- T7-S: Train 243 leaves from Milano Centrale at 8:20 p.m.;
it arrives at Roma Termini at 6 a.m.
Do you need additional information about this train?
- ...

Figure 1: Excerpt from the Dialogos corpus

As it was explained above, telephone recognition is error-prone. For preventing recognition errors the dialogue system sends to the lower levels of analysis information about the domain objects focused during each turn of the interaction; this information allows the triggering of context dependent language models that help to constraint the lexical choices at

the recognition level (see section 4). Moreover, in order to detect recognition or interpretation errors that occurred in previous turns, the dialogue system takes advantage from the global history of the interaction and it only accepts interpretations of user's input that are coherent with the dialogue history.

For example, let us consider the excerpt from the Dialogos corpus shown in Figure 1. In the example, on the left, the letter 'T' stands for 'Turn', the letters 'U' and 'S' stand for user and system, respectively. In each user's turn we reported in Italian the original user's utterance and its English translation (in italics). Then we have transcribed the best decoded sequence (in ALL CAPS), that is the recognizer output. The translations into English of the best decoded sequences are shown in ALL CAPS (in italics). The task-oriented semantic frames (the input of the dialogue module) have been put between angles. The system turns have been only reported in their English translation.

In T2-U the user utterance contains an hesitation when uttering the name of the departure city, "Milano". The first part of the word, "Mila-" was misrecognized as a noise, and the last syllable was recognized as "no", that the parser interpreted as the negation adverb "no". In this initial dialogue context there were no parameters to be denied, and the dialogue module was able to discard this information related to the negative adverb. It addressed the user with the request of confirmation of T3-S. T4-U was a confirmation turn of the user. After having consulted the data in the railway database, the system realized that the number of connections between Milano and Roma in the evening was high, and it suggested the user to choose a more precise departure time (T5-S). In T6-U the utterance segment "mah ... mi dica se c'e' qualcosa" (who knows ... tell me if there is something) was misrecognized as "mattino Ginosa" (morning Ginosa, where "Ginosa" is the name of an Italian village). As a consequence, the parser output contains another value for the part-of-day parameter and a departure hour. However, the dialogue discarded the information about the part-of-day, since this conflicted with a parameter value that the user had already confirmed, and only the second part of the utterance interpretation was retained (that is, the departure hour). In this case the insertion of a concept due to misrecognition was repaired at the dialogue level ¹.

As we can see from the example, the dialogue module makes use of confirmation turns because it deals with potentially incorrect information. However, the need for confirmations may result in a lack

¹This version of Dialogos only considers the first best solution. If the expected information is not found in the semantic representation of the current user's utterance, the dialogue system hypothesizes that something went wrong in the previous analysis, and it interprets that situation as an occurrence of non-understanding.

of naturalness of the telephone human-machine dialogues. In order to reduce the number of confirmation turns, we use the following strategies:

- the dialogue system avoids confirmation turns when the acquired information is coherent with the dialogue history and with the current focus
- the dialogue system asks for multiple confirmations of the acquired parameters (as in T3-S)
- the dialogue system asks for implicit confirmations whenever it is possible (as "From Milano to Roma. When do you want to travel?")

3.2 Repair of miscommunication

Sometimes the detection of some recognition errors (for example, the substitution of an uttered word with another one of the same class) is outside of the capabilities of both the parser and the dialogue modules. On the contrary, in principle the user is able to detect and correct such errors, and often she does it immediately or in subsequent turns. In (Danieli, 1996) an analysis of such phenomena is offered, and the method described in that paper is currently implemented in Dialogos. The approach is based on pragmatic-based expectations about the semantic content of the user utterance in the next turn. The theoretical background is the account of human-human conversation given by (Grice, 1967), re-interpreted in the context of human-computer conversation. In particular, the dialogue system is able to deal with *non-understandings* and *misunderstandings* (see (Hirst et al., 1994) for the classification of non-understanding, misunderstanding, and misconception), and it may recognize the occurrence of a miscommunication phenomena on the basis of the occurrence of two pragmatic counterparts, that is the deviation of the the user's behaviour from the system expectations, and the generation of a conversational implicature.

Non-understanding is recognized by the dialogue system as soon as it happens, because the system is not able to find any interpretation of the current user turn. On the contrary, misunderstandings are more difficult to detect and solve, because usually the dialogue system may get an interpretation of the user's utterance, but that interpretation is not the one intended by the speaker. If the user's correction of a misunderstanding occurs when the parameter is focused (that is, it occurs as a *third-turn repair*, see (Schegloff, 1992)), the focusing mechanism and the dialogue expectations allow to grasp the correction immediately. However, it is more problematic to detect user's corrections if they happen some turns after the occurrence of the errors. The dialogue system initially interprets user's correction with respect to its current set of expectations. As soon as it realizes that there is a deviation of the user behaviour from the expected behaviour, it hypothesizes a misunderstanding, and it re-interprets the current utterance

on the basis of the context of the misunderstood utterance (thanks to a focus-shifting mechanism).

Finally, the output of the parsing module may be only partially determined. In that case the dialogue module initiates clarification subdialogues. Let us discuss the excerpt shown in Figure 2.

T1-S: Hello, This is Train Enquiry Service.
Please speak after the tone.
Please state your departure and your destination.
T2-U: A Firenze.
To Firenze.
BLOW FIRENZE
BLOW FIRENZE
<city= FIRENZE>
T3-S: Are your leaving from Firenze or going to Firenze?
T4-U: Devo andare a Firenze.
I have to go to Firenze.
DEVO ANDARE A FIRENZE
MUST GO TO FIRENZE
<arr-city= FIRENZE>
T5-S: To Firenze. Where are you leaving from?
...

Figure 2: Example of miscommunication

In T2-U the arrival city is recognized and understood as a generic city, the dialogue strategy does not reject this information but it enters a clarification subdialogue in order to solve the ambiguity (T3-S and T4-U). The last turn of the system is a dialogue act that fulfill two communicative goals, that is the (implicit) confirmation of the arrival city and the request of the departure city. However, clarification subdialogues may be avoided if the dialogue expectations allow to choose an interpretation of ambiguous input.

Clarification subdialogues may also occur in case of parser outputs that contain inconsistent related information. Those may be so either because of recognition errors, or because of user's misconceptions. In our application domain misconceptions, i.e. errors in the prior knowledge of a participant, usually concern the expression of departure dates, as in the dialogue excerpt shown in Figure 3. The conversation took place on Thursday February, 27th 1997:

T1-S: What date do you travel?
T2-U: Voglio partire domenica ... il primo marzo.
I want to leave on Sunday ... March, 1st.
VOGLIO PARTIRE DOMENICA @ PRIMO MARZO
WANT LEAVE SUNDAY @ FIRST MARCH
< w-day= SUNDAY day= FIRST month= MARCH >
T3-S: Are your travelling on March, 1st or next Sunday?
...

Figure 3: Example of miscommunication due to misconception

The dialogue system recognizes the misconception in T2-U because the week day, the day, and the month are not interpretable with respect to its knowledge of the year's calendar (and the computer

presumes that its calendar is correct). Moreover the dialogue system finds that different chunks of the information supplied by the user could be coherently interpretable. Since it has no principled way to decide between them, it initiates the clarification subdialogue of T3-S.

4 Dialogue State Dependent Language Modeling

The dialogue module makes use of pragmatic-based expectations about the semantic content of the next user's utterance. In order to improve speech recognition performance the contextual knowledge may be used as a constraint for the language models. Contextual information is sent to the lower levels of analysis by communicating the dialogue act produced by the system for addressing the user. In Dialogos there are four classes of dialogue acts (request, confirmation, clarification, and request plus confirmation). Each class is further specialized with the indication of the focused parameters: for example "request:date of departure" if the system is asking the user to provide a departure date, "request-plus-confirmation: departure-plus-arrival", when the system is addressing the user with a feed-back about the departure city and a request of the arrival city, and so on. The information about the system dialog act is called "dialogue prediction". The recognizer makes use of the predictions by selecting a specific language model which was trained on a coherent partition of the training corpus (Popovici and Baggia, 1997).

The training set was collected in previous experiments of the system: users' responses to specific system dialog act were classified for training different language models. At present, the speech recognizer makes use of a class-based bigram model; then, in order to re-score the n-best decoded sequences, it uses a class-based trigram model.

In order to measure the improvements in recognition performance obtained using dialogue state dependent language models, we compared the differences in the Word Accuracy (WA) and Sentence Understanding (SU) rates obtained using different language models on the same test set. The test set included 2,040 utterances, randomly selected from corpus data collected during a field trial of Dialogos with 500 unexperienced subjects. In the first experiment, a single context-independent language model was used; it was trained on a set of 15,575 utterances (produced by Italian native speakers). In the second experiment, a set of dialogue state dependent models, trained on the same training set of the first experiment, was used; however in this case the training set was encoded according to the different dialogue states, as explained above. The error rate reduction between the two experiments is 8.6% of WA and 10.9% of SU. Moreover, with an improved *acoustic*

model (trained on a domain dependent training-set) the error reduction is even greater (over 12%) for both WA and SU.

5 Experimental Data

The Dialogos corpus consists of 1,404 dialogues, including 13,123 utterances. All the calls were performed over the public telephone network and in different environments (house, office, street, and car).

The WA and SU results on the global utterance corpus were: 61% for WA and 76% for SU. These results were greatly influenced by the quality of the telephone acoustic signal, and by the noise environment. Moreover, several city names contained in the dictionary of the system could be easily confused.

The overall system performance was measured with the Transaction Success (TS) metric, i.e. the measure of the success of the system in providing users with the information they require (Danieli and Gerbino, 1995). The TS rate was 70% on the 1,404 dialogues. By excluding from the corpus a set of dialogues that failed for users' errors, we obtained a TS result of 84%. The average successful dialogue duration is about 2 minutes: in most of the dialogues all the parameters were acquired and confirmed during the first minute of user-system interaction.

It is an open issue if a spoken dialogue system has to generate a clarification subdialogue when faced with ambiguity or unclear input. For example, the system described in (Allen et al., 1996) was designed on the basis of the principle that it was better to assume an interpretation and, of course, to be able to understand corrections when they arise. On the contrary, Dialogos was designed to enter clarification subdialogues when faced with input that cannot receive a single coherent interpretation in the dialogue context. Actually, we think that *in general* the strategy implemented by (Allen et al., 1996) may be more effective for the naturalness of the dialogue, however we believe that the effectiveness of that choice greatly depends on the ability of the users to grasp inconsistencies in the system feed-back. In the Dialogos corpus we observed that while subjects were usually able to correct errors in confirmation turns that concern a single information, or two semantically related information (such departure and arrival); on the contrary, some errors were not corrected when the feedback was offered together with a system initiative, or when the system asked to confirm information that had not strong relationships. The dialogue shown in Figure 4 is a typical example.

The acoustic decoding of "Allora" (a word that is used in Italian for taking turn) was erroneous: it was substituted with "All'una" (at one o'clock). This was interpreted as a departure hour. A conjunct confirmation of departure hour and arrival city was asked and the user confirmed both of them. In next section we will elaborate more on users' error.

```
T1-S: Hello, This is Train Enquiry Service.
      Please speak after the tone.
      Please state your departure and your destination.
T2-U: Allora ... vado a Milano.
      Then ... I am going to Milano.
      ALL'UNA VADO A MILANO
      AT ONE GO TO MILANO
      <hour=UNA, arr-city=MILANO>
T3-S: Do you want to go to Milano leaving at 13:00?
T4-U: Si.
      Yes.
      SI
      YES
      ...
```

Figure 4: Example of erroneous confirmation

For the sake of the present discussion, this example shows us that users are not always able to correct errors: on the contrary, we have seen above that the percentage of users' errors is high. In order to evaluate the effectiveness of the different approaches to face ambiguity we should experiment the different strategies on the same domain, or at least with the same interaction modality (phone or microphone). However, we have obtained some data that may give some insights on the issue. In the Dialogos corpus we have calculated the number of turns necessary for acquiring departure and arrival cities in the successful dialogues. While 64% of the users were able to give them in two turns (that is without experiencing recognition errors), the remaining 36% took from three to eight turns, i.e. these users' spent in correction from three to eight turns. Since the percentage of users that was not able to detect recognition errors is around 16%, we may hypothesize that a part of the subjects that experienced clarification subdialogues would have failed to give the correct values of the task parameters. Moreover, if we consider the cost of clarifications and repairs in terms of time, that is not awfully high: giving departure and arrival in less than three turns (that is, without clarifications or repair) takes from 20 to 29 seconds, while the entering of repair subdialogues increased this time of an average of 25 seconds on the total average time of the dialogues.

6 Discussion

By analyzing the Dialogos corpus, we identified some topics that require further work.

The first one is related to the need for goal management: in the task domain of Dialogos the goal is fixed during all the dialogue, but that is a simplification of the travel inquiry domain, introduced to control the complexity of interaction in order to meet the real-time requirement. However, we believe that the ability of the dialogue system to support goal management would greatly increase the naturalness of dialogue, as the work by (Allen et al., 1996) shows.

The second problematic issue is related to the impact that recognition errors have on the user be-

haviour. By examining the Dialogos corpus we collected evidence that some critical situations occur when the users make experience of repetitive recognition errors. The recognition errors seem to cause a cognitive overload in the users that influence their degree of co-operativeness. In our task domain most of the recognition errors occur during the recognition of departure city and arrival city. When users are asked to repeat a city name that was misrecognized by the system, some of them modify their way of speaking: they repeat the name louder, or spelt it, or even accept a misrecognized name proposed by the system. In this corpus 15.1% of the dialogues failed for users' errors. Similar behaviour has been described in (Bernsen, Dybkjaer, and Dybkjaer, 1996).

During the evaluation of the Dialogos corpus we classified the users' errors into two main classes:

1. the user confirmed a task parameter value that derived from a misrecognition (76% of the users' errors) after having experienced several recognition errors;
2. the user accepted a task parameter derived from a word that was inserted by the recognizer and interpreted by the parser (24%).

We applied the above classification in a different experimental set-up: in May 1996 a laboratory test was designed to study the reaction of users to different speech technologies and dialogue strategies applied to the railway information domain. 48 naive subjects took part in the experiment: they were equally distributed by sex; the age range was from 18 to 62 years. The subjects of this experiment called Dialogos and another spoken dialogue system. Each of them played two different predetermined scenarios with each system; after the trial, they were interviewed (the discussion of this experiment may be found in (Billi, Castagneri, and Danieli, 1997)). By analyzing the 96 dialogues between Dialogos and the experimental subjects, we found that the occurrence of users' errors could be classified into the two classes described above, that is users' errors were always concomitant with substitutions or insertions in the best-decoded sequence. In particular, in both the experiments subjects reaction to recognition errors was characterized by an alteration in the way of speaking. In the interviews collected in the second experiment, the subjects that made errors expressed the fatigue in experimenting repetitive recognition errors. However, we hypothesize that the non-cooperative behaviour may be partially due to the artificial experimental conditions: we are planning to experiment the current version of Dialogos in a real environment with users that really need to take trains for travelling all around Italy, and that will use the system for having timetable information.

While there are clearly many aspects in which our current approach requires further work, we may

claim that speech is a viable interface if we provide spoken systems with robust dialogue management. We have shown that the use of contextual information in different system modules may reduce the recognition errors, and increase the usability of telephone human-computer dialogue.

7 Acknowledgements

The work of the first and the second authors was partially supported by the LE3-4229 project, ARISE (Automatic Railway Information System for Europe), promoted by the Commissions of the European Communities, DG XIII (Telecommunications, Information Market and Exploitation of Research). The third author is a researcher of ICI (Institutul de Cercetari in Informatica, Bucarest, Romania): the work reported in this paper was implemented while she was visiting CSELT.

References

- Albesano, Dario, Paolo Baggia, Morena Danieli, Roberto Gemello, Elisabetta Gerbino, and Claudio Rullent. 1997. A Robust System for Human-Machine Dialogue in Telephony-Based Applications. In *International Journal of Speech Technology*, Kluwer Academic Publishers. To appear.
- Allen, James F., Bradford W. Miller, Eric K. Ringger, Teresa Sikorski. 1996. A Robust System for Natural Spoken Dialogue. In *Proceedings of the 34th annual Meeting of the ACL*, Santa Cruz, California. Association for Computational Linguistics, Morristown, New Jersey.
- Bernsen Niels O., Laila Dybkjaer, and Hans Dybkjaer. 1996. User Errors in Spoken Human-Machine Dialogue. In *Proceedings of the ECAI-96 Workshop on Spoken Language Systems*, Budapest, Hungary.
- Billi, Roberto, Giuseppe Castagneri, and Morena Danieli. 1997. Field trial evaluations of two different information inquiry systems. In *Speech Communications*. To appear.
- Danieli, Morena. 1996. On the Use of Expectations for Detecting and Repairing Human-Machine Miscommunications. In *Proceedings of AAAI-96 Workshop on Detecting, Preventing and Repairing Human-Machine Miscommunications*, Portland, OR., pages
- Danieli, Morena and Elisabetta Gerbino. 1995. Metrics for evaluating dialogue strategies in a spoken language system. In *Proceedings of the 1995 AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pages 34-39.

- Grice, Paul H. 1967. Logic and Conversation. In Cole, P., and Morgan, J. eds. 1975. *Syntax and Semantics*, New York and London: Academic Press.
- Hirst, Graeme, Susan McRoy, Peter Heeman, Philip Edmonds, Diane Horton. 1994. Repairing conversational misunderstandings and non-understandings. In *Speech Communication*, volume 15, pages 213-229.
- Lamel Lori, Jean Luc Gauvain, Samir K. Bennacef, Laurence Devillers, Saliha Foukia, Jean-Jacques Gangolf, and Sophie Rosset. 1996. Field Trials of a Telephone Service for Rail Travel Information. In *Proceedings of IVTTA-1996*, IEEE Third Workshop Interactive Voice Technology for Telecommunications Applications, Basking Ridge, New Jersey.
- Popovici, Cosmin and Paolo Baggia. 1997. Specialized Language Models Using Dialogue Predictions. In *Proceedings of ICCASP-97*, Munchen, Germany.
- Schegloff, Emanuel A. 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. In *American Journal of Sociology*, Volume 97, Number 5, pages 1295-1345