Group Interaction in the Cockpit: Some Linguistic Factors[1]

Manfred Krifka

Silka Martens

Florian Schwarz

For a number of years it has been recognized that the social dynamics of group interaction is an import factor in the origin of accidents and in the way how accidents or accident-prone situations are handled in aviation (cf. Helmreich 1997a, 1997b). Factors related to interpersonal communication have been implicated in up to 80% of all aviation accidents over the past 20 years. As a reaction to this, Crew Resource Management (CRM) has been developed with the goal of rating and improving crew performance in aviation and in other fields in which professional groups interact in situations of high taskload and potential risk (cf. Helmreich ea. 1999). As far as this can be estimated at all, installing CRM techniques in the major American and European airlines has resulted in a definite improvement in the safety of commercial aviation. In spite of this success of CRM, practitioners in the field feel that, beyond the general social dynamics of group interaction, there might be potential problems relating to language and communication in such settings.

In this article, we first summarize some aspects of previous research in this area. Then we report findings from a project that one of us, Manfred Krifka, has carried out, using transcripts of flight simulator sessions with pilots of a commercial American airline. We will discuss some of the problems of this project. Finally, we describe an ongoing continuation of that project that uses flight simulator sessions with pilots of a commercial German airline.

1 Previous Research

Communication in situations with high task load and its problems can be described from various angles. In particular, the following two perspectives come to mind: First, we can investigate properties of the language that serves as the medium of communication. Certain features of language may be detrimental to the goals of a group in situations of high task load, and they may be improved by carefully designing appropriate terminology and rules of language use. Secondly, we can investigate whether the social structure in the group leads to certain biases in the communication within the group that impede proper group interaction. Professionals then could be trained to recognize these biases and their danger to the flow of information within the group. In the following, we will describe two prominent examples of these types of research in the general area of aviation.

[1] Work on this project is supported by the Gottfried Daimler und Karl Benz Stiftung, in the context of the project *Group Interaction in High Risk Environments*. We thank the members of this project for extensive discussion, especially Brian Sexton for providing the data for the first part of the project, Dagmar Silberstein for help with discussing and analyzing the data, and Gerhard Fahnenbruck for the ingenious design of the simulator situation of the second project and for providing us with the data.

1.1 Problems of language

As for potential problems of language, investigations of accident reports in aviation have indeed uncovered that structural properties of language can lead to misunderstandings, especially in situations in which the cognitive resources of participants are limited because of high task load.

Cushing (1994) has collected a number of cases, from reports of the National Transportation Safety Board (NTSB) and especially from informal reports of the pilots newsletter *Callback*, in which properties of language and language use caused accidents or near-accidents in aviation. Typically, such cases arise because the cognitive apparatus of the pilots is preoccupied and overloaded with paying attention to problems such as equipment failure or adverse weather conditions; also, communication can be severely impeded in the noisy cockpit or in the radio talk with the air traffic control (ATC).

Cushing identifies several types of cases in which structural properties of language, and aviation language in particular, may lead to problems or actually have led to problems. A particularly important problem is ambiguity of language. Ambiguity was arguably involved in the most severe accident in commercial aviation, the collision of two airplanes at the Tenerife airport in 1977. In this event, the pilot of the starting KLM plane reported to the ATC *We are now at takeoff*, and got the reply *OK...Stand by for takeoff*, which was partly masked by noise. Clearly, the pilot meant to say 'we are now on the takeoff roll', or 'we are now taking off', whereas ATC understood it as 'we are waiting at the takeoff point'. There are a number of such word ambiguities or structural ambiguities in aviation language – Cushing mentions, for example, the verb *hold*, which means in aviation parlance 'stop what you are now doing', but in ordinary English also, 'continue what you are now doing', or the use of *PD* as 'pilot's discretion' or 'profile descent'. While such instances of ambiguity could be eliminated by improving the terminology of aviation English, others are more deeply rooted in the natural language, English, on which it is based. For example, Cushing reports a case that nearly led a midair collision that was caused by the phonological similarity of *two* and *to*: A pilot misunderstood the command *climb two five zero* as *climb to five zero*.

In addition to lexical ambiguities, and often in connection with them, structural ambiguities of language play a role, and it is difficult to see how they could be eliminated or even reduced. Cushing mentions a case in which *back on the power* could be parsed as [*back on*] [*the power*] or as [*back*] [*on the power*].

Perhaps even more important than lexical and structural ambiguities are ambiguities that are related to the use of pronouns and other anaphoric devices of language, such as contextual ellipses. One could imagine trying to avoid such anaphoric expressions if possible; but on the other hand they have developed to make human language more efficient. A rule like "avoid pronouns" or "avoid ellipsis" would very likely be counter-productive in making communication too verbose and slowing it down. However, it is feasible that awareness of the potential dangers of pronoun use and ellipsis could improve communication in the cockpit.

1.2 Problems of Communication

In addition to the problems that in one way or other relate to structural problems of language and its relation to situations in the world, there are specific problems of communication in the cockpit (see, e.g., Sexton & Helmreich 2000, Nevile 2001, Silberstein 2001). These can arise, again, due to overload of the cognitive system of the pilots in situations of high taskload or danger, or because of the specific social situation in the cockpit with one experienced pilot with higher status and legal power, and one typically less experienced first officer with lower status (and, formerly, a flight engineer with yet fewer rights and lesser status).

Problems stemming from the asymmetric social situation in the cockpit have been analyzed by Goguen & Linde (1983), see also Linde (1988). In a detailed analysis of NTSB transcripts, Goguen & Linde showed: (i) Speech acts by the First Officer are often mitigated, that is, more indirect, and often belittling potential or real problems, than speech acts by the Captain. (ii) Speech acts that do not lead to their intended effect are more often mitigated than speech acts that do. Goguen & Linde explicitly discuss the United Airlines accident in Portland of 1978, in which the officers failed to communicate the growing emergency of the fuel situation to the captain. For example, when the captain requests from the engineer a projection for the fuel situation for the next fifteen minutes, which he estimates to be 3000 or 4000 pounds, the engineer answers: *Not enough. Fifteen minutes is gonna --- really run us low on fuel here.* When, later, the captain reports the fuel situation by *showing a thousand or better*, the first officer challenges this just by *I don't think it's in there*. When the fuel is, as a matter of fact, nearly used up, the flight engineer reports *Not very much more fuel*. Five minutes later, the engines stopped.

Goguen & Linde analyzed flight recorder tapes of real accidents using analytic categories derived from speech act theory. The main interest was on mitigation and aggravation of speech acts, on speech acts of planning and explanation, and on the topic success or topic failure of speech acts. The speech act categories used were rather coarse. Goguen & Linde distinguish the following types: Requests (which includes orders, requests, suggestions, and questions), Reports (which includes simple reports, supporting and challenging reports, and psycho-ostensive reports), declarations (like declarations of emergency) and acknowledgements. The data, which were from real accidents as transcribed from flight recorders by the NTSB, certainly were authentic; however, the type and development of the accidents were naturally very different.

2 The First Project Phase

2.1 General Description

In 1999 one of us (Manfred Krifka, then University of Texas at Austin) started a project of investigating the fine structure of linguistic communication in the cockpit. The general aim was to look at properties of language use, in the general tradition of Linde and Goguen. However, we intended to use a more fine-grained and linguistically informed set of categories. Secondly, the data investigated should be more comparable, which meant investigating flight simulator transcripts of simulated flights that follow a comparable scenario. The project was carried out with Carrie Clarady as an assistant. It was supported by the Gottfried Daimler and Carl Benz Stiftung within the setting of the GIHRE project. The data we worked with were transcripts of flight simulator recordings on B-727 aircraft of a study on the effect of captain personality on crew performance carried out by NASA-Ames Research Center in 1987 (see Chidester ea. (1990) for a description). The transcripts of the video and audio recordings were done by Robert Helmreich and Brian Sexton of the Texas Airospace Crew Research Project.

The data available consisted of 17 transcribed simulation flights. The flights followed roughly a similar scenario, even though the reactions of the pilots could lead to quite different situations. There were five segments to each flight. A first segment was, unfortunately, not recorded; the four following segments were labeled A, B, C and D. Segments A and C were segments with, roughly, medium task load; segments B and D were segments with high task load. Segment A consisted of a climbout from San Francisco Airport and a descent and landing into Sacramento Airport. Segment B consisted of a descent and landing at Los Angeles Airport, with simulated events like a jammed stabilizer and low oil pressure. Segment C simulated a descent to Sacramento Airport, with a missed approach. In Segment D, the plain was diverted to San Jose Airport, with simulated events like hydraulic malfunction and a split flap malfunction. The performance of the crews was independently rated. In the limited time available for this project (April 1999 to July 2000) we could investigate altogether 5 flight simulator sessions completely, were we picked the two best-rated crews and the three worst-rated crews. Altogether, we investigated 6900 "thought units", or units of speech as provided by the transcribers, which we tried to follow in order to secure comparability with potential other research using the same material. This was altogether about 9 hours of conversation.

Our goal was to investigate whether there are features of communication that correlate with, on the one hand, the well-performing crews and the poorly performing crews, and, on the other, with situations of medium task load and of high task load or potential danger. It was obvious early on that given the nature of the material, in which the scenarios differed quite substantially, and our limited project resources, we could not hope to arrive at assertions that could attain a level of statistical significance. Rather, we intended to find out whether there are phenomena that could be investigated later in a more rigorous way.

The data set analyzed here is a subset of the one that was analyzed by Sexton & Helmreich (2001), using the *Linguistic Inquiry and Word Count* (LIWC) program described in Pennebaker & Francis (1999). The main findings in this study were: Captains used a higher number of words per segment; the numbers of words were higher in segments of high task load; Captains more often referred to the group by *we*, especially in segments of high task load; large words (greater than six letters) was negatively related to performance and positively related to error.

2.2 Analysis and Findings

We investigated the material in a number of dimensions. First, we were interested in the actual involvement in conversation. The crew consisted, as it was common in the 1980's on commercial aircraft of this size, of three members, the Captain (CA), the First Officer (FO) and the Flight Engineer (FE); in addition, there is the Air Traffic Controller (ATC).

2.2.1 General Characteristics

The following chart (Figure 1) gives a typical picture of the communication density during a simulation flight (here, the flight of the "good" crew 8). The number of contributions of each participant of a particular flight, which were called "thought units" by the transcribers, were added for each minute during the whole duration of the flight. We see that communication density is higher in the difficult segments B and D, that the captain in general has the most contributions, and that the flight engineer has more contributions in the difficult segments B and D (see Figure 1).

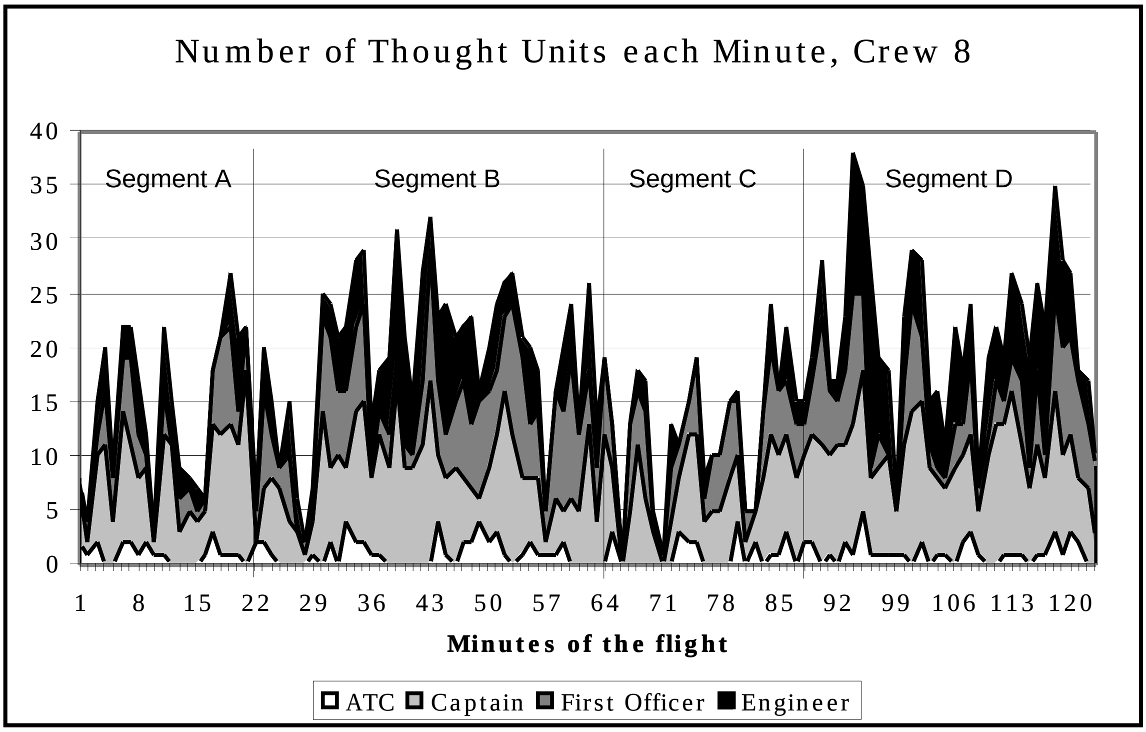

Figure 1: Number of Thought Units for each minute of the flight, Crew 8

In the following charts, Segments A and C are combined to the “Medium” Taskload Segments, and Segments B and D to the “High” Taskload segments. Also, Crews 3 and 4 are combined to the “Poor” crews, and Crews 5, 8 and 13 to the “Good” crews. “PoorM” refers to the medium task load segments of the poorly performing crews, “PoorH” to the high task load segments of the poorly performing crews, and “GoodM” and “GoodH” refers to the medium and high segments of the well performing crews. The charts give average values, and indicate standard deviations.

In general, high taskload segments were longer than medium taskload segments. Abstracting from the duration of the segments, there were more utterances per minute in the segments of high workload, and the utterances were longer, leading to a substantial difference of speech time per minute (see arrows, Figure 2). There was no clear relation to the performance level of the crews.

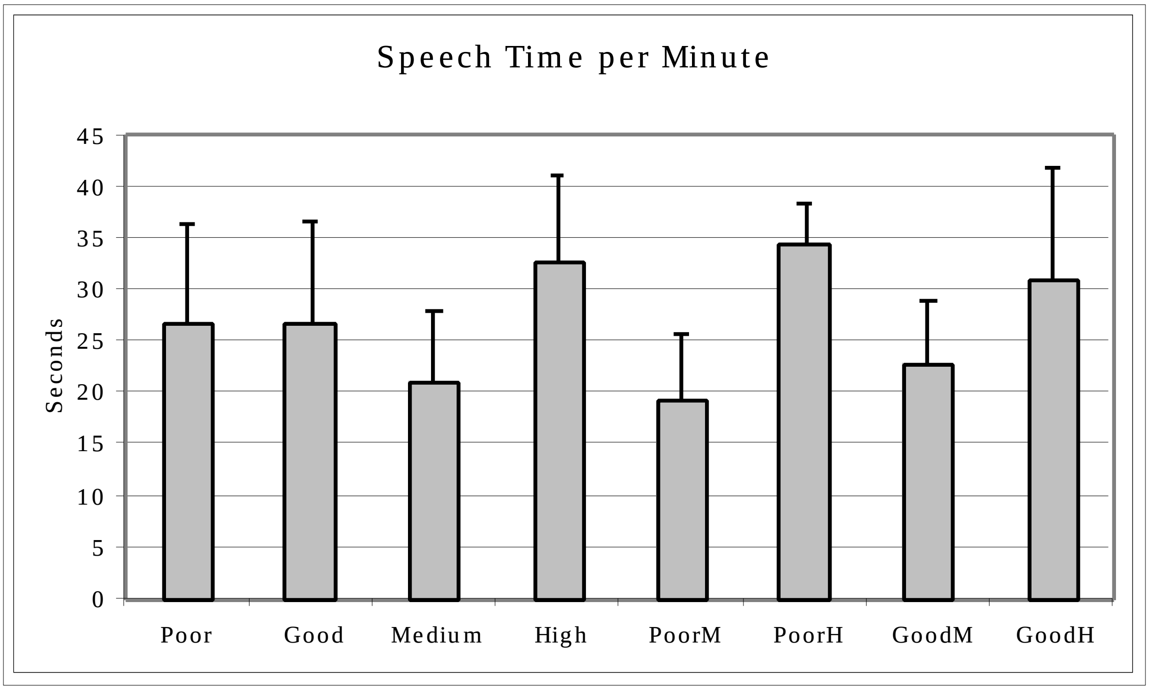

Figure 2: Speech Time per Minute, in Seconds
(for lack of complete data, Crew 13 is excluded)

The segments also differed with respect to speakers. While the captain assumed the speaker role most often overall (typically, about 40%), the engineer assumed the role of speaker more often in high task-load segments. The engineer was more often the addressee in the difficult segments, again with no clear differences between crews. Figure 3 and Figure 4 show the Speaker and Addressee role per thought unit (The figures for the addressee role do not add up to 100% as the addressee of some utterances cannot be identified properly, or is the whole crew).

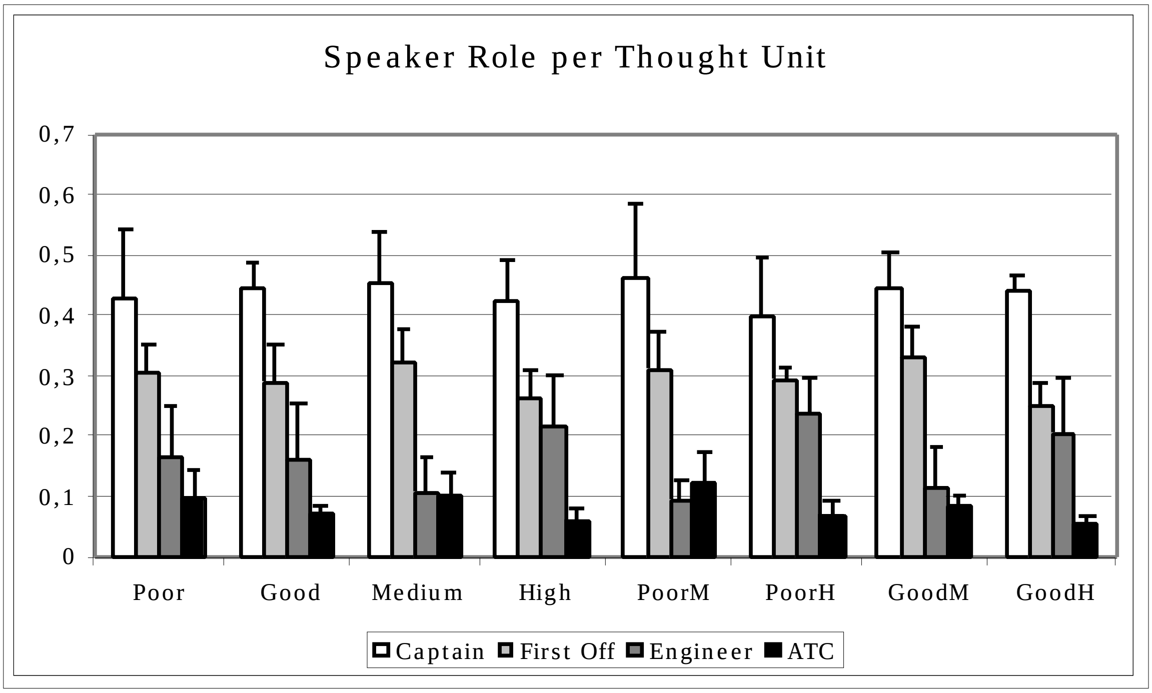

Figure 3: Speaker Role (Captain, First Officer, Engineer, Air Traffic Control) per Thought Unit

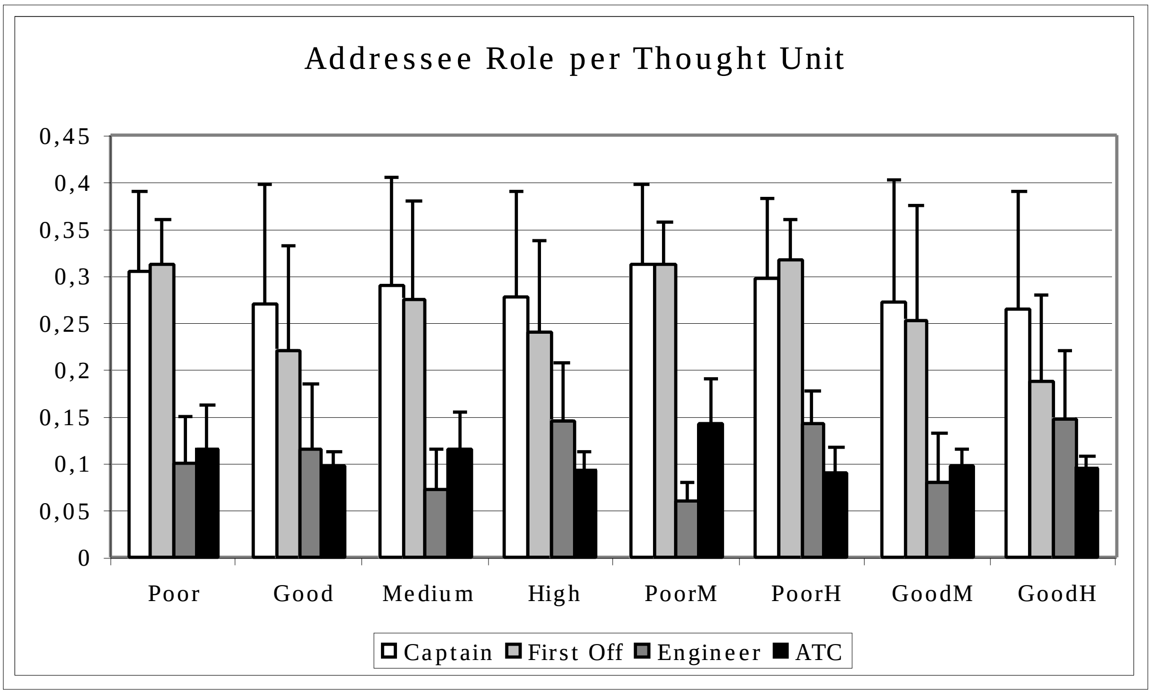

Figure 4: Addressee Role per Thought Unit

2.2.2 Speech Act Analysis

Our major efforts concentrated on an analysis of the utterance types informed by speech act theory because we hoped that in this area we could find differences between crews and task loads in the types of speech acts used, and that recommendations concerning the use of particular speech act types could conceivably become part of training programs for pilots. Our categories were inspired by the classical work on speech act theory, most notably Searle (1975), but we felt that the specific area of cockpit communication required more detailed speech act categories in particular areas. We decided to use the following categories (with examples added from Crews 4 and 8). In the transcript, *abc 801* refers to the plane.

- Status Reports (report current state of equipment, weather, location, etc.)
  [4.17] *uh roger uh we are out at 1443 and off at 53 from san francisco*
  [4.62] *uh, abc 801, be adviced the meter locator outer marker is unreliable, i'll call you at meter*
  [8.28] *we have an underexcited fault light on, uh, number 3*
- Reports of Action (speaker gives a report his own action)
  [4.75] *i keep looking for the altitude right here and see 801*
  [4.111] *i'll put it on the missed approach altitude*
- Reports of Report (rephrase information expressed before)
  [4.461] *well, he told us be ready for ils 24 right*
- Prognoses (express likely future course of events)
  [4.373] *okay, so i think if we do the two engine, the two generator operation that'll be uh, the next thing.*
  [8.214] *he's setting you up it almost looks like for a lot of bracketing here, bob, so…*
  [814] *we dump fuel it makes the, uh, out of trim situation worse though.*
- Diagnoses (express likely cause of past events)
  [8.92] *i think we drifted up.*
  [8.111] *i think what that light was, was when that field relay tripped.*
  [8.1926] *so we just weren' going far enough.*
- Commands (request an action by addressee)
  [8.94] *hey, with this fault light, would you switch essential power off of number 3, please for me*
  [8.212] *abc 801, that turn looks like it's going to take you through, so continue the right*
  [8.356] *press on*
  [8.64] *okay, now you fly the airplane*
  [8.224] *why don't you make it 190.*
  [8.305] *final flaps should be 30*
- Permissions (give permission to perform action to addressee)
  [8.19] *abc 801's cleared for traffic. you can descent at pilot's discretion.*
  [8.15] *oh, whatever you want to do here, bob*
- Complies (verbalizes an action that is performed to carry out a command)
  [8.414] [*how about that brake? now try it.*] *no.*
- Reports of Intention (express intention to act in a certain way)
  [8.32] *meanwhile i'll take care of the approach/descent here*
  [8.63] *we'll…we'll take care of that*
- Expressives (express an emotional state of the hearer)
  [4.185] *uh, shit*
  [8.639] *little devil*
  [8.904] *boy that is a lot of pressure*
  [8.989] *great, excellent, no, excellent*
  [8.1019] *come on, baby, come around here*

- Acknowledgements (express that preceding speech act was understood)
  [4.13] [*abc 801, expect ils 16 approach to sacramento*] *okay*.
  [8.410] [*i'll just crank it and you tell me when to stop*]. *yeah*.

- Affirmations (acknowledge and affirm a preceding speech act)
  [8.1075] [*well we got… we're committed to the right*] *that's right*
  [8.1120] [*can't do it*] *no* [as an agreement].
  [8.1169] [*shouldn't be that much of a problem*]. *nope, no no no.*

- Rephrases (acknowledge and rephrase preceding speech act)
  [8.50] [*one six right, approach.*] *one six right, yeah.*
  [8.1640] [*ground spoilers, out spoilers*] *out spoilers.*

This classification combines speech act types with content features (e.g., status reports, prognoses, diagnoses are all assertive speech acts that differ in their content). The last three categories concerned with the proper flow of information in conversation are often neglected in speech act theory, but we considered them to be of potential importance in the present context. In general, speech act theory (as opposed to discourse analysis) underestimates that many speech acts are parts of well-defined sequences (e.g., acknowledgements of statements, verbal complies to commands, etc.).

Certain generally recognized speech act types do not form separate categories, but rather cross-classify other categories. This is most obvious with questions and answers that can be related to assertive speech acts, i.e. state reports that can be true or false, or to commands. For example, *You got any problems?* was classified as a Status report / Question, as the answer would count as an assertion, or Status Report, whereas *You want me to brief with you?* was classified as Command / Question, as a positive answer would count as a command. The question *Was that what screwed it up?* was classified as Diagnosis / Question, and the answer *Might have* as Diagnosis / Answer. The speech act of Command (and Comply to Command) could also have been cross-classified; a command like *Tell me about the weather condition* requests a linguistic action, and hence is equivalent to a question. However, we didn't do so as such commands appeared to be quite rare.

A well-known problem for the classification of utterances into speech act types is that the linguistic form often does not determine the act type. A command can be expressed directly, by an imperative (*Close the door!*), or more indirectly, by a modal declarative statement (*The door should be closed*) or a question (*Why don't you close the door?*). We would classify such utterances as commands. Only a very indirect speech acts that require inferencing (e.g. *It is cold in here*) would have been classified directly (here, as a status report), as it would not be clear whether the speaker intended this as an assertion or as a command.

Another problem was that the segmented utterances in the transcripts (the "thought units") quite often comprised more than one speech act. As we wanted to maintain the integrity of the transcripts to ensure comparability with other studies, we did not break up such utterances. Nevertheless, we were forced to classify each utterance as belonging to one speech act. For example, the utterance *ABC 801, that turn looks like it's going to take you through, so continue to the right turn to 180* could have been broken up into a status report and command, but was classified as a command (the first part can be seen as a motivation for the command). Take another example: *I'm on the air, so you get it.* This could be analyzed as a command, or as an assertion and a command, or perhaps as a new speech act, "motivated command". Again, as the main function of this utterance seems to be to get the addressee to do something, we classified it as a command.

2.2.3 Other Features of Communication

Beyond the analysis of speech acts we looked at a number of other features that appeared likely candidates for the correlation with the quality of the crew or the task load.

- Explicit reference to addressee, speaker, or group
  [8.31] *why don't* ***we*** *go to the book and see what* ***you*** *can do on it.*
  [8.32] *meanwhile* ***i****'ll take care of the approach/descent here*
  [8.41] ***you*** *got any problems at all,* ***bob****, back there with that?*

- Correction of previous information
  [8.175-177] [*lots of gas. about* ***29***] ***28.***
  [8.266] *do you want abc 801 to the* ***marker*** *er…* ***tower****?*

- Evidence for misunderstanding
  [4.45-46] [*atc: abc 801, contact sacramento approach on* ***1256****.*] CA: *801, uh,* ***276****, roger.*

- Hesitation
  [4.121] *abc 801 is meter,* ***uh****, inbound.*

- Hedges (*sort of*, *I think*, etc.)
  [4.156] ***well****, that is true.*
  [4.164] ***i think*** *what he was saying was be prepared for a fillmore 4 and uh a.*

- Encouragement
  [8.134] *looking good, bob.*
  [8.190] *this is excellent.*
  [8.552] *you can fly it… hold it…*

- Emotional words
  [8.620] *we don't have a hell of a lot of gas to screw around with either.*
  [8.640] *little devil.*
  [8.905] *boy is that a lot of pressure.*
  [8.82] *ah, shit*

- Politeness
  [8.94] *hey, with this fault light, would you switch essential power off of number 3, please for me.*
  [8.105] *yessir, we just heard some other voices in the background, thank you.*
  [8.905] *pull back just a touch please.*

2.2.4 Results

We now report some of our findings. One should keep in mind that the numbers of cases we are working with are too low to achieve statistical significance, which is also clear by looking at the large standard deviations in many cases. But we do hope that some of the results may lead to worthwhile hypotheses for future work.

First, we found more prognoses and diagnoses in good crews and, to a lesser degree, under high task load (Figure 5). In particular, speech acts that can be thought of prognoses of one's own future behavior, which we called "report of intention", and which were not counted as prognoses, were more frequent in the good crews and, less dramatically so, in the segments of high task load. (Figure 6).

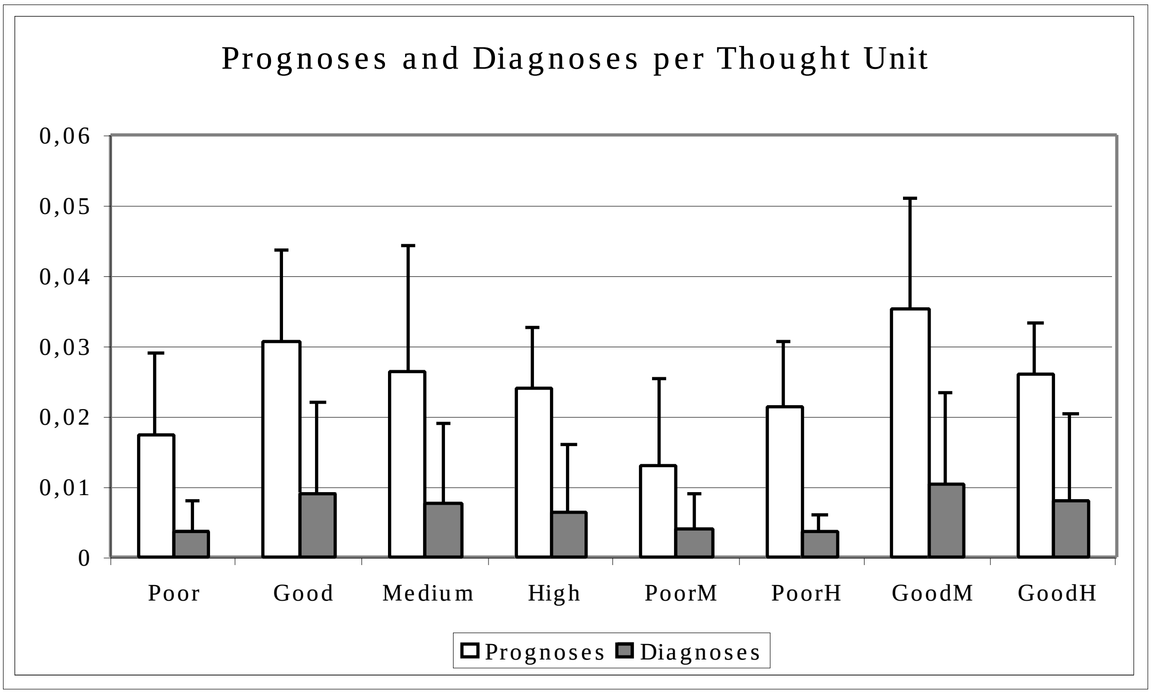

Figure 5: Prognoses and Diagnoses per Thought Unit

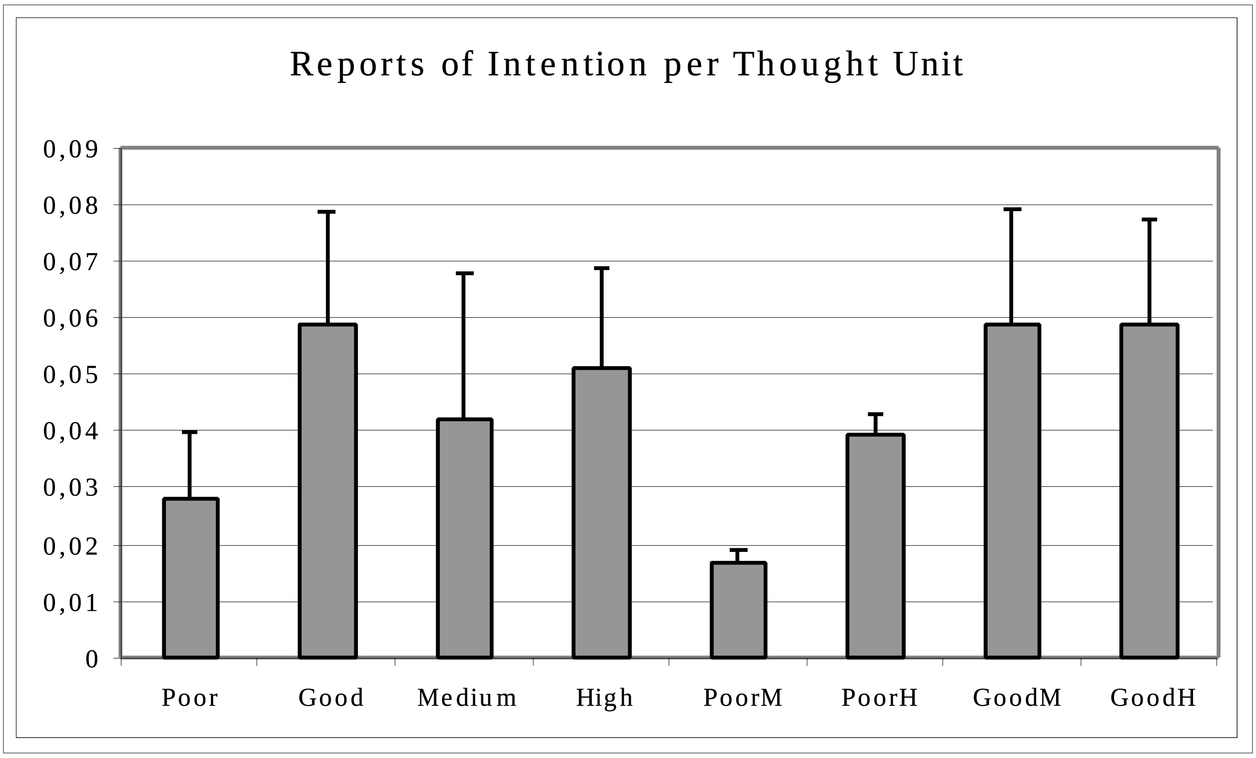

Figure 6: Reports of Intention per Thought Unit

Questions occurred slightly more frequently in the good crews, and the proportion of questions that were answered was considerably higher there. Interestingly, poor crews had more questions in the high task load segments, whereas good crews had more questions in the medium task load segments (see Figure 7).

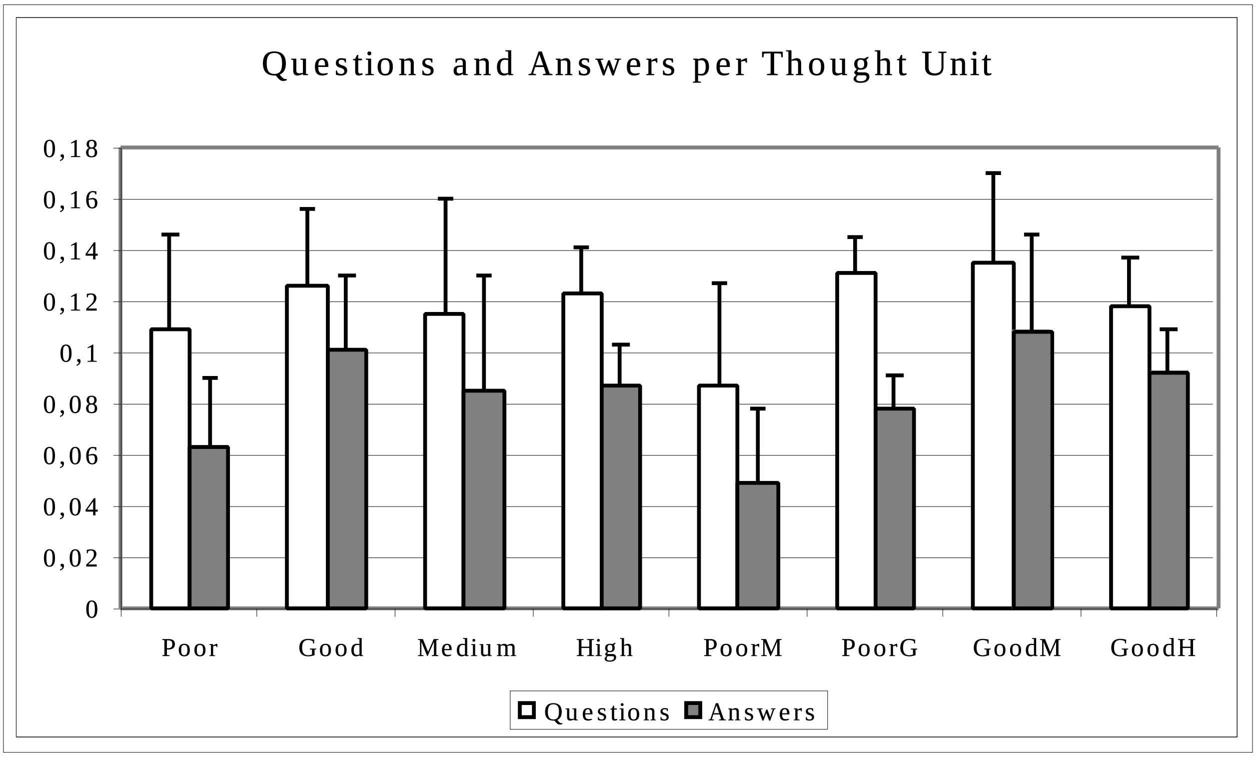

Figure 7: Questions and Answers per Thought Unit

Commands occurred more frequently in the poor crews, as well as overt linguistic verbalizations that commands are carried out. There was no relation with task load. (See Figure 8).

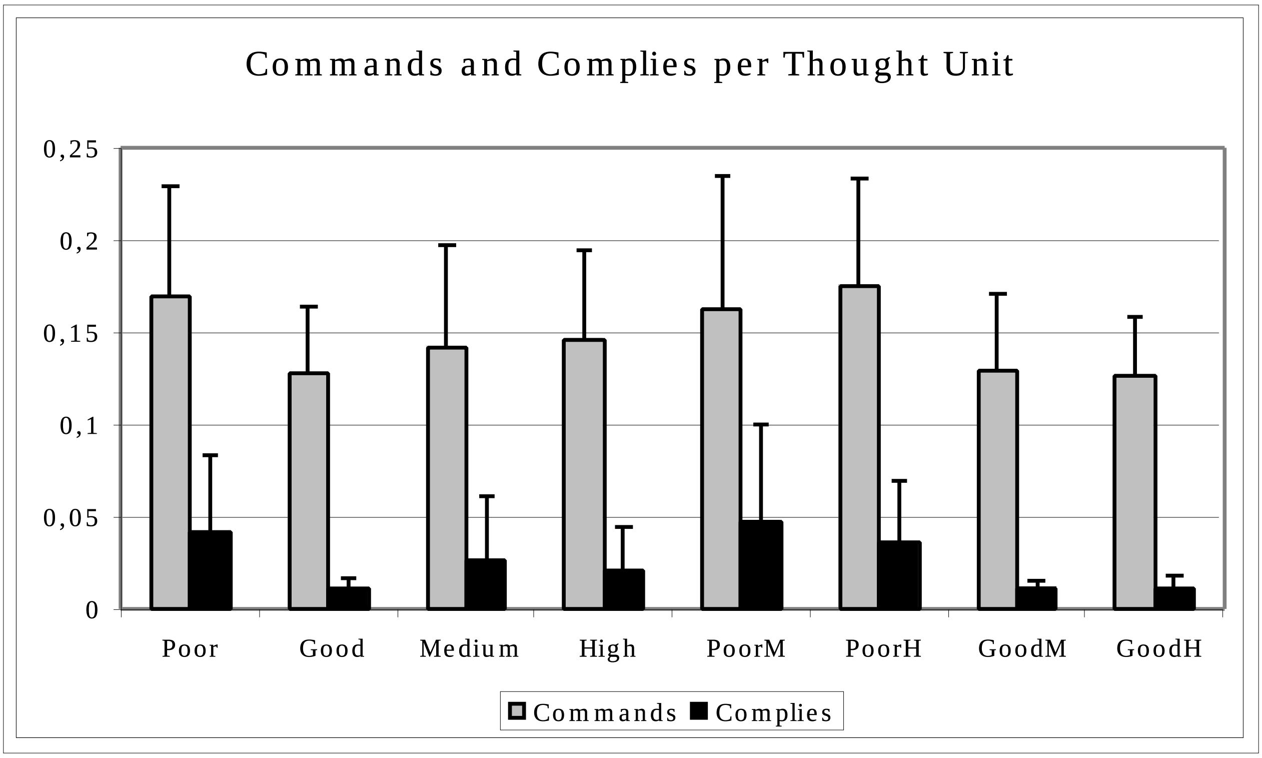

Figure 8: Commands and Complies to commands per Thought Unit

As for the speech act types that acknowledge or confirm another speech act type, which are quite characteristic for cockpit talk, we find generally more simple acknowledgements in the good crews, and fewer in the segments with high task load (Figure 9).

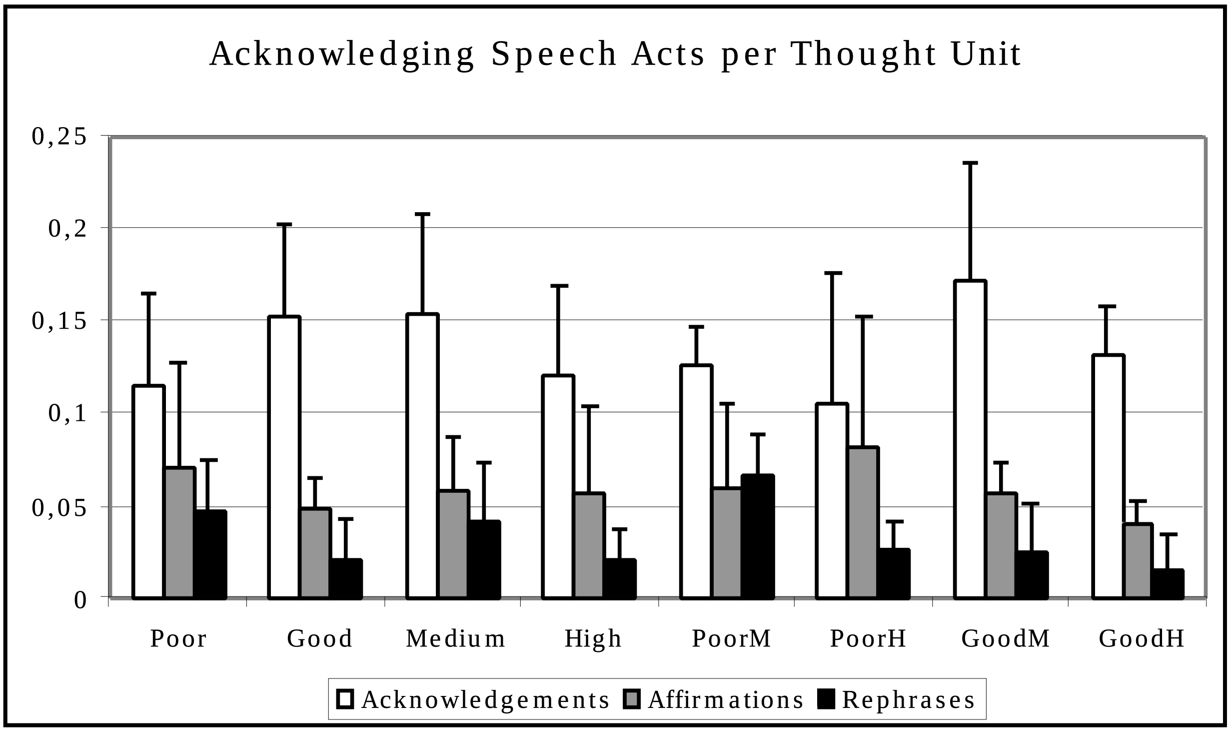

Figure 9: Acknowledgements, Affirmation and Rephrases per Thought Unit

Expressive speech acts and emotion words occur more frequently in the poor crews. Interestingly, in those crews segments of high task load showed fewer expressive speech acts and emotion words (see Figure 10).

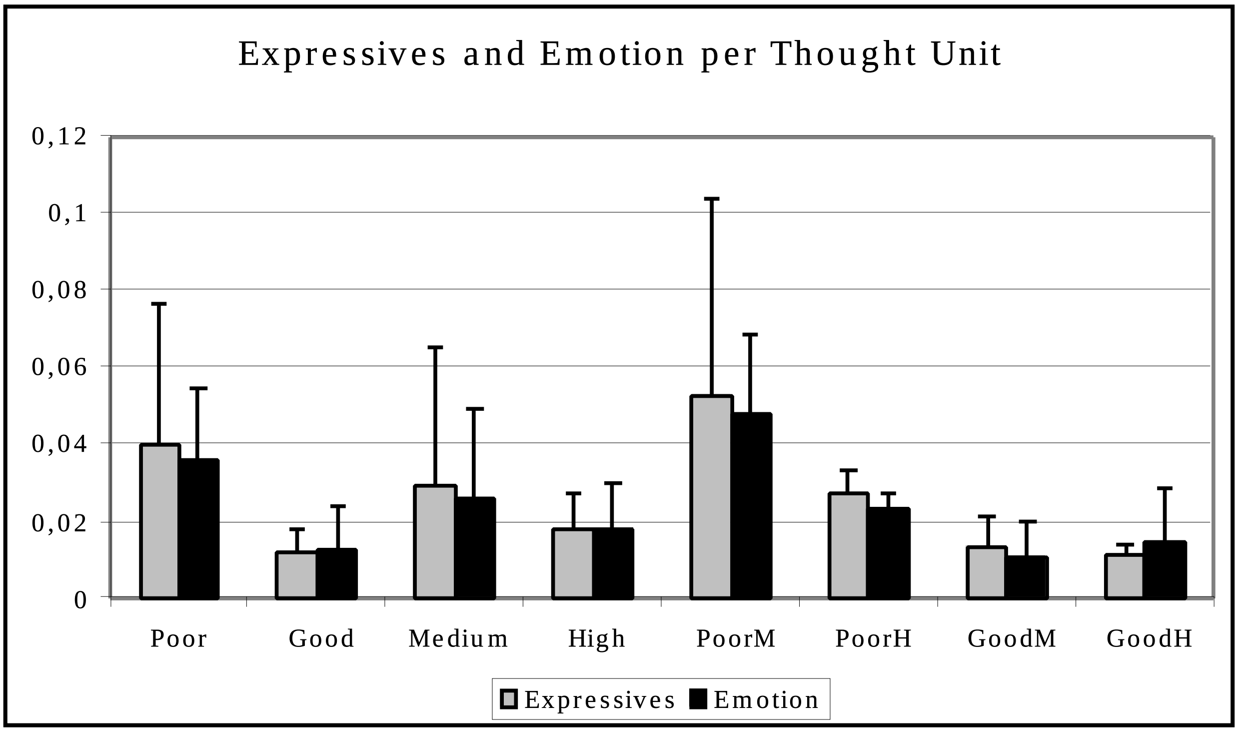

Figure 10: Expressive Speech Acts and Emotional Expressions per Thought Unit

Linguistic evidence for misunderstandings was found especially in one poor group, more frequently in segments of medium task load. Evidence for corrections occurs in all segments and crews about equally (Figure 11).

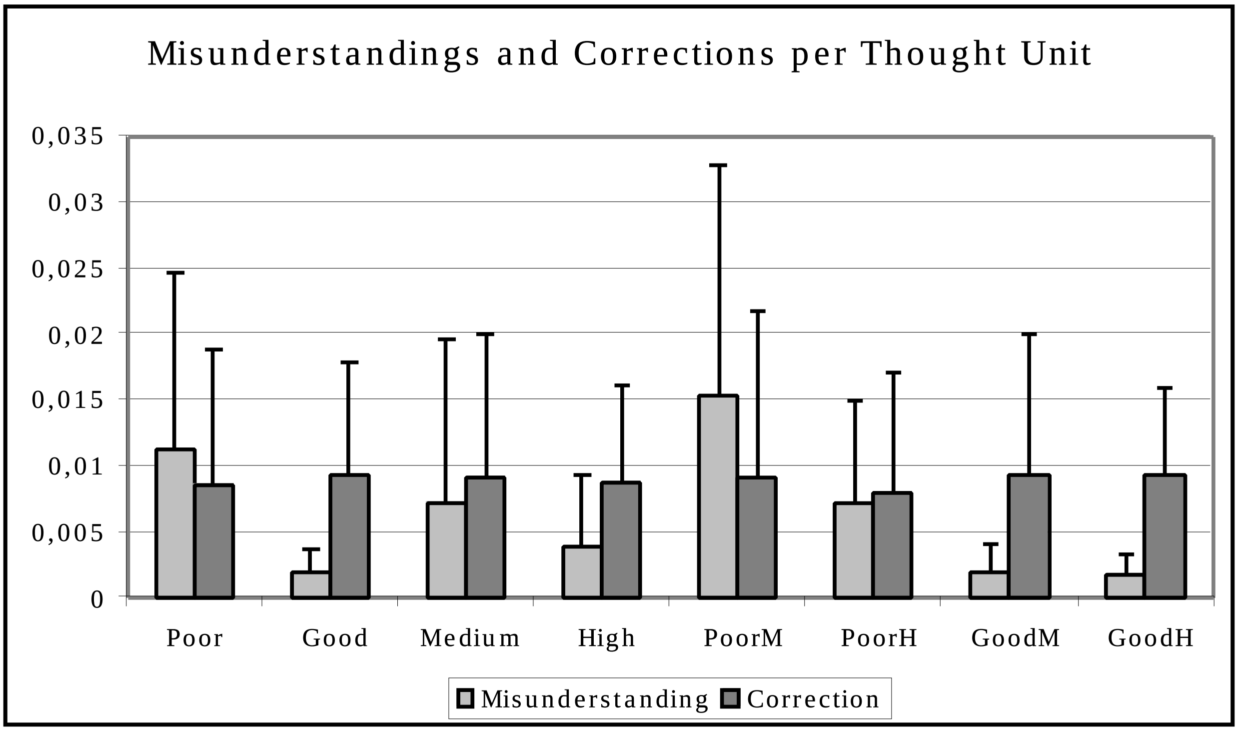

Figure 11: Evidence for Misunderstandings and Correction per Thought Unit

With hesitations and hedges we find the following picture: Hesitations occurred slightly more frequently with the poor crews. In these crews, the incidence of both hesitations and hedges increased with high taskload. (Figure 12).

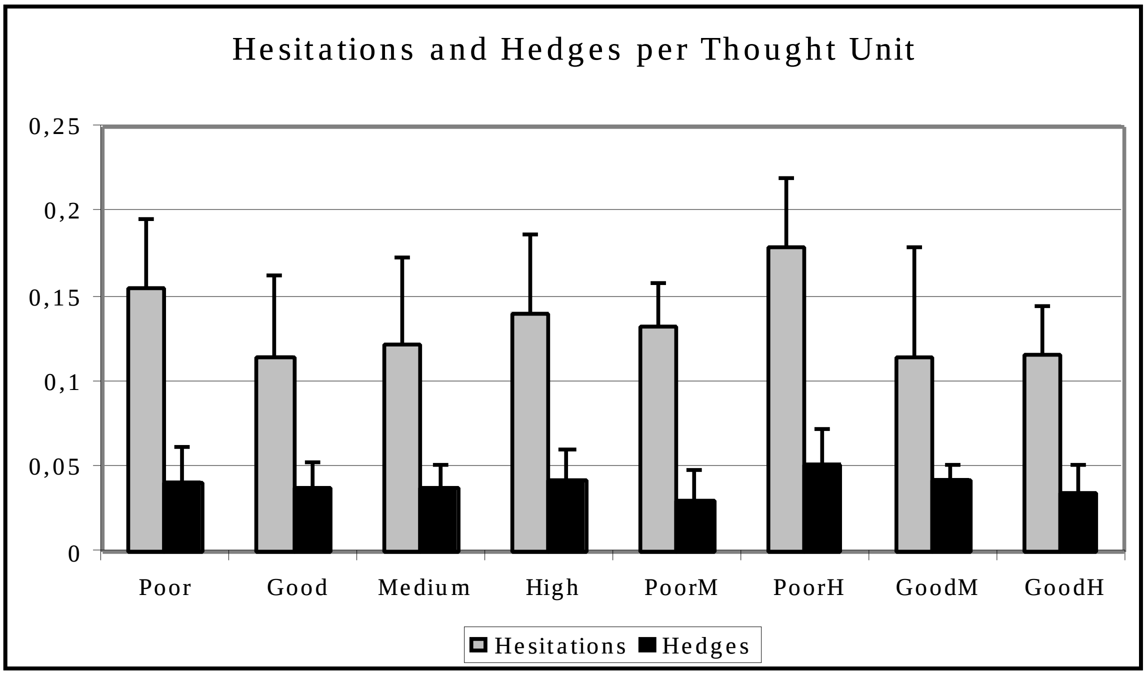

Figure 12: Hesitations and Hedges per Thought Unit

Politeness elements occur slightly more often in good crews. Interestingly, good crews have fewer politeness elements in segments of high task load, just the opposite to poor crews. (See Figure 13).

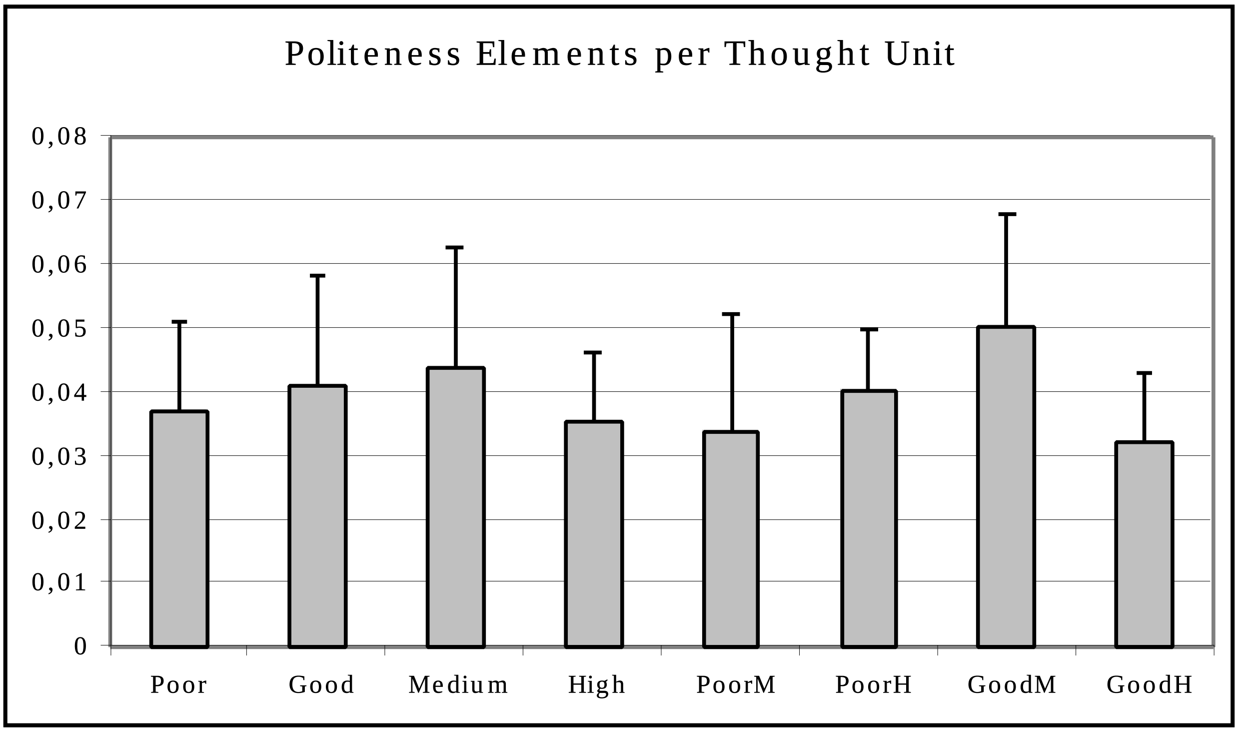

Figure 13: Politeness Elements per Thought Unit

We also counted explicit references to the speaker, the addressee and the group (see Figure 14). Especially reference to groups can be considered a group-building measure, Sexton & Helmreich (2001) found out that reference to the group increases over the life of a crew, and that captains use more reference to crews than first officers or engineers. They all occur somewhat less frequently with poor crews in medium task load segments, but increase in high task load segments. For good crews, they stay about the same.

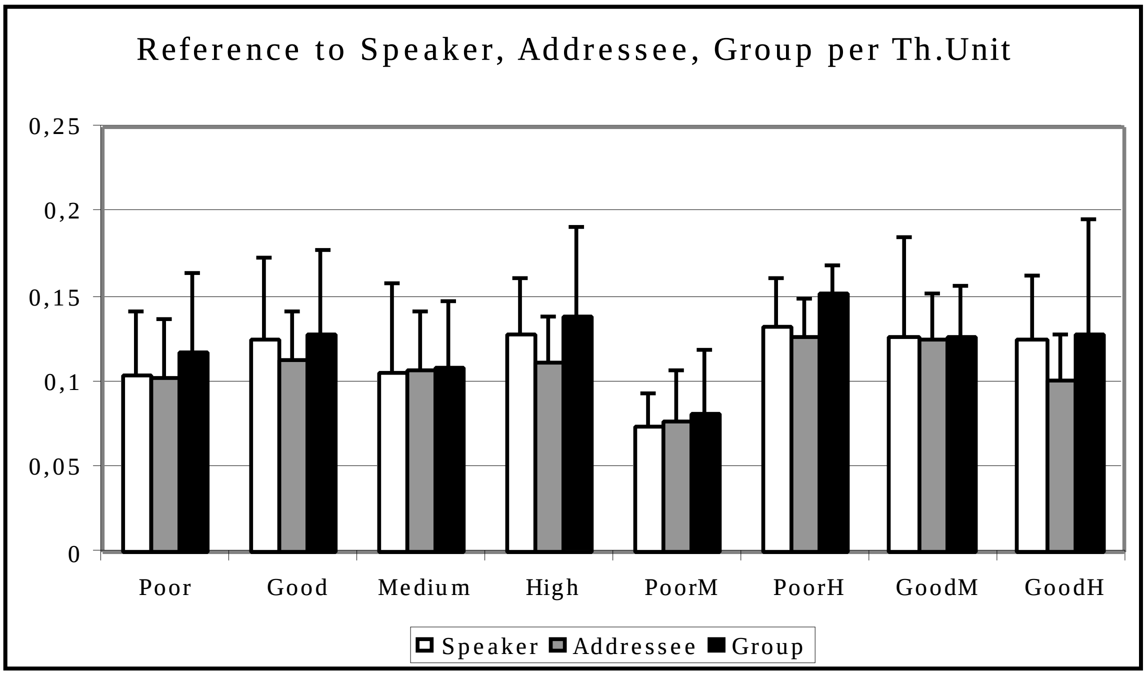

Figure 14: Reference to Speaker, Addressee, and Group per Thought Unit

Of the other features that we investigated, we would like to mention encouragement. This category occurred quite frequently in the best-performing crew (in about 5% of the thought units, more frequently in those with medium task load), and not at all in the worst-performing crew.

2.2.5 Limitations

The scope of the first part of the project was more limited than originally envisioned (we planned to analyze more flight transcripts), and therefore the results of this part have to be understood as indicators of tendencies at best. Nevertheless, we could identify some linguistic dimensions in which there is a relatively clear influence of workload and crew performance. For some cases, like the category of reports of intention and of prognoses in general, which was higher in difficult segments and with good crews, one could develop specific advice for crew resource management training.

But it also became clear in the course of the study that this type of research could be improved in a number of ways, some of which are listed below.

- The classification of segments A/C and B/D as low workload vs. high workload does not capture the fact that high workload stretches also occur in A/C, and low workload stretches occur in B/D. Thus, the correlation of workload with linguistic behavior may appear less pronounced in our analysis than they actually are.

- The number of crews that were analyzed is far too low to arrive at statistically significant conclusions with respect to crew performance. This is also reflected in the standard variation in the charts, which often is quite high.
- The complete analysis of simulation flights (about 7 hours), which included relatively routine stretches like checklists, was, in hindsight, not the most efficient way to apply the limited resources of the project.
- The quality of the transcripts we worked with was rather varied, e.g. in the recording of hesitation signs or other paralinguistic behavior.
- The information available about the flight scenario and the precise timings of malfunctions was limited.
- We did not have the means for extensive independent classification of the communicative features, like the speech acts. Some informal sample tests showed an interrater consistency of 70% and higher, depending on the category.

We also had the impression that quite a few types of linguistic behavior were not identified by our system of categories. For example, the following acts quite obviously invoke crew resources; we found such passages often in the best Crew 8, and rarely, if at all, in the worst Crew 4:

> [8.1511] *you can, you can listen with me, I might misunderstand.*
>
> [8.2162] *what else? what else have we missed?*
>
> [14.1118] *Okay, do we have everything?*

The technique that we have applied is particularly bad with coherent passages of discourse. For example, the following passage, a dialogue between captain and engineer, is remarkable because the interlocutors try to secure a precise understanding by asking back and forth:

[8.146]E-C: *climb to 500 feet, proceed directly sacramento v-o-r hold south 180 radial, 5,000 feet.*
C-E: *5,000?*
E-C: *yeah.*
C-E: *after 5000, direct to v-o-r to hold?*
E-C: *right.*
C-E: *okay got it.*

And in the following passage the captain focuses the crew resources in an admirable way:

[8.2007] F-C: *i'm worried about the gas. we're pissin' out 15,000 pounds an hour.*
C-F: *don't worry.*
C-F: *don't worry.*
F-C: *ok.*
C-F: *i'll tell you when to worry.*

One problem that is difficult to address is the issue of whether simulator data are representative of real data. While the participants in the simulated flights obviously took their job very seriously, there is evidence that they were frequently aware of the fact that they were operating in a simulator. We hope that, nevertheless, the communicative behavior in simulated situations is similar enough to real-life situations to make these analyses relevant for such situations.

3 The second project phase: an outlook

Due to the various problems in the first project phase several changes and improvements have been made in the second phase that started in May2001. The improvements, which will be discussed in detail in the following, concern the scenario, the data corpus including the transcripts and the modified methods of analysis. At the time of writing of this article we are just beginning our analysis, as the technical organization of the data has taken extensive efforts. Therefore we can not yet present any results.

The new project phase differs from the first project phase not only in respect to the methods of investigation but also in terms of the points of emphasis within the investigation. The general goal of our research, however, remains the same: To identify specific linguistic features of cockpit communication that correlate with crew performance, especially under varying levels and types of taskload. All three parts of the flight-simulator scenario (description see below) are classified as medium to high task load segments, but the type of task load varies. One part, for example, is demanding mainly in terms of manual skills, while another requires analytical skills. All parts involve high time pressure.

Our goal is, as in the first project phase, to find out which concrete linguistic factors are crucial for the crew's communicative behavior and what differences there are relative to different types of task load and different crews. Comparability of the different crews, an essential factor, is ensured by the design of the new scenario, which is considerably more constrained and leaves less room for variation in the course of events. Furthermore, communication skills are crucial in resolving the occurring problems. In contrast to the first project phase, we do not know the performance ratings of aviation experts in advance, but we will compare our findings to find out if and in which ways crew performance and communication do correlate. Also, we will systematically test for interrater consistency.

3.1 The New Flight Simulator Scenario

The first point of improvement concerns the structure of the scenario: It is substantially more focussed, more clearly oriented towards communication. It was developed by a group of aviation experts around Gerhard Fahnenbruck and carried out by Lufthansa CityLine on a CanadAir Regional Jet simulator.

The scenario consists of three parts: First, a so called circling approach, where the runway has to be approached with the wind from behind due to an instrumental landing system (ILS) failure on the ground. With the runway in sight, the crew has to circle around the airport in preparation to landing. The pilots have to brief and prepare for this particular type of approach. When they are about to land, another plane is blocking the runway and the crew has to fly a go around. In the second phase that begins at this point, the ILS has been repaired and a normal approach to the runway can be flown. During the approach, however, the glide slope, a part of the instrumental landing system, breaks down at 1500 feet. Constant communication about the altitude both within the crew as interactional partners and with the tower is necessary to be able to land under these conditions. After the landing the third section starts with a take-off. Shortly after the plane is in the air, an instrument failure occurs while the plane is in the clouds: the artificial horizon (pitch) on the captain‘s side and the compass (heading) on the first officer‘s side display incorrect values, the captain's instrument being far more important. This section is especially interesting from a linguistic point of view, as there is an increased need for communication in order to exchange the correct values of the defective instruments.

Since each crew goes through the same scenario, with the possibility of different developments within the course of events being quite limited, the degree of comparability is much higher than during the previous phase of the project.

3.2 Data Corpus and Transcripts

The next point of improvement is the quality of the data. We have a corpus of 16 simulator flights (video and audio data), all consisting of the same scenario. Differing from the previous project phase, the crews consist of only two members, the captain and the first officer, because there is no engineer on this type of plane. The shortest flight lasts about 27 minutes, the longest one hour and 28 minutes, the average length per flight is 53 minutes. The whole corpus consists of 14 hours and ten minutes. As it will not be possible to transcribe, let alone analyze, the entire 14 hours of data, certain well-defined segments will have to be chosen for our purposes. We intend to transcribe and analyze three comparable parts of each scenario in which different crews have to manage very similar tasks. In the third part of the scenario, for example, the segment to be analyzed starts at the moment a warning signal indicates an instrument failure to the crew and ends when their analysis of the problem is completed. Thereby we ensure that the objective tasks that the crew has to solve and the contexts in which they have to be solved are very similar, if not equal.

3.3 Methods

The methods used in the second project phase will be substantially different from the ones in the first phase. In addition to the (adjusted and changed) speech act theoretic investigation we will also apply methods from conversation analysis and try combining the two approaches. This means that our strategy will include quantitative analysis, since the occurrences of speech act types can be counted, as well as qualitative analysis.

3.3.6 Applied Speech Act Theory

Due to the better quality of the data corpus and by applying recent developments in applied speech act theory, we will be able to use a more fine-grained system of speech act types. Also, we will not rely on a previously determined dissection of the communicative events into sometimes relatively large “thought units” but will be able to assume a more fine-grained division, which will also enhance our ability to assign speech act types in an unambiguous manner. We will go beyond analyzing single, isolated speech acts by considering and developing types based on longer sequences of dialogue. These will also be organized in well-defined speech act classes, which will result in improved possibilities for quantification. We will determine more clearly how often well-defined speech act types occur in which kind of task load type and in which crews. This is essential for our goal of comparing the results with the performance ratings for the crews.

3.3.7 QVA and more fine-grained speech acts

The assignment of certain defined speech act types to utterances is a difficult task and should be valid and reliable. Working with “authentic” material (if we may consider simulator data authentic), this step of analysis should not simply rely on the researcher’s intuitions, in order to guarantee reliability and validity. One method to operationalize this step is *Qualitative Verlaufsanalyse* (QVA), “Qualitative Progression Analysis), as described in Diegritz & Fürst (1999). QVA is a successor of the so-called *Pragmatisch-dynamische Methodenkombination* (PDMK, “pragmatic-dynamic method combination”) of Diegritz and Rosenbusch (see Diegritz & Rosenbusch 1995), developed to analyze the communication between teachers and students in the classroom. It combines different research approaches from linguistics and sociology (e.g. interaction research), with speech act theory playing a central role. It is currently the most manageable instrument to classify utterances into speech acts in a transparent and rater-consistent way. Diegritz & Fürst have developed a *Sprechakttypeninventar zur Analyse von Lehrerintentionen im Gruppenunterricht* (SALG) – a speech act inventory for the analysis of teacher's intentions in group lessons. The goal of SALG is to achieve a sufficient level of objectivity and reliability for the analysis of speech acts in a specific context. In analogy to this we are developing an inventory of speech act types for the analysis of cockpit communication, namely a *Sprechakt-Typeninventar zur Analyse von Cockpit-Kommunikation* (STACK), that, we hope, should be useful for further investigations of cockpit communication.

Following QVA, the segmentation and classification of utterances into speech act types will be broken up in ten microsteps of analysis, as follows: Basic step 0: explication of relevant background knowledge; (1) which illocutionary indicators occur? (2) division of the utterance into speech acts; (3) verbal and non-verbal context, reading back and ahead (optional); (4) explicit interpretative verbalization of the utterance; (5) perlocutionary effects; (6) exploration of the institutional determination of the utterance; (7) explication of aspects of development of thematic progression, the development of relationships within the group and of the group process in general; (8) separation of illocution and proposition and the determination of the individual speech acts and their types; (9) explication and discussion of possible alternative acts and utterances the speaker could have chosen (this serves to explicate the speaker's intention more clearly). See Diegritz & Fürst 1999: 60-65 for a worked example. On the basis of this procedure, the role of particular speech act types in the interactional context of the cockpit can then be investigated qualitatively and quantitatively.

While much of research in speech act theory has been concerned with a global taxonomy for the classification of speech acts, there has also been an increasing interest in subdividing the major speech act classes into a more fine grained system of speech act types. This has been done, with a focus on German, by Rolf (1983) for informatives, by Hindelang (1978, 1981) for directives and questions, by Graffe (1990) for commissives and by Marten-Cleef (1991) and Zillig (1982) for expressives. For an overview of all of these, see Kohl & Kranz (1992). Such a fine-grained system is highly desirable for an empirical application of speech act theory. While it may be advantageous for theoretical purposes to establish a very general taxonomy, the analysis of authentic dialogue seems to benefit from a more fine grained system, since this seems more adequate to the vast variety of possible things that can be done with words.

One disadvantage of a large number of speech act types could be thought to be that the quantitative analysis does not lead to statistically relevant numbers of occurrences of a certain type. But then it is unlikely that all speech act types do occur in the cockpit. We will therefore, as already mentioned, create an inventory of types likely to occur in the cockpit. This inventory should contain a systematic and perspicuous explication of the relevant speech act types. The speech act types will also be organized in a hierarchical structure. This will allow for great flexibility in the level of abstraction used in the quantitative analysis. For example, we make a distinction, following Rolf (1983), between *mitteilen* ("informing") and *melden* ("reporting"), where the characteristic of *mitteilen* is that the speaker acts without obligation and assuming that the information conveyed should be of interest to the addressee, whereas what is characteristic for *melden* is that the speaker conveys the information under a certain obligation. If this does not lead to any interesting correlations, we can collapse this distinction and investigate whether the group of transmissive speech acts as a whole shows such a correlation.

3.3.8 Extended Speech Act Theory

One desideratum of classical speech act theory is that it doesn't consider sufficiently enough the conversational context in which speech acts are performed. Many speech acts can be uttered only in certain positions within a conversation. We will follow the distinction of Rolf (1983) between **initiative**, **reactive** and **reinitiative** speech act types. Initiative speech acts can be uttered independent of previous utterances; reactive speech acts can only be uttered in response to a previous speech act (by someone else); and reinitiative speech acts can only be uttered in response to a reactive speech act (see Kohl & Kranz 1992: 10-11). Typical pairs of initiative and reactive speech acts in aviation are **command** and (verbal) **comply** (see also command and control discourse in Goguen & Linde 1983: 39-46), as well as **statement** and **confirmation**. Other reactive types are problematic ones such as **expressing_doubt** and rejective ones such as **refuting**.

The context-sensitive typology of speech acts introduces a level that goes beyond the single speech act. A systematic way of representing the structure of dialogue, in terms of an extended speech act theory, has been developed by Franke (1990). His approach attempts to account for the sequential embedding of speech acts in the communicative context. The basic unit on this level is the **minimal dialogue**. These minimal dialogues are then classified as a certain dialogue type, depending on the initial speech act.

The main idea of Franke's approach is as follows. With an initial speech act (ISA), speaker 1 (S1) states a communicative goal. Speaker 2 (S2) can accept or reject that goal. If, for example, S2's reaction did not constitute a minimal dialogue (that is, if S2 was not accepting S1's communicative goal, e.g. did not answer a question), S1 has three different options in the third move: (i) **A retractive speech act**: S1 gives up his original communicative goal; (ii) A **revised speech act**: S1 modifies his original communicative goal; (iii) a **re-initiative speech act**: S1 insists on his original goal and possibly attempts to reach it by different means. In the case of (i), the speech act constitutes a minimal dialogue. In the case of (ii) or (iii), the dialogue continues until one of the speakers has achieved his or her communicative goal or until it is clear that the other speaker does not accept it. Complex dialogues consist of a number of minimal dialogues.

This type of analysis, which takes into consideration the structure of minimal dialogues, allows for the quantitative analysis of aspects of cockpit communication that could not be described in terms of traditional speech act theory. It will enable us to test hypotheses that involve interesting aspects of the crew members' communicative behavior. For example, one could test how often crew members initiate dialogues, how often they give up their goals in a dialogue, how often they respond with a counter-initiative speech act, and how often (and perhaps also how quickly) they do (or do not) finish a dialogue, that is, constitute a minimal dialogue. In the end this, as in the first project phase, can then be related to crew performance and task load in order to find out whether there is a correlation.

3.4 Conversation Analysis (CA)

The extended speech act theory mentioned above should bridge the gap with another tradition of analysis of verbal interaction, Conversation Analysis. CA criticized Speech Act Theory as a tool for empirical research in a number of ways: For its tendency to isolate units without considering their context, for the fact that it is speaker-centered and does not consider the role of the addressee in an appropriate way, and for putting to much focus on the speaker's intentions: Asking questions like "what does the speaker really mean" to determine what a speaker really intended with an utterance seems to be a questionable approach when working with authentic material. Diegritz & Fürst therefore argue that a clear distinction needs to be made between the speech act as such, the illocution as speaker's intention and the communicative function of a speech act (see Diegritz & Fürst 1999:43f.). All of these criticisms are ultimately connected to speech act theory's origins in theoretical philosophy, where constructed utterances are at the center of attention. Applying a type system that is based on constructed utterances to authentic dialogue is very difficult, if not close to impossible.

Because of these points of criticism we will use conversation analysis in addition to (extended) speech act theory as a second, equivalent method. Conversation analysis, as opposed to speech act theory, has always been a strictly empirical approach. Its main purpose is the investigation of social interaction as a process of creation and stabilization of social order (Bergmann 2000:525), or the investigation of "interactional organization of social activities" (Hutchby & Wooffitt 2001:14). These authors pointed out that even the key role of talk in broader institutional processes "takes the essential CA starting point that talk-in-interaction is to be seen as its own social process, governed by its own regularities." (Hutchby & Wooffitt 2001:21). This is, of course, also valid for talk-in-interaction in the cockpit.

Due to its roots in ethnomethodology, CA is concerned with the reconstruction of data segments and not with psychological interpretations. Conversation or, respectively, talk-in-interaction is understood as a process, as a place in which social facts are produced. This process is always integrated into interaction – each single utterance is produced and modified by both interlocutors together. The speakers constitute their turns in interaction with the addressee.

If we want to investigate cockpit communication, this means that we will have to reconstruct the constitutive principles and mechanisms in the process of cockpit interaction during the simulator flight by analyzing it systematically, thereby exploring the structural principles which the individual crews follow. The social relationship between the crew members and the social order within the crew are important factors. We want to find out whether these structural principles are modified in relation to different types of task load and within crews and also in how far different structural principles of crews affect the interaction (and presumably the performance) in high task load situations. The goal is to reconstruct the practical methods the crews apply in solving interactional problems. The use of these methods generates the observable order within the interactional process in the cockpit. In other words: We do not describe homogeneous forms of behavior, but principles that represent real points of orientation for the acting individuals.

A basic interest of CA is the question, how it is possible for interactional participants to understand each other and to ensure that they do. This is an essential issue, especially in highly dangerous situations. Another topic concerns the organization of turn-takes and, relative to that, the activation of repair organization by the participants in a conversation, in cases where a problem of understanding occurs (see Schegloff, Jefferson & Sacks 1977). Consequently, the investigation of uptake and repair organization as a substantial factor in the turn-taking-system as well as of concepts of face and politeness (see Brown & Levinson 1987 and also Linde 1988 and Goguen & Linde 1983: 28-36 from a related speech act theoretic view) will be central to our analysis. After all, interaction in the cockpit takes place in a hierarchical setting and involves the creation and maintenance of social order, where these concepts play a rather significant role

Evidently cockpit communication differs from everyday conversation, as other forms of institutional communication, or rather: talk-in-interaction, do as well. Have (1999) writes about this difference „that for some institutional systems, there is a pre-established system of **turn allocation**, and quite often turn-**type** allocation" (Have 1999:163). The institutional form can be seen as "more 'restricted'" than everyday conversation. The asymmetrical distribution of questions and answers might be an interesting example for that (see Have 1999:164). Nevertheless, in institutional as well as in everyday talk-in-interaction the same principles are applied by the participants, even though the distribution may be different.

For our investigation of cockpit communication this means to reconstruct the process of cockpit interaction during the simulator flight through a comparative-systematic analysis, especially in respect to the social relationship, i.e. the social order within the crew, and to determine which structural principles are applied by the crew in which ways. What is to be found out is (i) whether the structural principles vary in relation to the type of task load (and if they do, how), and (ii) whether differences can be determined in terms of how different structural principles among crews influence the interaction in situations with different types of task load. The idea is to reconstruct the "practical methods" (Bergmann), which serve the crew as a solution for interactive problems. It is the use of those practical methods that generates the observable order in the cockpit-interaction. In other words: The point is not a "description of uniformities of behavior" but rather "in which way the participants themselves consider these formal principles in their utterances and actions" (Bergmann 2000: 533, translation by the authors). For the concrete goals of our investigation this means: Which variations are possible in the different types of task load, i.e. how and to what extent does the crew's behavior vary, and is there a particularly efficient form of communicative behavior or a successful strategy that could be determined, which would enable crews to solve aviation problems in a more result-oriented and therefore more successful way?

4 Conclusion

In this article, we reported some of the findings in a number of studies concerned with communication in aviation. There is evidence that linguistic features of communication correlate with the performance of crews and with the level of task load. We reported, in particular, results of a study based on flight simulator sessions done in the USA on a B-727, and the design and analysis of a new study, also based on flight simulator sessions done in Germany on a Canadair Regional Jet that promises to yield more definite results.

5 Literature

Bergmann, J. (2000): Konversationsanalyse. In: Flick, U. & E. von Kardorff & I. Steinke, Hrsg. (2000), 524 – 537.

Brown, P. & S. Levinson (1987): Politeness. Some Universals in Language Use. Cambridge: University Press.

Chidester, T.R., Kanki, G.G., Foushee, H.C., Dickinson, C.L. & Bowles, S.V. (1990). Personality factors in flight operations. Volume 1. Leader characteristics and crew performance in a full-mission air transport simulation. (NASA Technical Memorandum 102259). Moffett Field, CA: NASA-Ames Research Center.

Diegritz, T. & C. Fürst (1999): Empirische Sprechhandlungsforschung. Ansätze zur Analyse und Typisierung authentischer Äußerungen. Erlangen: Univ.-Bilbiothek.

Diegritz, T. & H. Rosenbusch (1995): Die pragmatisch-dynamische Methodenkombination (PDMK) zur Erforschung von Kommunikationsprozessen. In: König, E. & P. Zedler, Hrsg. (1995), 435 – 461.

Flick, U. & E. von Kardorff & I. Steinke, Hrsg. (2000): Qualitative Forschung. Ein Handbuch. Reinbek bei Hamburg: Rowohlt.

Franke, W.(1990): Elementare Dialogstrukturen. Tübingen: Niemeyer.

Fritz, G. & F. Hundsnurscher, Hrsg. (1994): Handbuch der Dialoganalyse. Tübingen: Niemeyer.

Goguen, J.A. & C. Linde (1983): Linguistic Methodology for the Analysis of Aviation Accidents. Technical Report, Structural Semantics. Nasa Contractor Report 3741, Nasa Ames Research Center. National Aeronautics and Space Administration. Palo Alto, California.

Graffe, J. (1990): Sich festlegen und verpflichten: Die Untertypen kommissiver Sprechakte und ihre sprachlichen Realisierungsformen. Münster: Waxmann.

Have, P. ten (1999): Doing Conversation Analysis: A Practical Guide. London: Sage.

Helmreich, R. L. (1997a): Managing human error in aviation. *Scientific American*, May issue, 62-97.

Helmreich, R. L. (1997b): Kommunikationstraining gegen menschliches Versagen im Cockpit. *Spektrum der Wissenschaft*, July issue.

Helmreich, R. L., Merritt, A.C. & Wilhelm, J.A. (1999): The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology* 9, 283-301.

Hindelang, G. (1978): Auffordern. Die Untertypen des Aufforderns und ihre sprachlichen Realisierungsformen. Göppingen: Kümmerle.

Hindelang, G. (1981): Zur Klassifikation der Fragehandlungen. In: Hindelang, G. & W. Zillig, Hrsg. (1981), 215 – 225.

Hindelang, G. (1994): Sprechakttheoretische Dialoganalyse. In: Fritz, G. & F. Hundsnurscher, Hrsg. (1994), 95 – 112.

Hindelang, G. (1995): Frageklassifikation und Dialoganalyse. In: G. Hindelang & E. Rolf & W. Zillig, Hrsg. (1995), 176 – 196.

Hindelang, G. & W. Zillig, Hrsg. (1981): Sprache: Verstehen und Handeln. Tübingen: Niemeyer.

Hindelang, G. & E. Rolf & W. Zillig, Hrsg. (1995): Der Gebrauch der Sprache. Festschrift für Franz Hundsnurscher zum 60. Geburtstag. Münster: LIT.

Hutchby, I. & R. Wooffitt (2001): Conversation Analysis. Principles, Practices and Applications. 3rd ed. Cambridge: Polity Press.

Kohl, M. & B. Kranz (1992): Untermuster globaler Typen illokutionärer Akte. Zur Untergliederung von Sprechaktklassen und ihrer Beschreibung. In: P.-P. König & H. Wiegers, Hrsg. (1992), 1 – 44.

König, E. & P. Zedler, Hrsg. (1995): Bilanz qualitativer Forschung. Band II: Methoden. Weinheim: Deutscher Studien Verlag.

König, P.-P. & H. Wiegers, Hrsg. (1992): Sprechakttheorie. Münster: LIT.

Linde, C. (1988): "The quantitative study of communicative success: Politeness and accidents in aviation discourse." *Language in Society* 17.3, 375 – 399.

Nevile, M. (2001): Knowing who's who in the airline cockpit: pilot's pronominal choices and cockpit roles. In A. McHoul & M. Rapley (2001).

Marten-Cleef, S. (1991): Gefühle ausdrücken. Die expressiven Sprechakte. Göppingen: Kümmerle.

McHoul, A. & Rapley, M. (2001): How to analyze talk in institutional settings. A casebook of methods. London & New York: Continuum.

Pennebaker, J.W. & Francis, M.E. (1999): Linguistic Inquiry and Word Count (LIWC). Mahmah, N.J.: Erlbaum Publishing.

Rolf, E. (1983): Sprachliche Informationshandlungen. Göppingen: Kümmerle.

Schegloff, E.A. & G. Jefferson & H. Sacks (1977): "The preference for self-correction in the organization of repair in conversation." Language 53, 361 – 382.

Searle, J. (1975): A classification of illocutionary acts. *Language in Society* 5: 1-23.

Sexton, J.B. & Helmreich, R.L. (2000). Analyzing cockpit communication. The links between language, performance, error, and workload. *Human Performance in Extreme Environments* 5, 63-68.

Silverstein, D. (2001). Final Report of the Subproject "Initiating Crew Resources under High Cognitive Work Load", GIHRE project, http://www2.rz.hu-berlin.de/GIHRE/page/engl.html

Zillig, W. (1982): Bewerten: Sprechakttypen der bewertenden Rede. Tübingen: Niemeyer.