(NASA-CR-3741) LINGUISTIC METHODOLOGY FOR THE ANALYSIS OF AVIATION ACCIDENTS (Structural Semantics) 121 p HC A06/MF A01 CSCL 01C

N84-15135

Unclas

H1/03 11361

NASA Contractor Report 3741

Linguistic Methodology for the Analysis of Aviation Accidents

J. A. Goguen and C. Linde
Structural Semantics
Palo Alto, California

Prepared for
Ames Research Center
under Contract NAS2-11052

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1983

Table of Contents

List of Figures

ABSTRACT

This research develops a linguistic methodology for the analysis of small group discourse, and demonstrates the use of this methodology on transcripts of commercial air transport accidents. The methodology first identifies the discourse types that occur (these include planning, explanation, and command and control) and determines their linguistic structure; it then identifies significant linguistic variables based upon these structures or other linguistic concepts such as speech act and topic; next, it tests hypotheses that support the significance and reliability of these variables; and finally, it indicates the implications of the validated hypotheses. These implications fall into three categories: (1) training crews to use more nearly optimal communication patterns; (2) using linguistic variables as indices for aspects of crew performance such as attention; and (3) providing guidelines for the design of aviation procedures and equipment, especially those that involve speech.

1 EXECUTIVE SUMMARY

This section provides a non-technical summary of the entire report that follows. Further detail is available in the corresponding section of the report body; an Index and Glossary are provided in Appendix II.

1.1 Introduction

The basic motivation for the research reported here is to reduce the incidence of those air transport accidents caused wholly or in part by problems in crew communication and coordination. One important way to do this is to train crews to communicate more effectively. A major objective of this research is to determine those communication patterns which actually are most effective in specific situations; this requires developing methods for assessing the effectiveness of crew communication patterns. A second objective is to develop linguistic measures for assessing other aspects of crew performance, such as attention, fatigue, etc. A third objective is to provide guidelines for the design of aviation procedures and equipment, for example new technology permitting computer-generated verbal communication.

The main contribution of this study is a methodology to achieve these objectives and others of a similar nature. This methodology involves the following stages:

1. The research begins with a detailed investigation of how crews *actually* talk, yielding an empirically grounded formal description of communication patterns in the cockpit. Formal theories of the discourse types involved in air crew communication constitute a major part of the description; other linguistic concepts, such as speech act and topic, form additional parts. The present study is based on an investigation of eight aviation accident transcripts.

2. Variables based upon these theories are isolated, and in some cases tested for reliability. A number of such variables are discussed in this report.

3. Research hypotheses about normal crew communication and about the causes of communication failure are formulated using variables from the previous stage of research. They are then tested. Formulating hypotheses on one subset of a large sample and testing them on a disjoint subset reduces the likelihood of bias from the idiosyncratic nature of particular transcripts, and supports the view that the results are applicable to the larger population of all commercial air transport discourse; further arguments that the reults may generalize are given in Section 9. The reader who does not accept these arguments may instead regard the statistical results as descriptive summaries of a particular sample. This stage of research is not complete, and will be continued using flight simulator data. This is necessary because accident transcript data permits only limited correlations, between two linguistic variables, or between a linguistic variable and the gross performance data furnished by the NTSB reports. Flight simulator data will make it possible to test the current hypotheses more accurately, and also to test additional hypotheses, since this data will provide both repeated instances of the same situation, as well as detailed and accurate performance, behavioral, and systems data. In particular,

hypotheses about correlations of linguistic variables with crew and system performance variables can be tested.

4. In the fourth stage, the validated hypotheses on crew communication patterns are used in formulating proposals for crew training; these proposals can then be tested with flight simulation experiments. Applications to the evaluation of other research hypotheses are also possible, by using the linguistic variables as relatively inexpensive measures for aspects of the quality of crew performance. There are also applications to the design of aviation procedures and equipment.

1.2 Theory Creation and Adaptation

In order to provide an adequate description of cockpit communication, we have created or adapted a number of linguistic theories. These include: speech act theory, and formal theories for the discourse types of planning, explanation, and command and control. These theories support the linguistic variables used in hypotheses of the next phase. The variables include: mitigation/aggravation level, crew recognized emergency, crew recognized problem, operational relevance, and topic success or failure. We turn first to a brief discussion of the linguistic theories.

1.2.1 Speech Act Theory

Speech act theory, now well established in linguistics and the philosophy of language, focusses on the operational aspect of language -- how a particular sentence achieves some effect in the world. We call this the **social force** of the speech act. The fundamental insight of speech act theory is that some sentences, such as (1), *describe* or *report* a state of the world, while other sentences, such as (2), *create* a state of the world.

(1) **There's a thunderstorm ahead.**

(2) **I declare this bridge open.**

Speech acts may be either direct or indirect. Direct speech acts either use an unambiguous syntactic form to achieve their effect, as in (1), or explicitly name their own function, as in (2). Indirect speech acts like (3) and (4) place a greater interpretive burden on their addressee, forcing him to infer what effect the speaker wishes to accomplish.

(3) **What I need is the wind, really.**

(4) **Can you get that checklist?**

In these examples, the speech act of ordering is indirectly expresed by the form of a statement of need and a question about ability.

Speech act theory also provides a taxonomy of possible types of speech act. We have modified this taxonomy to provide an inclusive listing of the speech acts found in cockpit communication. These are: **Requests**, including orders, requests, suggestions and questions; **Reports**; **Declarations**; and **Acknowledgements**.

We have also provided several tests for determining whether a given speech act has actually succeeded in accomplishing its intended effect in the world. This is important because it furnishes a tool for measuring to what degree a given communication pattern functions effectively.

1.2.2 Command and Control Discourse

A speech act is a single sentence or turn produced by one speaker, and so is an atomic unit of social meaning. In order to understand the larger patterns of communication characteristic of command and control, it is necessary to move from the level of speech acts to the level of sequences of speech acts. We call such sequences **speech act chains**. A speech act chain consists of sequences of speech acts, and may also include the discourse types which are characteristic of operationally relevant cockpit communication -- planning and explanation.

Because the command and control chain is a well structured discourse type, it is possible to describe it with a formal grammar. Such a grammar defines correct and incorrect sequences of speech acts and embedded discourse types. This is valuable because it allows us to judge whether a given segment of talk follows the rules for command and control discourse, or whether it is deviant in some way. We hypothesize that correct command and control chains are the optimal pattern of communication in the cockpit, particularly in emergency situations. This hypothesis can not be tested directly on the present data, but it can be tested with simulator experiments.

1.2.3 Planning and Explanation

In this research, we focus on planning and explanation as linguistic activities, carried on by a group, rather than as individual mental activities. Planning and explanation are important in cockpit communication since they are one of the major means by which a group can solve novel problems. It is possible to give a formal grammar describing the discourse types of planning and explanation. This report extends our previous description of these discourse types [Linde & Goguen 78].

In addition to their importance in problem solving, planning and explanation form an important part of the process whereby a suggestion by a crew member is ratified by the captain, and becomes, in effect, an order issued by the captain. We call this process **ratification**. Such suggestions are frequently made as part of a plan. It is possible to make an addition to the formal grammar describing the various ways in which the captan can accomplish ratification.

1.3 Linguistic Variables Arising from These Theories

Using these linguistic theories, it is possible to define a number of variables used in our hypotheses about crew behavior patterns.

1.3.1 Crew Recognized Emergency

A **Crew Recognized Emergency** is a situation in which the entire crew attends to the situation or situations that caused the accident (using the NTSB determination of the cause of the accident.) Note that this variable does not indicate the *actual* onset of the problem. but rather the point at which the crew *recognizes* it as a problem.

This variable is required because we hypothesize that linguistic behavior differs when crew members know that they are facing an emergency situation. This definition allows us to test hypotheses of this form.

1.3.2 Crew Recognized Problem

A **Crew Recognized Problem** is similar to a Crew Recognized Emergency, but is less intense; it is a situation which the crew recognizes as potentially dangerous and not a normal part of flight operations. Like the Crew Recognized Emergency, this variable allows us to test hypotheses postulating differences in linguistic patterns during problem situations.

1.3.3 Operational Relevance

A distinction entering into many definitions and hypotheses is whether some utterance or discourse unit is **operationally relevant**. An operationally relevant utterance is one which is directly involved with the achievement of successful mission completion. This definition permits us to focus directly on the language of interest, and to exclude irrelevant remarks in a principled way.

1.3.4 Mitigation/Aggravation

The variable of **mitigation/aggravation** is necessary for this research because it provides one dimension for assessing the assertiveness of speech acts. Any utterance may be ranked on a scale of mitigation/aggravation. This corresponds to the degree of politeness or indirectness of the utterance. Thus, (5) is direct, (6) is mitigated, (7) is highly mitigated, and (8) is aggravated.

(5) **Close the window.**

(6) **Would you close the window?**

(7) **Please, would you mind closing the window?**

(8) **Listen, close that damn window right now.**

Mitigation softens the possible offense that an utterance might give. It is important for cockpit communication because we have found that the greater the degree of mitigation, the more likely it is that a given utterance will fail to accomplish its effect. In addition, we have found that speech acts by subordinates are more mitigated than those of superiors. A number of NTSB reports have noted that even when subordinates make correct suggestions in problem or emergency situations, these suggestions may not be accepted. The NTSB has suggested assertiveness training as a possible remedy. The present analysis of mitigation shows in detail some aspects of the nature of linguistic assertiveness and non-assertiveness. Moreover, these aspects seem to be particularly amenable to measurement and to crew training.

1.3.5 Topic and Topic Failure

A precise definition of topic is necessary to investigate why crew members sometimes fail to recognize or continue newly proposed topics, often topics of great operational importance. **Topic** is defined as the propositional content of a speech act. The propositional content is what the sentence predicates about the world, what the sentence is about, independent of its social force. Thus, (9), (10), and (11) have different social force but the same propositional content.

(9) Close the window.

(10) The window is closed.

(11) Is the window closed?

Using this definition, we have been able to define precisely instances of topic failure, and have also given a taxonomy of the major topics found in our sample of aviation discourse.

1.4 The Formulation and Validation of Hypotheses

The linguistic theories discussed above have been used as the basis for a number of hypotheses about the linguistic structure of cockpit communication and its relation to successful flight operations. Section 9 discusses the statistical issues involved in testing hypotheses on data like that of the present study, and then reports the results of these tests.

The hypotheses tested have two classes of implication. The first concern the basic structure of cockpit communication,including relations between variables of operational structure, social structure and linguistic structure, and hence, represent a basic test of the theory developed in this report. The second class of implications concern applications such as training.

The following subsections discuss the eight hypotheses in detail. In summary, these tests support our theory of cockpit communication, suggesting the essential correctness of the general reseach direction, and also suggesting the value of further research using suitable data for testing correlations between linguistic variables and variables of crew and system performance.

1.4.1 Speech Acts to Superiors Are More Mitigated

The first hypothesis states that the speech of subordinates is more tentative and indirect than the speech of superiors. This hypothesis has been accepted. It is important because it shows that there is a relation between the social hierarchy and the form of cockpit discourse, and it provides a foundation for later hypotheses that excessive mitigation is related to failure of proposed topics and suggestions.

1.4.2 Speech Acts Are Less Mitigated in Crew Recognized Emergencies

This hypothesis states that when crew members (including the captain) know that they are in a emergency situation, their speech is less tentative and indirect. This hypothesis has been accepted. It is important because it shows that crew members are able to vary their use of mitigation depending on their perception of the situation. This suggests both that experienced crews feel that mitigation is inapropriate in an emergency and that the level of mitigation used should be trainable.

1.4.3 Speech Acts Are Less Mitigated in Crew Recognized Problems

This hypothesis is similar to the previous one, stating that when crew members know that they are in a problem situation, their speech is less tentative and indirect. It has been accepted. Its significance is similar to that of the previous hypothesis.

1.4.4 Subordinates Plan and Explain More Often Than Superiors

This hypothesis tests, in an indirect way, possible inhibitory effects of the social hierarchy on contributions by subordinates. Rejection of this hypothesis would suggest that subordinates do not contribute as fully as superiors, because of their position in the social hierarchy. The test results show that not only do subordinates *not* plan and explain more than superiors, but they suggest that actually superiors may plan and explain *more* than subordinates. This result is interesting because modern management theory generally asserts that a group is more effective when subordinates contribute freely, perhaps more than superiors. It might be valuable to determine whether crew performance is improved by training subordinates to do more planning and explanation, and training captains to encourage this.

1.4.5 Planning and Explanation Are Less Common in Crew Recognized Emergencies

This hypothesis represents the intuition that when crew members know that they face an emergency situation, they will do less planning and explaining of possible courses of action, since an emergency calls for immediate action. This hypothesis has been accepted. It is possible that more planning and explanation would be desirable for successful mission completion in emergency situations. Testing the present hypothesis on data from successful flights would permit us to determine the optimal level of planning and explantion in CRE.

1.4.6 Planning and Explanation Are More Common in Crew Recognized Problems

This hypothesis states that when crew members are aware that they are in a problem situation, they do more planning and exlaining. This hypothesis has been accepted. This result is interesting because it shows that crew members do indeed reserve planning and explanation for appropriate situations, those in which the standard flight plan is no longer adequate.

1.4.7 Topic Failed Speech Acts Are More Mitigated

This hypothesis tests the idea that excessive mitigation can lead to undesirable consequences, specifically that a new topic is less likely to be picked up by other crew members if the speech act in which it is introduced is excessively mitigated. This hypothesis has been accepted. It is important because it suggests that the frequent situation of a subordinate failing to get a correct point accepted might be improved by training in linguistic directness.

1.4.8 Unratified Draft Orders Are More Mitigated.

This hypothesis tests the idea that when a crew member proposes a suggestion to the captain, the more indirect and tentative the suggestion, the less likely the captain is to ratify it. This hypothesis has been accepted. Like the preceding hypothesis, it is important because it suggests the possible value of training in linguistic directness.

1.5 Directions for Future Research

The present research suggests both immediate directions for future research and also possible practical applications of the entire research program. This subsection discusses first possible measures of crew performance arising from this research, and then some more speculative possibilities for improving crew performance.

1.5.1 Linguistic Measures of Crew Performance

One application of the present description of cockpit communication is the development of linguistic measures which correlate with performance or behavioral measures. This would be of particular interest in simulator studies where, it is hoped, linguistic measures could give a an earlier and more sensitive indication of degradation of crew performance than current behavioral measures. That is, current measures can only indicate actual crew errors, while linguistic measures might indicate earlier conditions tending toward impaired vigilance, co-ordination, etc. In some cases, the linguistic measures might also be less expensive.

One such measure that we have already developed, but not yet tested, is **degree of command and control coherence**. This variable attempts to formalize the intuition that it is possible to judge the degree to which a given sequence of utterances is well-integrated, tightly structured, and facilitates optimal crew communication. In such a well-integrated sequence, a request or report is followed by an acknowledgement, support, challenge, or request. No request or report is left without acknowledgement or comment. Such a pattern allows a crew member to know that his utterance has been heard and attended to.

This variable is directly based on the rules for speech act chains, giving a social interpretation to these formal rules for the sequencing of speech acts. Degree of command and control coherence can be computed most simply for any segment of text as the ratio of the number of command and control speech acts to the total number of speech acts. (A command and control speech act is one which forms part of a valid command and control chain; a non-command and control speech act is one which is part of any other discourse type, or which is isolated and does not form a part of any larger unit.)

The value of this variable is suggested by previous work [Foushee & Manos 81] showing that use of a greater number of commands and acknowledgements is correlated with mission success. The definition of the linguistic form of a proper command and control sequence makes this finding more sensitive, and hence, we believe, more useful. Command and control coherence should function as a linguistic correlate of resource management, attention, and vigilance, and should be valuable as an early warning sign of deterioration of these factors.

The command and control coherence variable may be viewed as a model for the form of linguistic variables and their potential correlation to problems of crew coordination and resource management. Other variables of this sort suggested by the present research include: rate of planning and explanation in Crew Recognized Problem and Crew Recognized Emergency situations, number of requests with a high number of possible interpretations, use of explanation in constructing false hypotheses about the nature of a problem situation, number of request-report-acknowledgement triples, etc. These variables, and others that are similar, could be validated with flight simulator data.

1.5.2 More Speculative Research Directions

Although further validation is necessary to allow the current theoretical and methodological framework to serve as a basis for training recommendations and other applications, it is possble even at this stage to suggest some directions for applications.

One training method would be to use films or videotapes illustrating the effects of certain patterns of communication on crew coordination and decision making. This approach could include the use of peer commentary in the training material.

More speculatively, it might be possble to design new speech acts having formal command and control status, in order to ameliorate particular communiation problems. For example, a formal **challenge** speech act might be created, which would be addressed by a subordinate to the captain, and which the captain would be legally obligated to acknowledge.

Moving further into the future, cockpit automation may well procede to the point where the system gives complex verbal information to the crew. If so, it would be desirable to have the speech of the system as similar as possible to the linguistic forms used by the crew. In particular, proper formulation of explanations would be particularly important in promoting effective crew utilization of on-board diagnostic systems, as experience with similar systems for medical diagnosis has shown [Swartout 81].

1.6 Conclusions

Based on this work, it may be concluded that a methodology is now available for the detailed analysis of cockpit discourse that can be applied to improving aviation safety. This methodology has produced a description of cockpit communication which has served as a basis for hypotheses about the linguistic behavior of crews. It has also been used to formulate a number of variables that might serve as indicators for various aspects of air crew performance such as vigilance and crew coordination, and to formulate a number of training suggestions for air crew communication.

In support of this methodology, the statistical hypotheses, while far from comprehensive, provide convincing evidence that the variables isolated are reliable and valid, and have powerful relations with one another and with the general structure of cockpit activity. There is also suggestive evidence even at the present stage of research that they may have powerful relations with crew and system performance levels. The important role of the linguistic variable of mitigation has been demonstrated, showing its correlation with a number of basic structural and crew coordination factors such as rank, topic failure, and draft order ratification.

The following subsections describe in detail the major contributions of this work.

1.6.1 Basic Contributions

1. A classfication of the discourse types occurring in cockpit communication:. command and control chain, checklist (a subtype of command and control chain), planning, explanation, narrative, and pseudo-narrative.

2. A theory of the structure of command and control chains.

3. A general theory of the structure of discourse, and a formalism for expressing it.

4. A scale of mitigation levels for speech acts in aviation discourse, and an experimental validation of this scale.

5. An empirically based theory of speech act misinterpretation.

6. A theory of draft orders and the process by which they are ratified.

7. A collection of variables summarizing various important characteristics of speech acts in cockpit communication.

8. A set of computational tools for testing statistical hypotheses, including LISP programs for checking the consistency of coded data sets, extracting relevant data, and performing the requisite statistical calculations.

1.6.2 Applied and Specific Contributions

This subsection describes the most important specific contributions of this research. Note that these contributions are limited by the nature of the present sample; future research using this methodology on simulator data should clarify many questions left open here.

1. It has been shown that the average mitigation level of requests by subordinates is higher than that of requests by superiors. It is hypothesized that this this assymmetry contributes to captain's misunderstandings of suggestions by subordinates.

2. It has been shown that there are significant regional differences in the interpretation of mitigation. Further research might determine whether or not this is a contributing factor to the misinterpretation of cockpit speech acts. Thi would indicate if it would be worthwhile training crews to recognize and compensate for these regional differences.

3. It has been shown that requests are less mitigated during Crew Recognized Problems, and still less mitigated during Crew Recognized Emergencies. This shows that crew members are able to vary their mitigation level depending on their perception of the situation, and hence suggests that mitigation level is trainable. It also supports the suggestion that such training might be valuable.

4. It has been shown that subordinates do not produce more planning and explanation than superiors. Further research is required to determine what the optimal ratio might be.

5. It has been shown that planning and explanation are more common during Crew Recognized Problems but not during Crew Recognized Emergencies. This suggests further research into the optimal levels of planning and explanation in both CRP and CRE.

6. It has been shown that more mitigated speech acts are more likely to have their topics fail: This demonstrates the importance of crew members using direct language to introduce operationally significant topics.

7. It has been shown that more mitigated draft orders are less likely to be ratified. This also demonstrates the importance of using direct language.

This research suggests the value of investigating the correlation of a number of other linguistic variables with system and crew performance variables. These include degree of command and control coherence, rate of request-report-acknowledgement triples, rate of planning and explanation, and rate of simple acknowledgements. Such correlations might be less costly indicators of objective performance measures, and might also have training implications.

Finally, this research should have many applications to the design of aviation procedures and equipment involving the use of language. Any equipment developed for the cockpit producing audio output, particularly complex linguistic output, should produce it in a natural way in order to ensure optimal utilization by the crew. The present research could serve as the basis for the design of such equipment.

2 INTRODUCTION

This section discusses the background and motivation for this research and the general applicability of results obtained from the data that was used. This section also contains the notational conventions used and acknowldgements for contributions to our research. It should be noted that the present study reports an entirely new theoretical approach to the issue of aviation safety. For this reason, the research is described in considerable detail, and the report provides theoretical background in several fields.

2.1 Background for This Research

The basic motivation for this research program is to reduce the incidence of air transport accidents. To this end, we are developing measures of the quality of crew coordination, and formulating suggestions for training procedures to improve crew coordination. Such measures involve interpersonal factors, and hence, linguistic factors. In support of this program, the present study provides a methodology for studying the language of the cockpit, including a theoretical framework, a number of linguistic variables, and tests of some hypotheses involving these variables. This study has used a data base of air transport accident transcripts in which crew coordination problems appear to have been a major causative factor.

Three previous NASA studies provide a motivation and foundation for the present research. [Ruffell Smith 79] identified management of resources, both human and material, as a major factor influencing the effectiveness and safety of crew operations, using B-747 full mission simulation studies. Frequent problems in communication, decision making, crew interaction, and crew integration were noted in this data. [Murphy 80] examined eighty four commercial aviation incident reports (collected through NASA's Aviation Safety Reporting System (ASRS)) from a resource management perspective, and found interpersonal communications with Air Traffic Control (ATC), task management, planning, coordination, and decision making to be major areas for concern. [Foushee & Manos 81] studied CVRs from the [Ruffell Smith 79] data and concluded that cockpit communication patterns are closely related to flight crew performance. A number of essentially linguistic factors (such as rates of commands and of acknowledgements in a text) were found to correlate strongly with various performance measures.

Basic theoretical work forms the largest part of this report, and should also be of value for other research on interpersonal factors in aviation, because it permits a more detailed and precise understanding of the mechanisms of interaction. It could, for example, be useful in designing other research programs that use CVR transcripts, or that use audio or video transcripts of flight simulator sessions, or that consider other hypotheses about crew performance involving variables similar to those in this study. For example, this work should be useful in studies of crew fatigue during extended missions, and in studies of air to ground communication. Possible applications are discussed in more detail in Section 10.

2.2 Applicability of This Research

As stated above, this research attempts to provide a methodology that can be used to study any form of data on aviation communication, including transcripts from CVR recorders and audio or video records of simulator sessions. However, it is also important to note certain restrictions on the applicability of this research.

Because transcripts are available only for flights that ended in an accident, there is no control data on the nature of communication for successful flights, and most importantly, for flights in which some problem arose and was dealt with successfully. Similarly, because of the absence of video records, there are many cases where it is impossible to tell what actually happened. For example, in a situation in which the captain gives an order and does not receive a verbal reply, it is not possible to tell whether he was answered with a nod.

These restrictions on the data limit the nature of the hypotheses that can be tested about correlations between linguistic phenomena and performance phenomena. In later studies, using data from flight simulators, it should be possible to remedy this lack. However, these restrictions on the data should have no effect on the basic form of the theory, which is intended as a formal description of aviation discourse. Additional data may motivate additions to the theory, but should not necessitate any fundamental changes to the theory.

2.3 Notational Conventions

The notation used in the official NTSB transcripts is neither entirely consistent nor entirely suitable for the purposes of the present report. In this report, the following conventions are used:

1. NTSB transcript citations are given in the form "airline/crash site/year," followed by the time in parentheses. However, since many examples used in this study are taken from United/Portland/78, citations from this transcript are abbreviated to just the time. (This transcript is used as a major source of examples because of its relevance to the purpose of this project and because of its familiarity to the aviation community.)

2. Individual turns of speakers are identified as to source and speaker. CAM indicates that the source was the cockpit microphone; RDO indicates a radio transmission. The following numbers are used for speakers: 1 = captain, 2 = copilot, 3 = flight engineer, 4 = third officer, 5 = jump seat occupant, 6 = head flight attendant, 7 = other flight attendant.

3. * indicates the omission of untranscribable material.

4. # indicates what the NTSB calls a "non-pertinent word;" usually these appear to be obscenity or profanity.

5. Parentheses indicate a word not completely clear to the transcriber.

It should be noted that the transcripts contain many imperfections. For example, at (approximately) 1751:29 and 1754:23 of the United/Portland/78 transcript, the word `will` appears where it evidently should be `we'll`. Also, attribution of speaker and punctuation is inconsistent and sometimes confusing. Nevertheless, in all cases, the NTSB transcription is used, since it has not been possible to compare the transcripts with the actual tapes.

2.4 Acknowledgements

We would very much like to thank Miles Murphy of NASA, Ames Research Center for his help in getting this project started and keeping it focused. We also wish to thank Renwick Curry, Clayton Foushee, Al Lee, John Lauber, Robert Randle and Trieve Tanner of NASA for their valuable comments, and our consultants Tora Bikson, Richard Frankel, George Lakoff, Michael Moerman and Captain John Raabe for their help with experimental methodology, speech act theory, conversational analysis, and aviation practice. Finally, we thank William Labov for his encouragement and inspiration.

PART I: THEORETICAL BACKGROUND

3 SPEECH ACTS

This section discusses speech act theory, one of the major theoretical tools used in this report to understand aviation accident transcripts. This section also indicates some modifications required to make speech act theory fully applicable to the present data.

3.1 Language and Social Force

It is possible to view any utterance from two perspectives -- the perspective of **language**, focussing on its linguistic form, and the perspective of **social force**, focussing on its effect in the world. Investigations at the level of language are concerned with the form of what is actually said, using the precise tools furnished by linguistics. Investigations at the level of social force are concerned with what an utterance accomplishes or fails to accomplish. The level of social force is of great importance in the present study, allowing us to ask such questions as what linguistic units were taken as orders and carried out, what explanations led to a resolution of a problem, what proposed actions were lost and never discussed, etc. (These two levels have also been termed "what was said" and "what was done" [Labov & Fanshel 77].)

Since it is the level of social force that is clearly of the greatest relevance for this project, one might ask what value there is in studying the level of language. It is necessary to study both levels since the level of social force is derived from the level of language; we must understand the form of what was said before we can make the interpretation of what effect it had in the world.

3.2 Speech Act Theory

Speech act theory is the first theory of language which focusses in a systematic way on the level of social force. The fundamental insight of speech act theory is that certain utterances can be viewed as **performing** actions in the world [Austin 62, Searle 69]. For example, (12), (13), (14) and (15) can be seen as performing actions, rather than simply describing them.

(12) I christen this ship the Argos.

(13) I now pronounce you man and wife.

(14) I promise you I'll get to your party on time.

(15) I bet you five dollars the Yankees will lose.

Thus, (15) does not describe the act of betting, but rather performs it. For examples like these, the social force, or to be more precise, the probable or potential social force is obvious, since the verbs of the sentences themselves correspond to the social act being performed -- christening, promising, betting, etc. This is one way to accomplish a speech act directly. Another way is to match the social force of the sentence to its syntactic form -- expressing a directive with an imperative form or a request for information with a question form. Section 3.2.2 discusses the complex matter of indirect speech acts.

Recent discussions of speech act theory have broadened the scope of the notion of speech act, so that any utterance may be considered to be a speech act of some kind. Thus, an utterance such as

(16) **The sky is blue.**

may be considered to perform the speech act of asserting or informing. This is of great importance for the present study, since the act of reporting is an important and frequent speech act in the cockpit.

3.2.1 Propositional Content

Speech act theory permits us to separate the social force of an utterance from its propositional content. The **propositional content** of an utterance is some proposition which it makes about the world. Depending on the social force of the utterance, this propositional content may be reported, requested, denied, etc. Thus, the following examples have the same propositional content, but different social forces.

(17) **Let me inform you that the sky is overcast.**

(18) **I have to warn you that the sky is overcast**

(19) **The sky doesn't look overcast to me.**

(20) **I agree that the sky looks overcast.**

In these examples, the propositional content is **The sky is overcast**; the social forces are reporting, warning, challenging, and agreeing or acknowledging.

3.2.2 Indirect Speech Acts

Thus far, all the examples given have been speech acts which express their social force directly. However, there are also speech acts which express their most probable social force indirectly. These use a linguistic form which is not to be interpreted literally. For example:

(21) **CAM-1 What I need is the wind, really**
(1755:13)

This is literally an **expressive**, in Searle's terms, in which the captain expresses a psychological state of "needing" information about the wind. However, given the context in which it was spoken, its social force might be given as the directive

(22) **Give me the wind.**

Clearly, the use of the linguistic form of one speech act to convey the social force of another presents opportunities for misinterpretation that can have serious consequences in the cockpit situation.

The primary question for indirect speech acts is how it can happen that one speech act gets interpreted as another. To answer this, speech act theory uses **felicity conditions**, which are conditions that must be satisfied in order for a speech act of a given kind to be uttered "felicitously" (also termed "non-defectively"). These conditions include preparatory conditions, propositional content conditions, sincerity conditions, an essential condition, and possibly some others. **Preparatory conditions** cover what must be satisfied before the utterance is made; for example, for an order, that the speaker must have appropriate authority over the addressee, and that the addresee is able to perform the act; or for a promise, that it is not obvious that what is promised would otherwise occur. **Propositional content conditions** express constraints on the propositional content; for example, for a promise, that it express a future act by the speaker. **Sincerity conditions** concern the speaker's internal states, including his intentions. For example, in a request that the addressee perform an act A, the speaker should really want the addressee to do A. The **essential condition** defines the desired effect of the speech act upon the addressee.

The most obvious way to accomplish a speech act indirectly is to make reference to one of its felicity conditions. For example, one of the felicity conditions for a request that the addressee make a report is that the speaker should really want to know the contents of this report. This gives us an explanation of how (21) can indirectly convey (22).

Figure 1 gives a list of felicity conditions for directives, which include orders and requests; Figure 2 gives a list of "generalizations" for the indirect accomplishment of directives. Both figures are adapted from [Searle 79].

Preparatory:	**Addressee is able to perform act A**
Propositional Content:	**Speaker predicates a future act A of the addressee**
Sincerity:	**Speaker wants the addressee to do act A**
Essential:	**Utterance counts as an attempt by the speaker to get the addressee to do act A**

Figure 1: Felicity Conditions for Directives

There is a very large body of literature on indirect speech acts in the fields of linguistics, philosophy of language, artificial intelligence, and psychology. (See, for example, [Searle 79, Gordon & Lakoff 71, Gazdar 79, Labov & Fanshel 77].) The foregoing discussion is a summary of the approach of [Searle 79], which underlies most of these approaches.

---

1. Speaker can make an indirect directive to do act A either by asking whether a preparatory condition concerning the addressee's ability to do A holds, or by stating that it does hold.

2. Speaker can make an indirect directive by asking whether the propositional content condition holds, or by stating that it does hold.

3. Speaker can make an indirect directive by stating that the sincerity condition holds, but not by asking whether it holds.

4. Speaker can make an indirect directive to do act A either by stating that there are good or overriding reasons for doing A, or by asking whether such reasons exist, except where the reason is that the addressee wishes to do A, in which case the speaker can only ask whether the addressee wishes to do A, but can not assert that he does.

Figure 2: Strategies for Indirect Directives

3.3 The Success or Failure of Speech Acts

At the level of social force, the crucial question about any speech act is whether or not it has succeeded. This requires a definition of what success means. The account of success given in speech act theory is insufficient for the present project. This section first sketches this account, and then gives the broader definition of success needed for this research.

3.3.1 Success of Speech Acts within Speech Act Theory

Speech act theory uses the **linguistic form** of the speech act, without any external factors, to determine the effect of the speech act, that is, the effect it would have were it to be successful in the world. This is termed the **illocutionary force** of the speech act. The illocutionary force represents the speaker's intention, what he wishes to accomplish with his utterance [Searle 69, Searle 79]. [Searle 69] claims that "the syntactic structure of the sentence" which performs a speech act contains an "illocutionary force indicator" which "shows how the proposition is to be taken, or to put it another way, what illocutionary force the utterance is to have; that is, what illocutionary act the speaker is performing in the utterance of the sentence."

In order to determine whether the illocutionary force of some sentence succeeds, speech act theory moves beyond the form of the sentence to "felicity conditions" that involve non-linguistic factors such as the nature of the propositional content, the intention and abilities of the speaker, the desires of the addressee, etc. For example, (according to [Searle 71]) in order for a promise to succeed, the promised action must be one which the speaker is able to perform, intends to perform, and which is to the advantage of the addressee. Thus, if someone says

(23) I promise to give you the moon on a silver platter.

that person has not performed a proper promise, since the action can not be carried out, and hence the addressee can not subsequently accuse the speaker of going back on a promise. Similarly, according to this theory, if someone says

(24) I promise to blow up your car if you come to my party.

this can be considered to be a successful promise only if, for some reason, the addressee wishes to have his car blown up. ((24) can, of course, easily be considered an indirect form of threat, which is a different speech act). [Searle 71] claims that it is possble to give necessary conditions of this kind for the success of every type of performative utterance. At present, such felicity conditions have been formulated for a number of speech acts, including all those major types present in the data of this study. As an example of this type of condition, Figure 1 gives felicity conditions for directives.

There are several reasons why this account of the success of speech acts is insufficient for the present project. One is that it concerns only the successful establishment of a particular speech act. Thus, it permits us to determine whether a particular speech act, for example, a promise has been made, but does not extend to determining whether that promise actually is carried out. The actual carrying out of a speech act in the world is termed, within speech act theory, its **perlocutionary force**, and all writers on speech act theory have deemed it beyond the scope of the theory's consideration.

A second, and more serious problem with this way of determining the success of speech acts is that it crucially depends on knowledge of mental events such as the intention of the speaker, the desire of the addressee, etc. This focus is inappropriate in the present study for a number of reasons. One is that there is no reliable way of ascertaining the intention of a speaker, or any other such postulated mental entity. Speech act theory relies on the judgment of the analyst in making this determination. This is a reasonable move in cases where the example sentences have been constructed by the analyst, and represent relatively simple cases. But in the more complex cases which occur in actual transcripts, analysts differ in their interpretations, and a definitive interpretation can not be determined in this way. One might argue that the speakers could be asked what their intention was. In the aviation accident situation, of course, this is rarely possible, since many of the speakers died in the accident. Even when the speaker can be asked about his intention, his memory of an intention is not fully reliable, and can not be given privileged status. In fact, his account of his intention is more data to be analyzed, and data of a more complex type than a direct transcription of an utterance.

3.3.2 Success and Failure of Speech Acts in a Real World Context

Since the CVR transcripts, taken together with the NTSB reports, provide a context that contains a wide range of information about the actual effects of speech acts, several different ways of determining speech act success are possible in the present research. The first and simplest measure of success is to look at later utterances to see what effect the speech act had.

For example, if we are interested in whether a suggestion by a subordinate was accepted, we can try to judge if the captain accepted or rejected it, based on what he said. (This process, called ratification of draft orders, is discussed in section 7.3.) If we are concerned about whether the proposal of a new topic succeeds, we can check whether the utterances immediately following this topic continue it. This simple test is possible because the transcripts of the entire interaction are available. It is also possible because we are concerned not with the speakers' and addressees' beliefs, intentions, etc, but only with their actions. Thus, with a speech act of persuading, we are concerned not with whether the addressee actually feels convinced, but only with whether he acts as though he were convinced.

This method of simple inspection is not sufficient when the failure is more complex, for example when the addressee appears to misinterpret the speaker's speech act. For example, someone may say

(25) It's cold in here.

intending it as an extremely indirect form of the request

(26) Close the window.

The addressee may misinterpret the speaker's intention, and merely respond

(27) Sure is.

If the speaker and the analyst are the same person, then the speaker can give an account of what he intended by his utterance; this gives a basis for analyzing the response as a misinterpretation. But as discussed above, for data like that of the present study, there is no reliable access to the intentions of speakers.

Despite the difficulty, some account of misinterpretation is necessary, because there are cases that we wish to analyze as misinterpretations. For example, the sequence (28a-b) gives us as analysts the sense that something has gone wrong, whether through misunderstanding or through deliberate stubbornness.

```
(28a) CAM-2   Do you have any idea of what the frequency of the
                Paris VOR is?
(28b) CAM-1   Nope. dont really give a #.
                              (Texas/Mena/73; 28:20.5)
```

The speaker of (28a) could have been making a request for information about the addressee's state of knowledge. Or he could have been making a request for action -- either that the addressee find out the frequency, or that he actually use the VOR. Of these three possibilities, we as analysts, without access to the speaker's intentions, are fairly sure that the first of these possibilities, the request for information, was *not* what the speaker intended, and that in so taking it, the addressee was in fact misinterpreting it In order to justify such a claim, we must introduce a new distinction, between the **prior** force of a speech act, before response, and its **posterior force**, after some response has been made. The **prior force** of a speech act derives from:

1. its linguistic form;

2. the previous linguistic context;

3. the identity of its speaker and intended addressee; and

4. shared information available to speaker and intended addressee.

Some speech acts are relatively unambiguous, such as

> **(29) CAM-1 Ah call the ramp, give em our passenger count including laps tell em we'll land with about four thousand pounds of fuel. (1751:35)**

Most readers or hearers will judge that (29) is an order, and that no other interpretation is tenable. However, there are other speech acts which are more ambiguous, so that an analyst will see several possible interpretations of their force. For example, (30), spoken by the captain to the flight engineer, may be interpreted as an order, or as a question about the flight engineer's feelings and plans.

> **(30) CAM-1 Do you want to run through the approach descent, yourself? So you don't forget something (1754:18)**

Example (30) has two recognizable prior forces: order and question. Furthermore, analysts can judge the relative possibility that each alternative actually was chosen by the participants in the situation. These judgements of possibility may be expressed in terms of fuzzy set membership [Zadeh 65, Zadeh 77, Goguen 69][1]. Thus, we can say that (30) has a .9 membership in the set of orders, and a .4 membership in the set of questions. (These values are based on a gedanken experiment performed by the analysts, supplemented by judgements of researchers at NASA Ames. It would be perfectly feasible to use members of the aviation community as subjects in an actual experiment to determine degree of membership of selected examples.) We call this range of interpretations of social force, together with their possibility values, the **prior spectrum** of the speech act.

Similarly, the **posterior force** of a speech act is an interpretation of its social force together with its relative possibility value, as judged by an analyst on the basis of the addressee's response to the speech act. Thus a response to (30) like (31) would assign to (30) the posterior force of a question.

> **(31) No, I'm pretty solid on that procedure.**

The actual response, (32), assigns to it the posterior force of an order.

> **(32) Yes sir. (1754:25)**

Posterior force can also give rise to a fuzzy set of social forces, called the **posterior spectrum**.

---

[1] Fuzzy set theory differs from probability theory in its reference to *possibility* rather than to *probability*; more technically, the events involved need not have values that add up to 1, as indeed they do not in the example given here.

This will often have a non-zero value for only one possibility, but an ambiguous response can give rise to a spectrum having more than one posterior force with non-zero possibility value.

To restate the notions of prior and posterior spectrum in more social terms, the prior spectrum of a speech act is its fuzzy set of possible interpretations given everything that the participants know up to and including the moment of utterance, while its posterior spectrum is the fuzzy set of interpretations taking into account whatever subsequent talk the participants actually produced.

Thus, in example (28), the interpretation assigned after knowing the addressee's response is the same interpretation assigned the highest degree of set membership before knowing that response, i.e., the interpretation that it is an order. For example, we judge that (28a) has membership of .8 in set of requests for action -- contacting the VOR, membership of .7 in the set of requests for action -- finding out the frequency of the VOR, and membership of .2 in the set of requsts for information about the addressee's state of knowledge (see Figure 3). The response, (28b) assigns a posterior force of request for information about the addressee's state of information. Since that interpretation had the lowest degree of membership of all those in the prior spectrum, the sense of misinterpretation can be described as a mismatch between our judgment of the prior force of the utterance and the posterior force actually given to it.

It is important to note that this analysis takes an extremely literal approach to the interpretation of speech acts. There is no way to tell whether the speaker of (28b) actually misunderstood (28a), or whether he understood it and was being deliberately obstreperous. With only the transcript, there is no way to choose between these possibilities, and the analysis of prior and posterior force does not operate at this level of speaker motivation.

3.4 Classification of Speech Act Types

Having established the theory of speech acts, it is now possible to make a taxonomy of possible speech acts. [Searle 79] offers a classifiction which is intended to be complete, for all contexts. In this work, we make use of those categories which actually occur in the CVR transcripts.

Searle's general classification is as follows:

1. **Assertives**, which "commit the speaker (in varying degrees to ... the truth of the expressed proposition." Verbs used for assertive speech acts include **believe** and **conclude**.

2. **Directives**, which are attempts (in varying degrees by the speaker to get the hearer to do something." This class includes orders and suggestions.

3. **Commissives**, which "commit the speaker to some future course of action." Typical verbs used for commissives include **promise** and **offer**.

4. **Expressives**, which "express a psychological state ... about a state of affairs specified in

```
---------------------------------------------------------------------------
1.

.8      x
        x          x
.6      x          x
        x          x
.4      x          x            x
        x          x            x
.2      x          x            x
        x          x            x
---------------------------------------------------------------------------
    Request      Request      Request      Acknowledgement    Report
    for          for          for Info
    Action -     Action -     on Adressee's
    Use VOR      Find Out     State of
                 Frequency    Knowledge

    "Do you have any idea of what the frequency of the Paris VOR is?"
```

Figure 3: Prior Spectrum of a Speech Act

```
---------------------------------------------------------------------------
1.                              x
                                x
.8                              x
                                x
.6                              x
                                x
.4                              x
                                x
.2                              x
                                x
---------------------------------------------------------------------------
    Request      Request      Request      Acknowledgement    Report
    for          for          for Info
    Action -     Action -     on Adressee's
    Use VOR      Find Out     State of
                 Frequency    Knowledge

      "Do you have any idea of what the frequency of the Paris VOR is?"
      "Nope, don't really give a #."
```

Figure 4: Posterior Spectrum of a Speech Act

the propositional content." Verbs used for expressive speech acts include **thank** and **apologize**.

5. **Declarations**, which, if successfully performed, "bring about the corespondence between the propositional content and reality." For example, if the captain declares a MAYDAY, then, indeed, the flight has MAYDAY status.

The CVR data contains the following types of speech acts:

1. **Request**. This class includes orders, requests, suggestions, and questions, that is, all speech acts which call for the addressee to perform some action, either a physical act or a speech act (as in answering a question.) It corresponds to Searle's class of directives.

2. **Report**. A report is an indication of some state of the world. This class corresponds to Searle's assertives. In the current data, it includes the following distinguishable subtypes, in addition to simple reports:

   a. **Support**. This is a special type of report that occurs most characteristicaly in explanations. It is a report of some state of the world which is offered as supporting evidence for a statement within an explanation (see Section 7).

   b. **Challenge**. Similarly, a challenge is a type of report which occurs most characteristicaly in explanations. It is a report which is offered as a challenge to some statement within a explanation.

   c. **Psycho-ostensive** This is a report, direct or indirect, of the speaker's psychological state [Matisoff 79]. An examples is:

      ```
      Less than three weeks to retirement, you better get me
      outta here
                          (1748:17)
      ```

   As we use this term, psycho-ostensives are specificaly not operationally relevant. There also are reports of internal states which are at least potentially operationally relevant. For example

      ```
      I'm so tired I can't keep my eyes open.
      ```

   Such cases would be considered as simple reports, not as psycho-ostensives.

3. **Declaration**. This is the direct equivalent of Searle's class of declarations. In the aviation context, declarations may be of MAYDAY or PAN.

4. **Acknowledgement**. This speech act acknowledges either that the speaker has heard some report, or that he will perform the action indicated by a request. In its latter function, it corresponds to Searle's class of commissives.

3.5 Speech Act Charts

Thus far, the discussion has focussed on single speech acts, or upon short sequences of speech acts. In addition, it is sometimes desirable to study larger patterns of speech acts. To do this, we use the **speech act chart**, a graphic device for displaying selected features of speech acts (such as aspects of their propositional content, their speech act type, their speaker, and their addressee) as a function of time. Speech act charts are especially useful for displaying all speech acts having some particular propositional content, such as fuel level or altitude.

Figure 5 is a speech act chart for the United/Portland/78 accident, showing all speech acts whose propositional content is fuel level. Fuel level was chosen as the relevant propositional content for this accident since its probable cause was determined by the NTSB to have been "failure of the captain to monitor properly the aircraft's fuel state, resulting in fuel exhaustion to all engines." On this chart, the actual fuel level is assumed to be the linear function of time determined by two given points: 7000 pounds reported by the captain to company at 1740:47, and nominal zero fuel level at 1813:38, when all engines flamed out. The chart has three scales for fuel level: one for the actual level, a second for the reported level at time of speaking, and a third for the projected level at some time later than the time of speaking.

We now give a narrative of the events shown in Figure 5, based on the actual utterances of crew members having fuel level as propositional content:

> The first speech act on the chart occurs before the CVR transcript begins, but the NTSB report on this accident mentions that recordings show that at 1740:47 the captain reported 7,000 pounds of fuel on board to company dispatch and maintenance personel. We take this point (7,000 pounds at 1740:47) as one end of a line showing projected linear decrease of fuel level. The other end is the point at which all engines flame out (0 pounds at 1813:38), which we take as nominal zero fuel level.
>
> Beginning at 1746:52 is the first of three request-report-acknowledgement triples: the first officer requests fuel level, the flight engineer reports 5,000 lbs. and then the first officer acknowledges the report. In the second of these triples, beginning at 1748:54, the first officer requests fuel level from the captain, who reports `Five`, and then the first officer acknowledges this by repeating `Five`.
>
> At about 1750:16, the captain requests from the flight engineer `a current card on weight figure (for) about another fifteen minutes`, and at 1750:30 elaborates this with `Yeah, give us three or four thousand pounds on top of zero fuel weight`. We interpret this as having the force of a projection, that in fifteen minutes, i.e. 1805:30, there will be 3,000 to 4,000 pounds of fuel.
>
> The next speech act on the chart is a challenge of this projection by the flight engineer, who says at 1750:34, `Not enough.  Fifteen minutes is gonna --- really run us low on fuel here`.
>
> This doubt apparently has no effect, for the flight engineer at about 1752:30 says to the

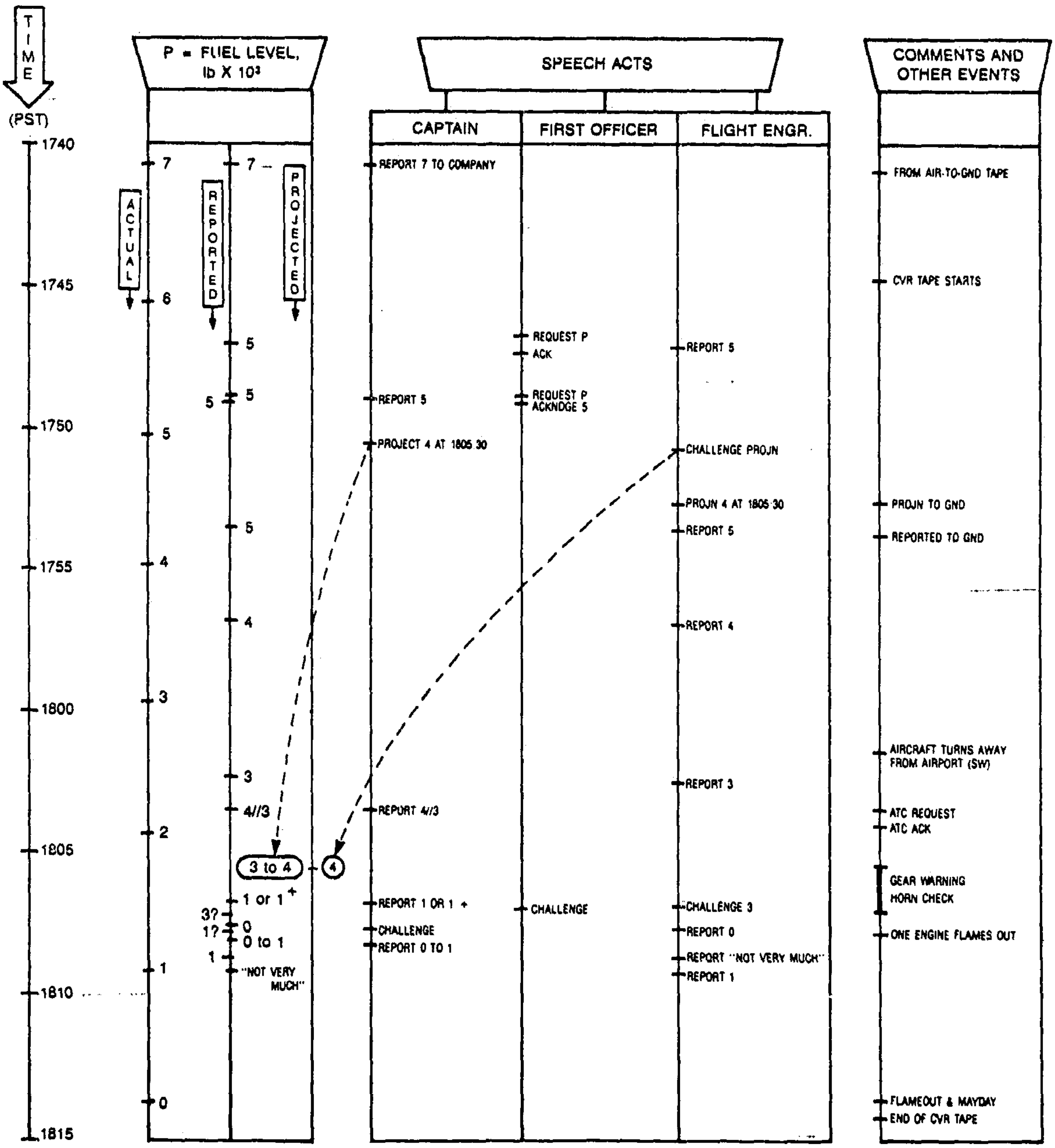


Figure 5: United/Portland/78 Speech Act Chart

company, **We'll be landing with about four thousand pounds of fuel**, which we interpret as a projected fuel level for 1805:30. (Note that this is actually on the *high* side of the range previously projected by the captain.) Slightly later in the same conversation, the flight engineer reports 5,000 pounds of fuel to the company. (This value may be a little high, but it is not significantly so, in contrast to the previous two projections.)

At 1756:53, the first officer initiates another triple with the flight engineer. The value reported is 4,000 pounds.

At 1802:22, the flight engineer reports, without having been requested, **We got about three on the fuel (and that's it)**, apparently an aggravated report.

At 1803:23, Portland Approach Control requests **amount of fuel** from the captain, who reports **about four thousand well, make it three thousand pounds of fuel**, which is acknowledged by **Thank you**. The captain's report is first hedged (**about** and **well**) and then corrected downward to 3,000 pounds, which appears to be quite accurate.

It is interesting to notice that from just after 1804:04 until 1806:10 the crew members are involved with a check which they had forgotten to do (of the gear warning horn) and do not attend to the question of fuel level. They are in fact flying almost directly away from Portland Airport at this time. At about 1806:40, one engine flames out.

At 1807:00 the captain reports **showing a thousand or better** and the first officer challenges this report with **I don't think it's in there**. The flight engineer says **Showing three thousand isn't it** which we interpret as a mitigated report.

At 1807:31 the flight engineer reports **It's showing zero** and the captain responds **You got a thousand pounds, you got to**. At 1807:51 the captain reports **Showing down to zero or a thousand**, which is acknowledged by the flight engineer with **Yeah**.

At around 1808:50 the flight engineer repots **Not very much more fuel** which represents a vague range of values, and at 1809:10 he reports **We're down to one on the totalizer** and then **Number two is empty**.

A number of observations can be made using this chart. One is that the actual reports of fuel level were fairly correct; it was the projected fuel levels that were serious underestimations. The second observation is that there is an interesting erosion of the flight engineer's challenge of the captain's erroneous projection of fuel weight. This sequence begins at 1750:30 with the captain's mistaken projection that there would be three or four thousand pounds of fuel at 1805:30. The flight engineer challenges this projection: **Not enough. Fifteen minutes is gonna really run us low on fuel here**. But the flight engineer does not maintain his challenge, and reports to ground a projected figure of four thousand pounds. (Our estimate of the actual fuel level at 1805:30 is about 1700 pounds.) A final observation is to note the attention given to checking the gear warning horn in the important period from approximately 1804 to 1806, during which time the aircraft was flying away from the airport and running out of fuel.

It should be noted that similar graphic devices are used in NTSB reports to display particular utterances over time, altitude, etc. Charts of this type are valuable because they permit us to focus on the specific propositonal content of interest, and on the pattern of speech acts which express it. This is important since errors of resource management and crew coordination often occur not as the result of a single speech act, but in the course of a chain of speech acts.

4 MITIGATION AND AGGRAVATION

This section defines the notions of **mitigation** and **aggravation**, and introduces the scale which they form. An empirical validation of this scale is also given. Variables that range over the mitigation/aggravation scale play an important role in several of the hypotheses discussed in Section 9.

4.1 Definition of Mitigation and Aggravation

The definition given in this subsection attempts to capture the intuition that, while some sentences are quite direct, other sentences with the same (or similar) social force are more indirect; moreover, these differences in degree of directness correspond to differences in degree of politeness. Thus, most native speakers of English feel that (33) is quite direct, while (34) is quite indirect, and also more polite.

(33) CAM-1 **Reset that circuit breaker momentarily, see if we get gear lights** (1810:17)

(34) CAM-1 **Do you want to run through the approach descent yourself? So you don't forget something** (1754:18)

Mitigation and aggravation are possible because English (like all human languages) presents its speakers with a variety of means of expressing the same propositional content. A **mitigated** form is one which expresses a given propositional content in such a way as to avoid giving offense. An **aggravated** form, such as (35), has more potential for giving offense.

(35) CAM-2 **Get this # on the ground** (1801:45)

(Actually, (35) is not very likely to give offense in the context in which it was used, but its linguistic form is nevertheless aggravated, rather than direct.)

As many analysts have noted, aggravation is considerably rarer than mitigation in most social situations, and there are far more forms for mitigation than for aggravation [Labov & Fanshel 77]. Therefore, the following discussion focusses on mitigation.

There are many linguistic devices which function as mitigations: questions are more mitigating than imperatives; modal auxiliaries, such as **would**, **might** and **could**, are more mitigating than simple verbs; markers of request for agreement, such as **right** and **OK**, are mitigating. This list could be continued almost indefinitely.

However, in order to deal with all the mitigation devices and strategies ocurring in a given text, it would be preferrable to have some theory of why such a seemingly heterogenous group of linguistic phenomena should serve this function. Such a theo⁻ has been given by [Brown and Levinson 79]. (A similar theory of politeness has been developed by Robin Lakoff in a series of papers; we use Brown and Levinson's theory because of the convenience of their single unified presentation.)

Brown and Levinson's account is based on the notion that politeness is the attempt to avoid **face threatening action**, where **face** is the public self-image that every member of the culture wants to claim for himself/herself [Goffman 67]. There are two types of face, negative and positive. **Negative face** is "the basic claim to territories, personal reserves, rights to non-distraction -- i.e. to freedom of action and freedom from imposition." **Postive face** is the "positive consistant self-image or 'personality' (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants" [Brown and Levinson 79] p. 66. These two types of face give rise to two types of politeness, also called negative and positive. **Negative politenes** attempts to minimize the degree of trespass to the addressee's autonomy; **positive politeness** attempts to minimize the distance between speaker and addressee, so that the speaker's and addressee's desires appear to be the same.

Brown and Levinson also identify a third class of strategies for politeness, called **off record** strategies. These are modes of indirection which permit the speaker to avoid being held accountable for what he/she intends to convey. Such strategies are very rare in this data. This is fortunate, since they are particularly likely to be misinterpreted. No further discussion of off record strategies is necessary for the present study.

Figures 6 and 7 show the negative and positive strategies we have found, using as data all directives (i.e., requests and orders) in the United/Portland/78 transcript (excluding directives for acts that are purely speech acts). The mechanism of many of these strategies can be explained by the theory of indirect speech acts given in Section 3.2.2, but the present section is concerned with the dimension of mitigation, rather than with the mechanisms by which indirection is achieved. Figure 6 shows the negative mitigation strategies found in this data, and Figure 7 shows the positive mitigation strategies. Although directives are not the only speech acts that can be mitigated, they are among the most likely to be mitigated, since a request that someone do something, following Brown and Levinson, is a threat to the adressee's autonomy. However, it should be noted that in the cockpit situation, where there is a strict and known hierarchy of command, a request for action is less face threatening than would be the case in a more fluid and undefined social situation.

Since many examples contain more than one mitigation device, the device of interest is indicated by <u>underlines</u>. Speaker and addressee are denoted by numerals; for example 1 --> 3 is spoken by the captain to the flight engineer.

- Give Reason for Request

  Do you want to run through the approach descent yourself? <u>So you don't forget something.</u>

  1 --> 3 (1752:20)

- Give Options about Compliance

- Frame Request as Suggestion

  If I might <u>make a suggestion</u> -- you should put your coats on.

  4 --> 1,2,3 (1748:21)

- Frame Order as Request

  <u>Why don't you</u> put all your books in your bag over there Rod.

  1 --> 2 (1755:55)

- Minimize Extent of Action Required

  Do you have the signal for not evacuate, also the signal for protective position. <u>That's the only things I need from you right now.</u>

  6 --> 1 (1744:40)

- Make Request Hypothetical

  <u>If I might make a suggestion</u>, you should put your coats on.

  4 --> 1,2,3 (1748:21)

- Use Modal Auxiliary

  If I <u>might</u> make a suggestion, you <u>should</u> put your coats on.

  4 --> 1,2,3 (1748:21)

- Use <u>If</u> Clause

  <u>If I might make a suggestion</u>, you should put your coats on.

  4 --> 1,2,3 (1748:21)

Figure 6: Examples of Negative Mitigation Strategies

4.1.1 Psychological Status of Mitigation/Aggravation

It should be noted that mitigation and aggravation are linguistic categories, not psychological ones. Thus, when a speaker uses a particular instance of an aggravated form, we can not directly draw any conclusions about his psychological state at the moment, nor about his personality characteristics, although a speaker's long-term profile of use of mitigation/aggravation in different contexts is probably related to his personality characteristics.

Mitigation/aggravation as a linguistic phenomenon is related to the psychological notion of assertiveness, but is not identical to it. Use of few mitigation strategies, or of many aggravation strategies is one way of behaving assertively; there are, of course, many others.

```
- Minimize Distance Between Speaker and Addressee

- Use Informal Syntax
       How much fuel we got, Frostie?
                                 1 --> 3 (1746:52)

- Use Informal Lexical Choice
       But if anything goes wrong, you just charge back and get your
       ass off, OK.
                                 1 --> 4 (1748:40)
- Use us Rather than me
       Yeah give us three or four thousand pounds on top of zero
       fuel weight.
                                 1 --> 3 (1750:30)

- Seek Agreement
       You're going to take care of the shutdown, right.
                                 2 --> 1 (?) (1758:18)
```

Figure 7: Examples of Positive Mitigation Strategies

4.2 Scale of Mitigation/Aggravation

A number of the hypotheses suggested in this report require discriminating degrees in a scale of mitigation and aggravation. The degrees of this scale correspond to the sense felt by the native speakers of a language that some sentences are more polite or more indirect than others. The validity this scale has been established by checking the judgement of linguistic analysts against the judgements of members of the aviation community. (See Section 4.3 for a discussion of how this test was performed.) We have found that four degrees of mitigation/aggravation are the most that native speakers can reliably discriminate. This scale has a midpoint of zero, representing a direct, unmitigated utterance. There are two degrees of mitigation -- low and high. There is only one degree of aggravation, corresponding to the facts that aggravation is much rarer than mitigation [Labov & Fanshel 77], and that there are fewer strategies for effecting aggravation than for effecting mitigation.

4.3 Experimental Support for Scale of Mitigation/Aggravation

This subsection discusses an experiment conducted to demonstrate the reliability of an operationally defined scale for degrees of mitigation. This scale is used in coding data for certain hypotheses tested in this research. This demonstration is important, both to determine whether the linguistic phenomeon of mitigation/aggravation can indeed be viewed as a scale, and to check the reliability of coding. Whether or not mitigation/aggravation forms a scale is relevant to the issue of statistical testing for hypotheses utilizing this variable (see Section 9.2.4).

The reliability experiment on the scale of mitigation/aggravation trained six subjects, familiar with aviation but not with linguistics, collected their ratings of a set of speech acts, and then compared these ratings with the analysts' ratings of the same speech acts. The data set consisted of 31 reports and requests, chosen randomly from the six transcripts. Requests (which include orders) are a natural choice for this test because they are the speech acts most centrally involved with mitigation, since the act of requesting that someone do something is always potentially face-threatening. Reports are the next most important category of speech acts for mitigation. Although they are less often mitigated than requests, they too can play an important role in the misunderstandings that arise in command and control discourse.

The scale of mitigation/aggravation tested had the following four levels: Aggravated; Direct; Low Mitigation; and High Mitigation. Our original experimantal plan called for a sample with six reports and six requests of each mitigation level. However, this proved impossible because of the scarcity of examples in certain categories. Starting from the entire body of speech acts in the six transcripts, each with a mitigation rank assigned by one analyst, speech acts were choosen at random and their mitigation ranking was checked by a second analyst. This process continued until the desired number of speech acts were obtained in each of the most common categories. For the rare categories, separate pools were formed containing all the speech acts with that level of mitigation. Some speech acts were eliminated because they had ambiguous social force or because they used contradictory mitigation strategies; the remainder were included in the experiment. Ten of these "bad" sentences were also included in the sample, even though we did not intend to use them in the evaluation process, in order to check the assumption that this kind of sentence would pose special difficulties. A separate randomization step determined the order in which these 41 speech acts would be administered to subjects for coding.

The experimental subjects consisted of six commercial airline professionals, including two of rank captain, three of rank first officer, and one of rank flight engineer. (We had expected three of each rank, for a total of nine subjects, but three subjects failed to appear at the test site.) Before being asked to rank the speech acts, they were given pre-test training in the meaning of the categories used: A previously prepared explanation of the notion of mitigation was read to the subjects. They were then given some sample written examples to rate, and these examples were discussed by one of the analysts with the group. Finally, they were given the written speech act protocols to score.

An analysis was made of the match between the subjects' mitigation ratings and those of the analysts. The criterion which is generally used for reliability of such scales is a stringent one: there should be at least an 80% match between the subjects and the analysts; that is, the average number of agreements of the analysts judgements with the subjects exceed 8 out of 10. This criterion was just met in the present experiment, in which the average agreement of the six subjects with the analysts' judgement was .801. Although neither the number of subjects nor the number of stimuli were as great as originally planned, they are sufficient to support concluding that this is indeed a reliable scale for degrees of mitigation.

A more detailed analysis of the data provides further evidence that a scale of the kind required has indeed been defined. First of all, no two subjects had an agreement ratio with each other that was as high as their agreement ratio with the analysts. (In fact, the average agreement ratio among subjects was only .68.) This strongly suggests that much of the disagreement that did appear is simply due to variance among subjects less well trained than the analysts. (Indeed, the agreement of the analysts ratings with the modal response of the subjects is far higher than .8.)

Another factor affecting subject variance in coding is regional dialect differences. While data from six subjects can only be regarded as suggestive for this purpose, the following facts should be noted: there were two subjects each from California, New York State, and the South; the analysts are from the North-East (one from New York City and one from Western Massachusetts). The inter-subject agreement for New York subjects is higher than that for California subjects or Southern subjects (.81 versus .71 for Califonia and .68 for Southern subjects). The average agreement of the New York subjects with the analysts is higher than with any other region (.90 versus .76 for California and .71 for Southern). These figures suggest that further experimentation would be valuable, in order to determine whether regional dialect differences in aircrew composition could be a significant factor in speech act misinterpretations that could potentially lead to accidents. This would be a significant finding, because it would be possible to train crew members to recognize the intended mitigation values of speakers from other regions. Indeed, the fact that during the pretest period, subjects joked with one another about their regional mitigation peculiarities suggests that this factor should be easily trainable. We feel that the validity of the mitigation/aggravation scale in measuring a general linguistic phenomenon is strongly supported by the fact that finer grained regional differences can be detected.

5 SITUATIONAL VARIABLES FOR SPEECH ACTS

Thus far, the discussion of speech acts has focussed on language in the cockpit without any special consideration of the different types of situation which can occur, and which affect the form of the language produced by crew members. This section examines three types of special situation: Crew Recognized Emergency, Crew Recognized Problem, and operationally relevant versus non-operationally relevant discourse.

5.1 Crew Recognized Emergency

Crew recognized emergency (CRE) is a social, rather than a legal or factual category. The beginning of the crew recognized emergency is defined as the first point at which the entire crew begins to attend to that situation which led directly to the accident. There are several remarks to be made about this definition:

1. In order to identify the situation which led to the accident, we rely upon informed and documented opinion in the aviation community. In practice, this means that we rely on the National Transportation Safety Board's accident reports, but in disputed cases, it

would also be possible to use a minority report, other published materials, or oral reports from members of the aviation community.

2. The definition requires that the entire crew attend to the situation. It may be the case that individual crew members attend to the situation that led to the accident long before the crew recognized-emergency point, and may even have attempted to bring it to the attention of the rest of the crew. However, it is group attention that is being defined here.. Note that in practice, this means the attention of the captain, since in the command and control situation, the captain has the authority to direct the attention of the crew to any situation which he considers to be threatening, while other crew members may suggest but can not compel such attention.

3. In some accidents there may never be a crew recognized emergency. These are cases in which the crew never attends to those situationss that caused the accident.

The concept of crew recognized emergency is required since a number of our hypotheses postulate differences between periods during which the crew members believe that the flight is proceeding normally, and periods in which they believe that they are in an emergency situation. The captain's official declaration of a Mayday does not serve to identify this point, since this declaration often appears quite late, considerably after the point at which the crew begins to act as if they were in an emergency situation. **Mayday** is a legal category, specifying a situation in which there is "immediate danger to equipment and personnel."

A clear example of crew recognized emergency can be found in the United/Portland/78 transcript. The situation leading directly to the accident was the "exhaustion of fuel to all four engines." As the speech act chart in Section 3.5 clearly shows, there was continued attention to the current fuel level throughout the thirty minutes of transcript available. The possibility of running out of fuel is first raised by the flight engineer quite early, 24 minutes before the actual impact. However, the crew recognized emergency point does not occur until considerably later, 7 minutes before the impact. This is the point, beginning at 1806:34, at which the flight engineer reports the loss of an engine and first the copilot and then the captain begin to react to this situation.

```
(36a) CAM-1   Okay we're going to go in now, we should be
              landing in about five minutes.
(36b) CAM-3/2 I think you just lost number four engine, Buddy, you --
(36c) CAM-6   Okay, I'll make the five minute announce, announcement,
              I'll go, I'm sitting down now
(36d) CAM-2   Better get some cross feeds open there or something
(36e) CAM-3   Okay
(36f) CAM-6   All Righty
(36g) CAM-2   We're going to lose an engine Buddy
(36h) CAM-1   Why?
(36i) CAM-2   We're losing an engine
(36j) CAM-1   Why
(36k) CAM-2   Fuel
(36l) CAM-2   Open the crossfeeds man
```

```
(36m) CAM-1   Open the crossfeeds there or something
              ((simultaneous with above))
                                        (1806:34-52)
```

In this example, (36b) is the first utterance of the chain of reports and orders about the loss of the engine due to fuel exhaustion. While the copilot and flight engineer attend to this, the captain continues planning with the head stewardess about preparing the passengers for an emergency landing (due to possible landing gear failure). At (36h), the captain finally joins the other crew members in attending to the fuel level and engine state. (It might be noted that Mayday is not declared until 1813:50, about seven minutes later.)

5.2 Crew Recognized Problem

In addition to the Crew Recognized Emergency, we also use the notion of **Crew Recognized Problem** (CRP). This is a situation recognized by the crew as potentially dangerous and not a normal part of flight operations. It could be an actual problem, or some situation which is off-nominal, surprising, or not expected.

The concept of CRP helps to account for the distribution of mitigation in CVR transcripts. Characteristically, mitigation is not uniformly distributed in these texts. Rather, some segments are rich in mitigation, while others have few or no mitigated sentences. In fact, it is the CRP segments which contain the highest proportion of mitigated utterances (see Section 9.4.3 for a precise statement of this hypothesis and its verification).

The correlation of mitigation and CRP is not surprising in light of the function of mitigation. Mitigation in a request serves to minimize the posible offense generated by telling someone what to do. Under normal flight conditions, there is little or no possibility of offense in requesting someone to carry out a routine, expected action which is part of his regular duties. It is in the case of unexpected, non-routine actions that offense becomes a more salient possibility. Similarly, mitigation in reports serves to weaken the degree of certainty with which a speaker expresses some proposition. It is in unusual, unexpected situations that uncertainty is most likely to arise, and most desirable to express. However, mitigation is least frequent in CRE segments, because in the case of an actual emergency, crew members attend almost exclusively to the operational task at hand, paying almost no attention to the social possibility of giving offense by too direct a statement (see Section 9.4.2.)

5.3 Operational Relevance

A very pervasive distinction, entering into many of our definitions and all our hypotheses, is whether some utterance or some particular discourse unit is operationally relevant. Operational relevance means that the utterance is directly involved with the achievement of successful mission completion. This definition insists upon **direct** involvement; thus, a request for a snack would not be defined as being directly operationally relevant, even though it might have some effect on the state of a crew member, and hence an indirect effect on successful mission completion.

It should be noted that there is no value judgement involved in this definition. We do not wish to suggest that non-operationally relevant discourse should not occur in the cockpit. As the example of the request for a snack suggests, a non-operationally relevant utterance can have valuable indirect effects. Even utterances which do not have any apparent indirect effect on successful mission completion, utterances which could be described as 'just shooting the breeze', might be useful in maintaining alertness in low-workload flight segments.

The distinction between operationally relevant and non-operationally relevant utterances has been introduced because there are certain phenomena which are potentially of great importance in operationally relevant discourse, but have no serious consequence in non-operationally relevant segments. An example is topic failure, which is discussed at length in Section 10. If a speaker introduces a topic which is operationally relevant, and other crew members do not pick up this topic, the consequences can be quite serious. However, a topic failure of a non-operationally relevant topic is of much less concern. We wish to be able to focus on the failure of operationally relevant topics, without having to consider non-operationally relevant cases; this definition allows us to do so.

6 COMMAND AND CONTROL DISCOURSE

The command and control perspective on CVR transcripts involves determining the relevance of any talk in the cockpit to successful mission completion. This perspective gives primacy to the operational aspect of talk; that is, to how it helps to get things done. An important point in understanding operationally relevant talk is that it occurs in the context of a strict hierarchy of authority, in which each member's place is known. (Ambiguities do, in fact, occur, but both the legal definition of the situation and the crew members' understanding of it, is that it is unambiguous.)

These transcripts contain several distinct discourse types, the instances of which may be operationally relevant to varying degrees. The main purpose of this section is to give a precise theory of the structure of the discourse type with the greatest operational relevance to command and control. This is the **command and control speech act chain**, a sequence of command and control speech acts (i.e., orders, requests, acknowledgements, reports, declarations, plans and explanations) having the same major propositional content. This section first considers the general nature of discourse types, summarizing some previous work in this area, and then focusses on this specific discourse type.

6.1 Discourse Unit and Discourse Type

A **discourse unit** is a segment of spoken language, longer than a single sentence, with socially recognized initial and final boundaries, and a formally definable internal structure. (This definition generalizes the criteria given by [Labov 72] for the narrative of personal experience.) A **discourse type** is a theory of the structure of a class of discourse units; that it, it provides a way of recognizing whether or not a given segment of language is an instance of the type. Thus, we can think of a discourse type as the class of discourse units that satisfy a given theory. This corresponds to the familiar distinction between type and token.

Discourse types that have been studied, other than narratives, include pseudonarratives, i.e. spatial descriptions [Linde 74, Linde & Labov 75], plans [Linde & Goguen 78], jokes, and explanations [Weiner 79, Goguen, Linde & Weiner 81]. All these studies are based on an analysis of transcripts of tapes of spontaneous social interaction. It is possible to use this previous work for the present study because CVR transcripts provide exactly this kind of data.

This project requires a precise understanding of how people actually use discourse units, which in turn imposes further requirements on how the research should be conducted, and in particular, on the descriptions to be used for the discourse units involved. First, the work must based upon a careful empirical analysis of actual human discourse in natural situations. This means in particular that we cannot use invented examples to develop our theory (although such examples can be used to illustrate it). Secondly, it is necessary to have a mathematically precise description of the discourse structures of interest. Without this, we cannot properly test hypotheses involving variables that refer to discourse structure.

Third, a suitable theory must also provide a simple and natural taxonomy of the parts that can occur in a given type of discourse, and of how these parts relate to one another. Each of the discourse types that has been studied has certain characteristic parts, and also certain characteristic relationships of subordination among these parts. For example, the characteristic parts of plans include goals, plans, actions and actors, and the charactaristic relationships of subordination for planning include GOAL/PLAN, ACTOR/DO, IF/THEN, and EXOR (for exclusive OR). These subordinators each represent relationships that the parts of a given discourse unit may bear to one another.[2]

For example in an explanation, one statement may be subordinate to another statement by the relationship of providing a supporting REASON, as in the following example where the second statement supports the first.

```
(37a)  CAM-3  Not enough
(37b)  CAM-3  Fifteen minutes is gonna really run us low on fuel here
                                 (1750:44)
```

Other kinds of subordination that can occur in explanation include serving as an EXAMPLE (i.e., an instance) of a statement, and having several statements serve in conjunction, as examples of or as reasons for the same statement.

Such an organization of discourse units into parts that are connected by relationships of subordination is easily and naturally represented by **tree structure**. This offers a convenient, graphically suggestive, and mathematically precise way to represent hierarchical subordination. In this representation, the top node represents the whole discourse, and its immediate subordinates represent the first subdivision into parts. For example, in a plan the top node is a GOAL/PLAN node which indicates a division of the plan into two major parts, the first a goal

[2]However, the parts of discourse units do not readily correspond to any one syntactic structure; thus, a part may be expressed by a sentence, a clause, a phrase, or even by a single word.

part, and the second a plan part. Labels on nodes distinguish different kinds of subordination that occur; these labels are called **subordinators**.

A fourth feature of discourse that a theory must adequately model is the construction of discourse units in *real time*. To do this, it is also necessary to have a notion of the present focus of attention, in order to be able to indicate to what previous part a new part is to be subordinated. (This is discussed in the next subsection.)

6.1.1 Transformation and Focus of Attention

The real time aspect of discourse is especially important in the aviation context, because problems of crew coordination, resource management, speech act interpretation, and so on, actually occur in real time. The process of discourse construction is modelled by **transformations** on the tree structure which represents the discourse structure. Such a transformation can add, delete, or alter a discourse part.

For example, Figure 8 shows the transformation that constructs a tree representing a text of the form **Statement S1 since Statement S2** as in Example (37a-b) above. It begins with S1, **Not enough** in (37f), which is then subordinated by a STMT/RSN node as the transformation adds the statement S2 (**Fifteen minutes is gonna really run us low on fuel here**) that supports S1.

```
                         STMT/RSN
                          /    \
S1          ==>          /      \
                        S1       S2
```

Figure 8: A Transformation

Transformations are very familiar in the literature of linguistics [Chomsky 65]. However, they have previously been applied only to the structure of sentences, rather than to larger discourse structures. Also, such transformations have not been used to model the real time construction of syntactic structures, but rather have been postulated as part of an abstract mechanism for generating syntactic structures.

The **focus** of a discourse represents the presumed focus of attention of the participants at a given point in a discourse; it might be described intuitively as "where we are now."

Graphically, we represent the current focus as a "*" at a particular node on the tree.[3] [Grosz 77] discusses a notion of focus which is primarily semantic in its concern with the resolution of pronoun references; however, it involves a hierarchical structure of "focus spaces" that is similar to what embedded pointers do in our theory.

There is one very important connection between focus and transformations, a constraint on how discourse structure can be built up in real time: a transformation can be applied **only** at the node currently in focus. This constraint on the application of transformations corresponds to speaker's and hearer's expectations about what will occur next. In particular, a transformation cannot be applied to a part of the tree developed earlier without first moving the pointer back to the appropriate subtree. Some transformations, in fact, only accomplish pointer movement, i.e., they just change the focus of attention, and thus do not add any semantic content to the tree.

6.2 Command and Control Speech Act Chain as a Discourse Type

The command and control speech act chain is the basic discourse type for command and control in the cockpit. This section describes this discourse type in the general framework of the preceding section.

Let us begin with the basic definition: a **command and control speech act chain** is a sequence of speech acts, each of which has the same major propositional content. (38) is a typical speech act chain. Its component speech acts include requests, reports, explanations and acknowledgements, all concerning the topic of "fuel weight."

```
(38a)   CAM-1  Hey Frostie
(38b)   CAM-3  Yes sir
(38c)   CAM-1  Give us a current card on weight figure
               about another fifteen minutes
(38d)   CAM-3  Fifteen minutes?
(38e)   CAM-1  Yeah give us three or four thousand pounds on top
               of zero fuel weight
(38f)   CAM-3  Not enough
(38g)   CAM-3  Fifteen minutes is gonna really run us low on fuel here
(38h)   CAM-?  Right
                                                   (1750:16)
```

A possble difficulty in applying this definition lies in determining whether or not a given speech

[3] Actually, more than one pointer is needed for some transformations. We have found constructions in explanation much like those called "parallelism" in classical rhetoric, where there is not only an active node of focus, but also a passive node; in these constructions, some transformations reverse the active and passive nodes, so that addition can proceed alternately among two subtrees. Markers such as on the other hand are used to switch to the other subtree. There are even cases where more than two pointers are needed; for example, if one parallel construction is embedded within another. However, this kind of construction can be quite difficult to understand, and is not found in the CVR transcripts that we have studied.

act has the same major propositional content as those preceeding it. This can be a difficult problem for discourse domains with a wide or unlimited range of possible topics; however, aviation discourse presents a limited range of topics that are operationally relevant.

It should also be mentioned that speech act chains can appear to be **discontinuous**, that is, they can be interrupted by other discourse units, including other speech act chains. This does not mean that they are discontinuous structurally, but rather that, like all discourse units, they can be interrupted by actions in the physical world, by the introduction of new participants, or by some other discourse unit with a more urgent topic.

The following subsections respectively discuss, for speech act chains, the categories of utterance, the subordinators that are used, and the rules that govern sequencing; together these constitute a theory of the structure of speech act chains and may be called a "grammar."

6.2.1 Categories of the Command and Control Speech Act Chain Grammar

Operationally relevant speech act chains typically concern possible actions or actions which have already been performed (see Section 5.3). As Section 3 showed, speech acts can also be seen as acts, which alter the state of the world. This subsection presents a category system that includes both linguistic and physical acts; this is necessary for the formal description of the speech act chain.

The most general category is **acts**. This includes physical acts, command and control speech acts, and acknowledgements of such speech acts.

A more specific category is **speech acts**, the basic category of interest for command and control. This category includes requests, reports, and declarations. For example, (39), (40), (41) and (42) are all requests of various strengths, while (43) is a report, and (44) is a declaration.

```
(39) CAM-1 Open the crossfeeds there or something  _______
                                         (1806:52)

(40) CAM-1 Push the breaker momentarily
                                         (1806:52)

(41) CAM-1 Okay ah, what would you do?  Have you got any suggestions
           about when to brace?  Want to do it on the PA?
                                         (1744:50)

(42)  CAM-2 You plan to land as slow as you can with the power on?
                                         (1800:50)

(43) CAM-2 Its flamed out
                                         (1897:00)

(44) RDO-2 Portland tower United one seventy three heavy Mayday we're
           the engines are flaming out, we're going down, we're not
           going to be be able to make the airport
```

(1813:50)

Additional utterance categories of interest for command and control are plans and explanations. These are structurally more complex than the categories discussed here, and discussion of them is deferred until Section 7.

6.2.2 Subordination

This subsection discusses the elements used to construct speech act chains. These elements are of two types: the speech acts used in command and control; and the subordinators that indicate the relationships among them. We have already given an intuitive sketch of the meanings of the various categories of speech acts; the present discussion focusses on how they function within the formal grammar of speech act chains. An abbreviation for use in graphical representations is given for each subordinator; these abbreviation use "square brackets," i.e., [...].

1. **CHAIN:** This node type is the top level subordinator for a sequence of command and control speech acts having the same major propositional content and constituting a speech act chain. This node therefore marks the fact that a sequence of utterances is indeed a speech act chain; it is not usually indicated explicitly in the actual sequence of utterances. The abbreviation is simply [CHAIN].

2. **REQUEST:** Requests are the most typical command and control speech acts. They include questions, commands and suggestions. (A command can be viewed as a request that has been ratified by the captain. See Sections 6.3 and 7.3 for discussions of ratification). In the formal grammar, a request must have the form of a request node subordinating a single subtree, which is the act that is requested. (Searle's taxonomy calls these "directives.") The abbreviation is [REQ].

3. **REPORT:** A report is an indication of some state of the world. The abbreviation is [REP]. In the formal grammar, reports have the form of a [REP] node subordinating a single subtree giving the act or state reported. (45b) is an example.

   ```
   (45a) CAM-2 Ah, what's the fuel show now buddy?
   (45b) CAM-3 Five
   (45c) CAM-2 Five
                                   (1748:54)
   ```

4. **ACKNOWLEDGE:** A command and control speech act (e.g., a request or declaration) can be acknowledged; but challenges, supports, and other acknowledgements cannot be acknowledged. (This is the kind of constraint on sequencing that the rules below are intended to capture.) For example, (46b) is an acknowledgement. The abbreviation is [ACK]. An [ACK] node indicates the subordination of an acknowledgement to the speech act that it acknowledges.

   ```
   (46a) C-1 You gotta keep em running, Frostie
   (46b) C-3 Yes, sir
                          (1808:42)
   ```

Two interesting further points about [ACK] nodes are: (1) the speaker of an acknowledgement must be among the addressees of the request or report that it acknowledges; and (2) more than one addressee may produce an acknowledgement of the same speech act.

5. **STATEMENT/REASON:** Subordinates a request or report on the left, and a reason supporting it on the right. It is abbreviated [ST/RSN]. It may also occur in the opposite order, abbreviated [RSN/ST]. This node type is discussed further in Section 7.2.

6. **STATEMENT/CHALLENGE:** Subordinates a request or report on the left, and a challenge to it on the right. It is abbreviated [ST/CH]. It may also occur in the opposite order, abbreviated [CH/ST]. It is also discussed further in Section 7.2.

7. **GOAL/PLAN:** Subordinates a goal on the left, and a plan to achieve it on the right. Abbreviated simply [GOAL/PLAN]. It may also occur in the opposite order, abbreviated [PLAN/GOAL]. It is also discussed further in Section 7.2.

6.2.3 Rules

This subsection gives the rules of the grammar for speech act chains in simple English, and also in a graphical form in Figure 9. This grammar expresses how speech act chains are constructed in real time. It thus defines the sequences of operationally relevant speech acts that are possible in command and control discourse, and indicates some (but not all) of the sequences that are not possible. It should be noted that this is a grammar of social force rather than of linguistic form; that is, the rules apply to the social interpretations of utterances, rather than to the utterances themselves, or to the sequences of words or sentences which comprise them.

In this grammar, nodes that must subordinate other nodes have "square brackets," e.g., [ACK], and nodes that indicate categories that will later be filled have "pointed brackets," e.g., <REPORT>. The first two rules simply define subcategories of given categories. They are

1. A command and control speech act, abbreviated <SPACT>, may be a request, a report, or a declaration, abbreviated <REQ>, <REPORT> and <DECL> respectively.

2. An act, abbreviated <ACT>, may be a <SPACT>, an acknowledgement, or a physical act, abbreviated <ACK> and <PHACT> respectively.

The basic entity being formalized, the speech act chain, is indicated by a [CHAIN] node; all the speech acts that constitute a given chain will be subordinated to one such node. The beginning of the production of a speech act chain is a single [CHAIN] node with two subordinate <SPACT> nodes; the fact that there are two such nodes expresses the fact that there must be at least two speech acts in a speech act chain. The basic rule of development for speech act chains is simply:

3. A [CHAIN] node with n descendent nodes can be elaborated into a [CHAIN] node with n+1 descendents. This expresses the fact that a speech act chain may be of any length; that is, it may contain any number of speech acts.

The next two rules are basically parallel; they indicate how <REQ> and <REPORT> nodes can be elaborated:

4. A <REQ> node can be expanded into a [REQ] node subordinating an <ACT> node. This means that any request is a request for an action, either a physical action or a speech act.

5. A <REPORT> node can be expanded into a [REPORT] node subordinating an <ACT> node. This means that any report is a report of an action, either a physical or a speech act or of a state of the world.

Next is a set of three rules that may be applied to any node [XX] that is either a [REQ] or a [REPORT] node subordinating an arbitrary subtree:

6. An [XX] node subordinating a subtree may be replaced by an [ACK] node subordinating [XX] with its subtree on the left, and an <ACK> node on the right. This means that any report or request may be acknowledged.

7. An [XX] node subordinating a subtree may be replaced by either: a [ST/RSN] node subordinating the [XX] node with its subtree on the left, and subordinating an <EXPL> node on the right; or a [RSN/ST] node with the same subordinate subtrees in the opposite order. This rule means that any report or request may be supported by giving a reason (RSN), having the formal structure of an explanation.

8. An [XX] node subordinating a subtree may be replaced by either: a [ST/CH] node subordinating the [XX] node with its subtree on the left, and an <EXPL> node on the right; or else a [CH/ST] node with the same subordinates in the opposite order. This rule means that any report or request may be challenged by a speaker giving an explanation of why it is a bad idea.

The final rule has to do with the introduction of planning, and may eventually lead to ratification as discussed in section 7.

9. A [REQ] node subordinating an arbitrary subtree may be replaced by a [GOAL/PLAN] node subordinating the [REQ] node with its subtree on the right, and a <PLAN> node on the left. This means that any request may be incorporated as part of a plan; that is, the simple process of requesting an act, and having that act acknowledged can be elaborated into the process of planning.

These rules are all given graphically in Figure 9; graphical indications of focus of attention are also given there. An extended example is given in the following subsection, illustrating how these rules are used to analyze some actual cockpit discourse.

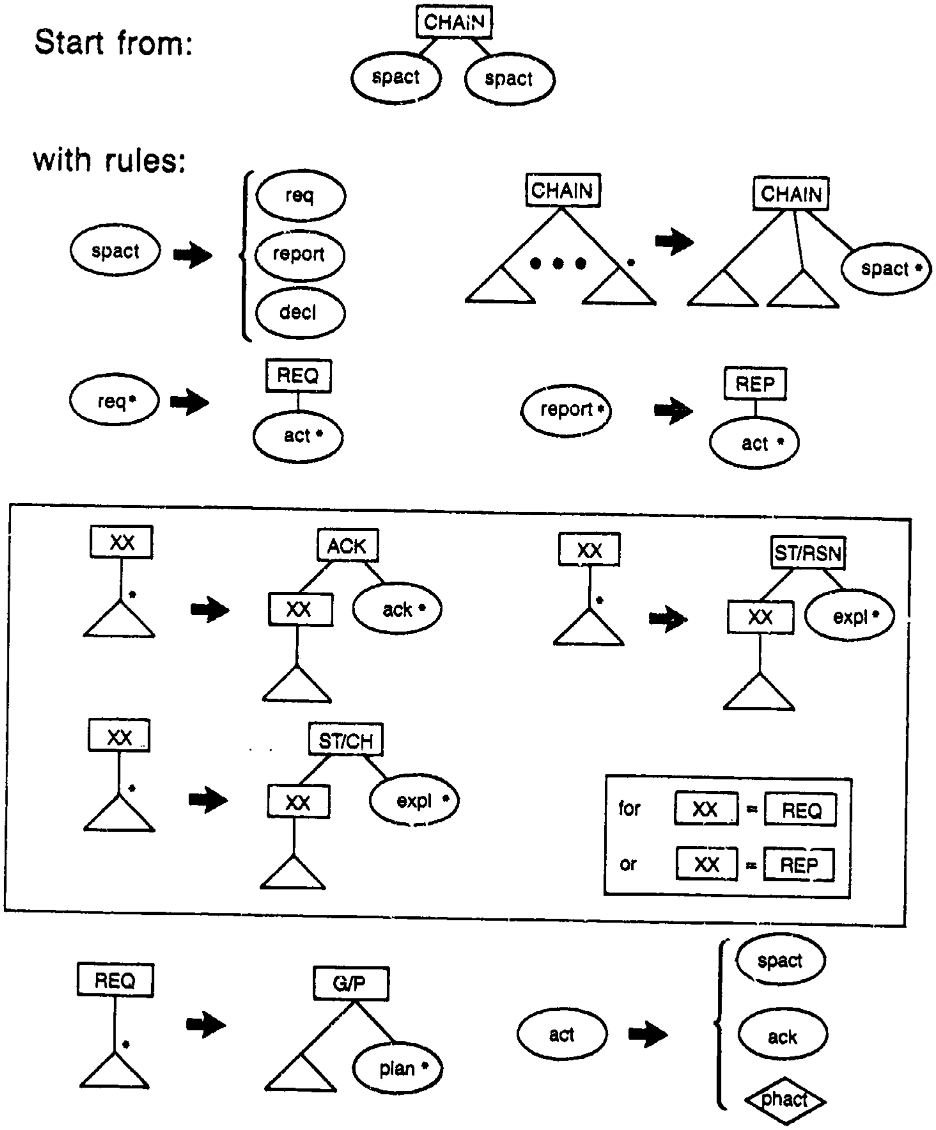


Figure 9: Graphical Presentation of Command and Control Rules

6.2.4 An Example of a Speech Act Chain

The purpose of the preceding discussion has been to describe some constraints on chains of command and control speech acts, and in particular, to indicate some possible and impossible embeddings of social force. That is, we have attempted to specify what sequences of speech acts count as command and control chains, and what sequences would not form command and control chains.

For example, an acknowledgement of a support of a request for an act A should not occur, although an acknowledgement of a request for an act A and a request for a support of a request for an act A may occur.

To illustrate this kind of sequencing, let us consider again the data in example (38).

```
(38a)  CAM-1  Hey Frostie
(38b)  CAM-3  Yes sir
(38c)  CAM-1  Give us a current card on weight figure about another
                fifteen minutes
(38d)  CAM-3  Fifteen minutes?
(38e)  CAM-1  Yeah give us three or four thousand pounds on top
                of zero fuel weight
(38f)  CAM-3  Not enough
(38g)  CAM-3  Fifteen minutes is gonna really run us low on fuel here
(38h)  CAM-?  Right
                                                          (1750)
```

First of all, (38a) and (38b) form what is termed a "call-response" pair, that is, a call for attention followed by an acknowledgement that the addressee is attending. Using the concepts of this study, this can be seen as a request having empty propositional content, followed by an acknowledgement; it cannot be seen as a command and control chain, because chains must have more than one subordinate node. Thus the pair (38a-b) is indicated as shown in Figure 10, where Ø indicates empty propositional content.

```
  [ACK]
   / \
  /   \
 Ø     Ø
```

Figure 10: A Call-Response Pair

Adding (38c-d) to this yields the tree shown in Figure 11, where **c** denotes the propositional content of (38c) and **d** that of (38d).

(38e) refines this propositional content to say that there will be three or four thousand pounds in fifteen minutes, denoted here as **e**. This is followed by an unusually strong challenge in (38f), the propositional content of which, **Not enough**, is indicated by **f** in Figure 12. Rather than repeating the two subtrees of Figure 11, we here denote them as t1 and t2, respectively.

```
          [CHAIN]
          /     \
        /         \
    [ACK]        [ST/CH]
     / \          /    \
    /   \        /      \
   Ø     Ø      c        d
```

Figure 11: A Challenge

```
          [CHAIN]
          /  |  \
        /    |    \
      t1    t2   [ST/CH]
                  /    \
                 /      \
              [REQ]      f
                |
                |
                e
```

Figure 12: A Further Challenge

Finally, (38g) is a supporting explanation of (38f), and (38h) is a support of (38g), and thus of (38f). Thus, the social force of this whole sequence could be notated as in Figure 13, where **g** is the propositional content of (38g) and **h** that of (38h).

```
          [CHAIN]
          /  |  \
        /    |    \
      t1    t2   [ST/CH]
                  /    \
                /        \
             [REQ]     [ST/RSN]
               |         / \
               |        /   \
               e   [ST/RSN]  h
                    /    \
                   /      \
                  f        g
```

Figure 13: A Complete Command and Control Chain

7 PLANNING AND EXPLANATION IN THE COCKPIT

This section discusses planning and explanation as discourse types and also introduces and discusses the important notion of a "draft order." It should be noted that the terms "planning" and "explanation" refer to linguistic activities performed by two or more people rather than to planning or explanation as individual, mental activities.

7.1 Importance in the Cockpit

Planning and explanation are important because they represent the process by which a group decides what to do, or what some unexpected situation means. Planning and explanation are found particularly in situations which are unexpected or off-nominal. This is not surprising. In a flight progressing normally, there are standard operating procedures and a pre-filed flight plan; hence, there is little need to make additional plans. Similarly, in a normal flight, the state of the equipment, weather, etc, is known (or believed to be known); hence, there is little need for the crew to reason about the state of things, or to explain it to one another at length.

Because planning and explanation correlate with unexpected and problematic situations (see Section 9.5.6), they are crucial to our understanding of aircrew behavior. For this reason we consider them in some detail.

7.2 Theory of Planning and Explanation

This subsection reviews some previous work on planning and explanation, and then discusses the additions required for application to the aviation context.

7.2.1 Review of Work on Planning

The linguistic study of small group planning [Linde & Goguen 78] has shown that the language used to accomplish planning is a discourse type in that: it has an initial boundary, consisting of the statement of the goal which the planning is intended to accomplish; it has a final boundary, which may consist of the group's evaluation of the probable effects of the plan, or of their approval and acceptance of it; and it has a precise internal structure, consisting of members' proposals of new sublans, or of their proposals to modify or replace parts of the plan previously proposed by others.

Formally, this internal structure of the planning discourse unit is described by a sequence of **transformations** on the plan being formed by the group. (See Section 6.1.1 for a discussion of transformations.) In planning, these transformations represent the real-time effects of proposals by members to add, delete, or modify plan parts. Similarly, the relations of logical subordination which hold among the plan parts are represented by a **tree** structure.

An example is given in Figures 14 and 15, a plan from the United/Portland/1978 accident. The major goal, stated by the copilot, is to **call out the equipment**; his plan for this is to **have the company call**. This PLAN/GOAL relationship is indicated in Figure 14. In Figure 15, the captain replaces the copilot's plan with a plan to **call dispatch in San Francisco**. In Figure 16, he adds a node which indicates that **maintenance down there will handle it that way**.

In these figures, what was said is shown on the left. On the right is shown the tree resulting from the application of the transformation invoked by that portion of text to the previous plan tree. The sequence of transformations starts with an initially empty tree, which is not explicitly shown, and ends after the captain has elaborated the copilot's simple plan.

```
CAM-2  He's going to have                     PLAN/GOAL
       the company call                          /   \
       out the equipment?                       / *   \
                                          have the      call out
                                       company call     the equipment
```

Figure 14: A GOAL/PLAN Node

```
CAM-1  We'll call dispatch                  GOAL/PLAN
       in San Francisco                       /    \
                                             /      \
                                      call out      ACTOR/SAY/TO
                                 the equipment       /    |    \
                                                    /     |     \ *
                                                  we     call  dispatch
                                                               in San
                                                               Francisco
```

Figure 15: Addition of an ACTOR/SAY/TO Node

The order of application of transformations is the same as the order of production of clauses in the text. However, the order of nodes in the tree may no longer correspond to the order in which they were produced, if deletion or rearrangment transformations have been applied.

There are a number of relations of logical subordination which have been found in plans. The first and most basic of these is the GOAL/PLAN relationship, which subordinates a plan to an announced goal. Next is the AND relationship, which can subordinate any number of subplans or subgoals. There is also EXOR, for "(mutually) exclusive or," either of goals or plans; IF/THEN, for a conditional plan or goal; and ACTOR/DO, and its special case ACTOR/SAY/TO, in which some actor says something to some other. Finally, there are the terminal nodes, which represent actions and goals which are not further logically decomposed, but which are instead filled by parts of the language produced by the speakers. Note that the compound nodes permute freely, depending on the order in the text; thus, we find

```
CAM-1  and maintenance                 GOAL/PLAN
       down there will                  /    \
       handle it that                  /      \
       way                    call out     ACTOR/SAY/TO
            (1754:27)     the equipment     /    |     \
                                           /     |      \
                                          /      |(call) \
                                         /       |        \
                                             ACTOR/DO
                                              /    \
                                             /      \ *
                                   maintenance      handle it
                                   down there       that way
```

Figure 16: Addition of an ACTOR/DO Node

GOAL/PLAN and PLAN/GOAL, IF/THEN and THEN/IF, etc. See Figure 17 for a display of all the subordinators found in planning.

```
      GOAL/PLAN                AND          EXOR
       /     \                /|  \         /  \
      /       \              / |...\       /    \


     SEQ         NOT        IF/THEN      ACTOR/DO
     /|  \        |         /     \       /     \
    / |...\       |        /       \     /       \


      OR                  ACTOR/SAY/TO
     /|  \                 /    |    \
    / |...\               /     |     \
```

Figure 17: The Subordinators Found in Planning

7.2.2 Review of Work on Explanation

Closely related to this work on planning is some later work on the structure of explanation [Goguen, Linde & Weiner 81, Weiner 79, Weiner 80] showing that it too is a discourse unit, having similar structural properties and expressible in the same formalism. By **explanation** we here mean a specific discourse unit, with a describable formal structure; we do not mean any piece of discourse which serves the function of explaining something. Informally, an explanation is a discourse unit consisting of a statement about the world to be demonstrated, and a structure of supporting reasons, often with further embedded relationships of subordination. This kind of discourse occurs, for example, in social contexts where a single person attempts to justify to an addressee actions he has already performed, or will perform later.

Figure 18 shows an analysis of a simple explanation (actually a report of an explanation) in which the flight engineer reports his justification of the decision not to recycle the landing gear.

```
                 STATEMENT/REASON
                   /          \
                  /            \
                 /              \
      don't recycle             ALT
      the gear                 /   \
                              /     \
                             /       \
                            /         \
                           /           \
                 (not bent or     REASON/STATEMENT
                      broken)       /          \
                                   /            \ *
                                  OR          not able
                                  /\          to get it
                                 /  \         down
                                /    \
                            bent      broken
```

... **and I said we're reluctant to recycle the gear for fear something is bent or broken, and we won't be able to get it down** (1751:16)

Figure 18: An Explanation Tree

The most important relationship of subordination in explanation is indicated by the STATEMENT/REASON node. In the explanation displayed in Figure 18, the main STATEMENT is **Don't recycle the gear**. Everything which follows is a REASON supporting this. The ALT node represents the speaker's postulation of two alternate worlds, which differ by whether or not the landing gear is broken. This ALT node is established by the **underlined** portion of the following text: ...**we're reluctant to recycle the gear for fear something is bent or broken**. The phrase for fear indicates indicates both the uncertainty about whether the gear is bent, and the decision to treat the alternate world in which it is bent as the one on which attention is focussed.

Figure 19 shows the node types which are found in explanation. It includes EXAMPLE, which is not present in the example of Figure 18. EXAMPLE is a node which takes as its subordinates one or more examples of a statement.

```
STATEMENT/REASON          AND          OR          NOT
   /         \           / | \        /| \          |
  /           \         /  |  \      / |  \         |
 /             \       /   |...\    /  |...\        |

STATEMENT/CHALLENGE      IF/THEN     EXAMPLE        ALT
   /         \           /    \       / | \         / \
  /           \         /      \     /  |  \       /   \
 /             \       /        \   /   |...\     /     \
```

Figure 19: The Subordinators Found in Explanation

7.2.3 Static Versus Dynamic Information

The difference between planning and explanation in the cockpit and the type of planning and explanation previously studied lies in the type of information which is available to the participants. These previous studies, [Linde & Goguen 78, Goguen, Linde & Weiner 81], examined situations in which the information available to the group is static during the period of interaction. Although individual members may have new information to pass on to the group, there are no cases where information new to all members of the group enters during the process of planning or reasoning. In the case of planning, we may call this **static planning**.

The situations in the transcripts used in this study differ from static planning in two ways. One is that there is a predetermined flight plan, which is in force unless something unexpected happens. The existence of this flight plan (and associated manuals of standard procedure) means that normal goals and the plans and procedures for achieving them need not be stated, since they are known to all participants. Only new goals, and new plans which are not part of normal operating procedures need be stated explicitly. The second difference is that new information may be needed, and there may be planning to acquire this new information. Because of these differences, we call this type of planning **dynamic planning**.

The differences between static and dynamic planning can be handled by slight modifications of the previous theory. GOAL/PLAN nodes must be admitted into plans trees at some previously unexpected locations, in order to include plans for acquiring new information. The formalism must also recognize that some particular subplans may be suspended while some other physical or linguistic activity occurs, such as actually carrying out actions, or assessing the implications of newly received information. Some of these suspensions may involve the embedding of other discourse units, while others may involve breaking off a plan in progress to check something else, resulting in a discontinuous discourse unit.

A simple example is shown in Figure 20. In this situation, the plan already announced by the captain is to make an emergency landing in about ten minutes, if the passengers have been properly prepared. Execution of this plan to land is delayed until the readiness of the passengers has been determined. A plan is made by captain to acquire this information through an inspection by the flight engineer. The issue of when to land is dropped until four minutes later, when the flight engineer returns with a report on conditions in the cabin; during this period the captain and copilot work on a different plan about what to-do after landing and just how to land. (However, the issue of when to land is not immediately resumed.)

```
CAM-1  You might -- you might                PLAN/GOAL
       just take a walk back                  /     \
       through the cabin and                 /       \
       kinda see how things                 /         \
       are going  Okay?            take a walk       see how things
                                   back through      are going
                                   the cabin

CAM-1  I don't want to, I                    GOAL/PLAN
       don't want to hurry                    /     \
       'em but I'd like to                   /       \
       do it [land] in                      /         \
       another oh, ten               land in          IF/THEN
       minutes    (1757:21)        ten minutes         /    \
                                                      /      \
                                                     /        \
                                               PLAN/GOAL     [land]
                                                /     \
                                               /       \
                                              /         \
                                    take a walk        see how things
                                    back through       are going
                                    the cabin
```

Figure 20: Planning to Acquire Information

This distinction between the static and dynamic forms of planning is similarly applicable to explanation. It is also necessary to distinguish between explanation produced by a single speaker, and explanation produced by a group. Figure 21 shows the possible combinations of these two variables, plus one description of each type of case.

One form of a single speaker justifying something under conditions of static information is explanation as defined in [Goguen, Linde & Weiner 81]. This is produced essentially as a monologue, with perhaps minor evaluations or questions from the addressee. Many participants attempting to justify one or more propositions under conditions of static information produce what is commonly called argument. One speaker justifying something under conditions of dynamic information is what might be called "thinking out loud." In this situation, the speaker produces the "new" information himself, as he works out the implications of various approaches to a problem. Situations of this kind have been described by a number of researchers as the

		Information	
		Static	Dynamic
Participants	one	single speaker explanation	"thinking out loud"
	many	"argument"	group explanation of new information: the air crew case

Figure 21: Taxonomy of Explanation Types

paradigm case of "reasoning," produced by asking subjects to describe their "thinking process" out loud as they attempt to solve problems in mathematics or chess (e.g., [Newell & Simon 72].) Such protocols are extremely aberrant linguistically, since the speaker is not interactionally responsible to any other person or group. Finally, many speakers justifying or working out the implications of someting under conditions of dynamic information is represented by the air crew case.

In the case of explanation, the difference between the static and dynamic cases has to do only with the nature of information, not with the method for acquiring it, and so no new node types are required to extend to the theory from static to dynamic explanation.

7.3 Theory of Ratification

Plans are important in the aviation context because they are the major means allowing the crew to discuss possible actions. A crucial question about this process is how decisions about what actions to take are actually made and expressed. This is a complex social process, requiring appropriate communications among the individuals involved, and depending, in part, on the fact that there is a strict social hierarchy, in which all the participants are highly trained and are moreover legally responsible for the correctness of the decisions made.

Studying the execution of plans means understanding planning as part of the command and control system. From the command and control perspective, a plan is a directive whose propositional content contains possible actions. If such a directive is made by someone other than the captain, or by the captain as a suggestion rather than as an order, then it must be **ratified** before it has the social force of an action which the crew understands is to be performed. Since the final authority rests with the captain, all possible actions must flow through him for ratification. Examination of the transcripts shows that such ratifictions can be either explicit or implicit. Thus, an action proposed by someone other than the captain may be viewed as a **draft order**, which requires the captain's ratification to turn it into an actual order. Actions proposed but not ordered by the captain are more complex; they may receive approval or modification by crew members, and then flow back to the captain for actual ratification. Under this description, all ratified actions are seen as orders issuing from the captain.

This area is interesting because of its relevance to air crew coordination. A general problem here is how it can happen that important and relevant actions are not in fact taken. One specific form of this is that an appropriate action is actually proposed but then not ratified. The subsection below gives an informal discussion of the rules by which suggested actions become orders in planning discourse. There is also a brief discussion of how explanation might be treated similarly. It might be noted that this is an area of research for which the data set of transcripts used for this report is not rich enough to permit the construction of a complete theory; such a theory must wait until data is available from appropriate controlled simulator experiments.

7.3.1 Informal Rules for Plan Ratification

A natural way to move from a theory of planning to a theory of group decision making is to add rules for ratification to the rules for the construction of plans by a group that have already been found [Linde & Goguen 78]. Moreover, this should occur within the overall context of command and control discourse, that is, of speech act chains as discussed above. The sequence that produces first a proposed action and then its ratification can be seen as a complex (and possibly discontinuous) speech act.

The rules for ratification found in examining the current set of transcripts, may be stated informally as follows:

1. No action proposed by the captain need be ratified by the crew in order to become an order; but some actions may recieve such ratification. Explicit ratification by a crew member is *likely* if the captain has used an imperative form, and then may take the form of an acknowledgement. That is, acknowledgement of an order can be viewed as ratification by the crew member giving it, although such ratification is not required to give the directive the social force of an order.

2. An action proposed by a crew member *must* be ratified by the captain for it to become an order, unless:

   a. The captain or other crew member can be seen or heard to be performing the action immediately after the utterance of the order, or

   b. the action is not under the command of the captain (for example, if the action is personal, or if the captain has delegated authority).

3. Ratification of an entire plan counts as ratification of all the actions embedded in it.

4. An action proposed by a crew member is (provisionally) ratified if the captain subordinates other nodes to it.

5. A proposed action A below an EXOR ("exclusive or") node is ratified if the captain (or other relevant speaker in the case of delegated authority)

    a. explicitly negates the other branch, or

    b. ignoring the other branch, subordinates nodes to A (note that this is a special case of rule 4 above).

6. A plan will be ratified at its end (thus ratifying all its subordinate actions, by rule 3) *unless* it contains an action A such that

    a. A must be completed to obtain information needed for completion of the plan, or

    b. A is an urgent action, or

    c. A is subordinate to an intermediate GOAL/PLAN node, in which case only the subplan subordinate to that intermediate GOAL/PLAN node is ratified.

In terms of the command and control grammar given in Section 6.2, the utterances occurring in ratification are plans, supports, or challenges before ratification, and become requests by the captain afterward.

Note that a simple form of ratification also occurs in command and control speech act chains. In this case, a suggestion by a subordinate is followed by either an acknowledgement or a support by the captain, constituting a ratification, or by any other speech act, constituting at least a provisional failure of ratification. This form of ratification is handled by the command and control speech act grammar.

7.3.2 Explanation

It is important to be precise about the status of the various kinds of rule discussed in this report. The rules for plans represent constraints on the form of language. The rules for plan ratification are rules of interpretation for the move from language to social force; and the rules for command and control discourse represent constraints on the ordering and embedding of such social forces.

We propose that a similar set of rules is possible for reasoning. These rules would take some proposition about the world, and through ratification by the captain and other crew members, transform it into a shared belief about the world; i.e., into what currently counts as reality. Our transformational rules for explanation construction [Goguen, Linde & Weiner 81] would play the same role for explanation ratification that our rules for plan construction played above for plan ratification.

We have not yet pursued research in this area because it appears to be of somewhat lesser practical importance. However, it should be noted that the problem of an air crew "sticking" on a false hypothesis may fall into this area of constructing and agreeing upon shared belifs.

8 TOPIC SUCCESS AND TOPIC FAILURE

This section introduces one final theoretical concept required to understand CVR transcripts and to formulate hypotheses.

8.1 The Definition of Topic

Intuitively, topic refers to members' notion of "what the conversation is about" or "what we are talking about." More technically, the topic of an utterence concerns the propositional content of that utterance. As was discussed in Section 3, propositional content is independent of social force; thus, the following sentences all have the same propositional content, although they have quite different direct social forces.

(47) **The window is closed.**

(48) **Close the window.**

(49) **Is the window closed?**

(50) **I think it would be nice if the window were closed.**

In our discussion of propositional content, we distinguished the **specific** propositional content from the **general** propositional content of an utterance. Thus in the order

(51) **CAM-1 Give us three or four thousand pounds on top of zero fuel weight.**
(1750:30)

the general propositional content is **fuel weight**, while the specific propositional content is **three or four thousand pounds of fuel**. Thus, we may define the **topic** of an utterance (or sequence of utterances) to be the common general propositional content (if there is one).

Negation does not change major propositional content, although it reverses specific propositional content. Thus, (52) and (53) have opposite specific propositional contents but the same topic, **closure of the window**.

(52) **The window is closed.**

(53) **The window is not closed.**

8.1.1 A Taxonomy of Topics

General topics, or topic themes, can be listed and classified for this specific aviation domain. We expect that there are a limited number of these, since there are a limited number of factors which are of operational relevance to the flight mission, and that these topics can be organized into a taxonomy of topics. The topics which have been found in the data set of this study are shown in Figure 22.

Psycho-ostensives [Matisoff 79] are remarks whose primary function is to show the state of mind of the speaker; although they may have the form of requests or reports, they can not be carried out, or add nothing to what has been said previously. Some examples are:

STATE OF THE AIRCRAFT
- Power
- Fuel
- Weight
- State of Equipment

POSITION OF THE AIRCRAFT
- Altitude
- Heading
- Route and Course
- Location
- Airspeed
- Flight PLan

COMMAND AND CONTROL
- Routine Procedures
- Emergency Procedures
- Command and Delegation of Command

OUTSIDE COMMUNICATION
- Navigational Aids
- Visibility and Landmarks
- Communication Systems

HUMAN SYSTEMS
- State of Crew
- State of Passengers

OTHER
- Psycho-ostensives and Meta Remarks
- Non-operationally Relevant Remarks

STATE OF THE FLIGHT CONTEXT
- Location of Aircraft
- Weather
- Terrain
- Schedule
- Airport
- Takeoff Information and Clearance
- Landing Information and Clearance
- Change of Flight Plan
- Location of Other Aircraft

Figure 22: Taxonomy of Topics

```
(54) CAM-4  Less than three weeks to retirement, you better get me
            outta here.
                                (1748:17)

(55) CAM-2  Get this # on the ground.
                                        (1808:42)
```

Meta remarks are comments evaluating some utterance, or talking about talking about some topic. The above list of topics is not exhaustive. As we analyze further transcripts in detail, we expect that further topics will be found; but we also expect that this taxonomy will remain relatively small.

8.2 Topic Introduction and Topic Failure

The notion of topic permits us to define topic success and topic failure, notions that are of considerable importance for our analysis, because they allow us to track whether or not matters of operational relevance have been successfully brought to the attention of the crew.

We may view the first mention of a topic as an attempt by the speaker to introduce the topic to the group. If some other crew member produces an utterance on the same general topic, then the attempted introduction is a success. If no one does, then the attempt is a failure.

Note that this definition would count as successful a case where a topic is mentioned and its addressee verbally refuses to consider it or denies its relevance; this is deliberate. We are most concerned with cases where an attempt to introduce a topic receives no attention from the rest of the crew. In the case of a refusal, there is at least evidence that the topic has been attended to and considered, even if its relevance is finally denied.

Note also that success of a topic cannot be achieved by its speaker alone, but requires social interaction. This view of topic as an achievement of a social group is common to many discourse linguists who have worked with the notion of topic [Schegloff & Sachs 73, Keenan & Schieffelin 75, Polanyi 79].

We may also make a more delicate distinction between the operational success and the discourse success of a topic. **Operational topic success** is full success. A crew member introduces a topic, of operational relevance, and it is continued by other crew members in a way that is operationally relevant. **Discourse success** is a kind of false success -- the topic is continued but not in a way that is operationally relevant. (56) is an example of discourse success but operational failure:

(56a) CAM-2 **If we keep this up indefinitely, we'll be in Tulsa.**

(56b) CAM-1 **I haven't been in Tulsa in years.**
(Texas/Mena/73; 19:3434.5)

Here we may say that the most likely reading for the topic of (56a) is **We shouldn't keep this up indefinitely**. In (56b), a less likely interpretation of the topic **be in Tulsa** is continued, but operational relevance (what the crew should and should not do) has failed[4]. All discussions in this report of topic success refer to full operational topic success; discourse success is of little interest in this context because by definition, it is not operationally relevant.

---

[4] Those who have read the interim technical reports [Structural Semantics 82] for this project may note that this notion of topic failure generalizes our previous notion of goal formation failure. Goal formation failure was defined as the proposal of an action which could serve as a goal, without a plan being subordinated to it. For a number of reasons, we have replaced the notion of goal formation failure with the notion of topic introduction failure. The major reason for this change is to facilitate the statistical testing. The notion of topic failure given below includes far more cases than does the notion of goal formation failure, and should therefore permit far more reliable testing. A second reason for the change is that it should give greater inter-coder reliability. Goal formation failure requires that the coder determine that a plan could have followed some utterance and did not. This is a more subtle determination than whether or not two utterances are topically cohesive. However, the concept of topic failure accounts for the same intuitions as the initial concept of goal formation failure, and should lead to the same operational recommendations.

PART II:
HYPOTHESIS TESTING AND RESULTS

9 FORMULATION AND TESTING OF HYPOTHESES

This study attempts to deal in a rigorous empirical manner with linguistic data collected in a natural setting, the commercial air transport cockpit. This section is devoted to stating, testing, and discussing eight research hypotheses about the use of language in this setting. The experimental procedure and statistical methodology are also discussed in some detail; particular attention is paid to discussing generalizability of the results obtained. Section 9.2 gives a table and graphs summarizing the numerical structure of the sample, and Section 9.6 gives a table summarizing the results of testing each hypothesis.

9.1 Sampling Procedure

This subsection discusses how the sample studied in this research was obtained. There are three main stages to this process: (1) the production of accident transcripts, (2) the selection of transcripts, and (3) the coding of selected transcripts. The sample space that results from these procedures consists of a large number of speech acts, rather than, for example, a small number of transcripts or of crew members. This choice seems well suited for studying how linguistic behavior changes as a function of general features of the cockpit situation. On the other hand, accident transcript data is less suitable for studying individual differences in the behavior of crews or crew members, because these transcripts do not provide a sample of crews tested in a single standard situation, but rather show a single crew for each of several unique situations.

9.1.1 The Production of Accident Transcripts

When a commercial air transport accident involving a U.S. carrier occurs, the "black box" containing the last thirty minutes of cockpit conversation is routinely transcribed as part of the NTSB investigation into the causes of the accident. These CVR (Cockpit Voice-Recorder) tapes are not of outstandingly good accoustical quality, nor are the transcribers employed particularly expert in linguistic issues. However, it appears that these transcriptions are adequate for the purposes of this study. (We have not yet been able to compare the transcripts with the tapes, since only the transcripts are in the public domain. We hope to be able to make this comparison in later research.)

One beneficial property of this method of acquiring data is that it is "unobtrusive," that is (see Section 9.3.1) it is produced for reasons that have nothing to do with the researcher. This means that there is no possibility of any systematic effects due to bias of the researcher.

9.1.2 Transcript Selection Criteria

This subsection gives the criteria used for selecting the transcripts from which the speech act sample of this study was drawn. These criteria were developed using categories and analyses from [Murphy 80].

1. The transcript contains a critical segment. **A critical segment** is a portion of transcript containing observable degradation or failure of crew coordination which is actually or potentially critical to the completion of the flight.

2. The entire stuation of interest must not be significantly longer than 30 minutes (since the maximum length of the tape is 30 minutes).

3. There must be sufficient background information to permit understanding all relevant aspects of the situation.

4. The language of the transcript should be suitable for analysis. This means that there should be enough talk to permit analysis, and that all the conversation should be in English, since we are not focussing on cross-linguistic problems.

5. There should be sufficient interest and agreement in the aviation community to support further investigation.

6. All other things being equal, more recent transcripts are preferred. (Note that this criterion plays a major role in determining whether or not criterion 4 is satisfied; older flights are of lesser interest since the procedures and equipment are more likely to have been superseded.)

7. If possible, the set of transcripts should include all flight segments -- taxi, takeoff, climb, cruise, approach and land.

NASA personnel preselected a number of potentially suitable transcripts, using criteria 1 and 5, and 6 and 7 whenever possible. These eleven were examined in detail for inclusion in the dataset. They were:

1. United Airlines/Portland/78;
2. Eastern Airlines/Miami/72;
3. Northwest Orient Airlines/Thiells, New York/74;
4. Allegheny Airlines/Rochester/78;
5. World Airlines/Cold Bay, Alaska/73;
6. Texas International Airlines/Mena, Arkansas/73;
7. Pan American Airlines/Bali/74;
8. Air Florida/Washington, D.C./82;
9. Southern Airways/New Hope, Georgia/77;
10. PSA/San Diego/78; and
11. Pan American Airlines/Teneriffe/77.

Accident	1. Critical Segment	2. Events Thirty Minutes	3. Facts Known	4. Language Suitable	5. Community Interest	6. Recent
United/ Portland	X	X	X	X	X	X
Eastern/ Miami	X	X	X	X	X	X
NW Orient/ Thiells	X	X	X	X	X	X
Allegheny/ Rochester	X	X	X	X	X	X
World/ Cold Bay	X	X	X	X	X	X
Texas Int./ Mena, Ark.	X	X	X	X	X	X
Pan Am/ Bali	X	X	X	X	X	X
Air Florida/ Washington	X	X	X	X	X	X
==========						
Southern/ New Hope	--	X	--	X	X	X
PSA/ San Diego	X	X	--	X	X	X
Pan Am-KLM/ Tenerife	--	X	X	--	X	X

Figure 23: Criteria for Transcript Selection

Eight of the transcripts of this set are suitable for inclusion in the dataset. Figure 23 shows the satisfaction or failure of the selection criteria for each transcript. (Summaries of these transcripts are given in Appendix I.) The transcripts shown in this figure above the double line have been selected as suitable. Those below the double line are unsuitable for the following reasons:

1. **Southern/New Hope**. Several of the major contributing events occur before the beginning of the tape, and indeed, before departure, i.e. the company's failure to provide up-to-date severe weather information, and the crew's "lack of significant attempt to seek information on current flight conditions" (NTSB report, p. 33). In spite of the intrinsic interest of the situation, the transcript available does not contain a situation in which crew coordination is probably critical to the successful completion of the flight.

2. **PSA/San Diego**. The NTSB report on this accident mentions the possibility that there were two small planes in the vicintiy of the PSA plane, rather than just one, as both the crew and ground control appear to have believed. After completion of the NTSB report, there were newspaper reports that the pilot of a second small plane came forward and claimed to have been in the vicinity at that time. This puts into question some of the factual determinations of the accident report, since it is not possible to determine accurately to which plane the PSA crew and ground control were referring at any given time.

3. **Pan Am-KLM/Teneriffe**. Unlike the other accidents chosen for selection, the cause of this accident appears to lie in failure of air-to-ground communication, rather than in crew coordination. Futhermore, some of the communication problems appear to arise from the fact that three different languages are involved -- English, Spanish, and Dutch. While both these factors make this accident of great interest for a study of a different nature, this accident is so unlike the others in the dataset as to make the present methods of analysis unsuitable.

9.1.3 Data Coding Procedures

Although the selection procedure described above was applied to transcripts, the unit of coding and analysis is the speech act. Every speech act in the eight selected transcripts was coded according to the categories below. For each category, the value "unknown" is used when it is not possible to determine any other value. Moreover, many categories have a context condition that must be satisfied before meaningful coding is possible; if the condition is not satisfied, the code "not applicable" is used.

1. Speech act number. Speech acts were numbered sequentially within each transcript.

2. Speaker. The following numbers were used for speakers: 1 = captain, 2 = copilot, 3 = flight engineer, 4 = third officer, 5 = jump seat occupant, 6 = head flight attendent, 7 = other flight attendent. Alphabetic abbreviations were used for ground control, tower, approach control, etc.

3. Addressee. The conventions for speaker were also used for addressees.

4. Speech act type. The speech act types coded were request, report, acknowledgement, greeting, support, challenge, declaration, and psycho-ostensive.

5. Discourse type. Discourse types coded were command and control chain, checklist (which

is a special kind of command and control chain), plan, explanation, narrative, and pseudo-narrative.

6. New topic. Each speech act was coded for whether it introduced a new topic. (See Section 8 for a definition.) This variable was coded with values true, false, not applicable, and unknown.

7. Topic success. Each speech act which took the value "true" for new topic was coded for whether or not this topic succeeded, where topic success was defined as use of the topic by any other next speaker. This variable was coded with values true, false, not applicable, and unknown.

8. Draft order. Every request by a subordinate was coded for whether it expressed a draft order. (See Section 6.3 for the definition of draft order.) This variable was coded with values true, false, not applicable, and unknown.

9. Ratification. Every draft order was coded for whether it was ratified by the captain. This variable was coded with values true, false, not applicable, and unknown.

10. Mitigation level. All requests and reports were coded for mitigation level. This variable was coded for the values aggravated, direct, low mitigated, and high mitigated, abbreviated A, D, LM and HM in the coding sheets and in the frequency tables given below. In the case of a sentence which was mitigated by following sentences, the sentence was coded as its own mitigation value plus one.

11. Crew Recognized Emergency. Each speech act was coded for whether it occurred during a crew recognized emergency. (See Section 5.1 for a definition.) This variable was coded for the values true, false, and unknown.

12. Crew Recognized Problem. Each speech act was coded for whether it occurred during a crew recognized problem. (See Section 5.2 for a definition.) This variable was coded for the values true, false, and unknown. Any speech act occurring during a crew recognized emergency by definition also occured during a crew recognized problem.

13. Operationally relevant. Each speech act was coded for whether or not it was operationally relevant to the completion of the flight. (See Section 5.3 for a definition.) This variable was coded with values true, false, and unknown.

14. Comment (optional). If in the opinion of the analyst, the speech act exhibited some special feature which might be of interest in future studies, a comment marking that feature was added. (For example, sentences containing profanity were commented, because this feature may be of interest in future studies of mitigation and aggravation.)

These data were entered into a separate computer file for each transcript. These files were then run through a program checking consistency with the coding conventions, and were manually corrected. Then, for each hypothesis, the files were run through a specially written program to extract the data needed for testing that hypothesis. For several of the hypotheses, auxiliary data were also printed to permit reference back to the transcripts in order to check the accuracy of the process and to enhance the researchers' understanding. Finally, for each hypothesis, the data were tabulated, aggregated, and subjected to the relevant statistical test; for the hypotheses given here, either Student's t test or the χ^2 test was employed.

9.2 Numerical Overview of the Sample

This subsection provides a general overview of the structure of the sample.

		Operationally Relevant Speech Acts						
transcript	length	N1	N2	N3	N4	N5	N0	total
Portland	343	91	41	56	0	4	9	201
Miami	158	44	28	13	0	4	3	92
Thiells	189	43	30	5	0	0	2	80
Rochester	71	10	10	0	0	0	1	21
Cold Bay	179	65	33	9	0	2	0	109
Mena	223	101	97	0	0	0	0	198
Bali	209	33	17	3	7	0	1	61
Washington	353	43	74	0	0	0	0	117
sums	1725	430	330	86	7	10	16	879

Figure 24: Operationally Relevant Speech Acts by Speaker

There are altogether 1725 speech acts in this collection of eight accident transcripts, 879 of which are operationally relevant. Figure 24 shows the number of operationally relevant speech acts, by speaker, in each transcript. The first column names the transcript (by city), and the second gives the total numbers of speech acts in that transcript. The next six columns give the number of operatonally relevant speech acts for each crew member in each transcript: the N1 column gives the number of speech acts produced by captains; N2 by first officers; N3 by flight engineers; N4 by third officers; N5 by jump seat occupants; and N0 by those denoted "?" in the transcripts. (No attempt has been made to improve or correct the attributions given by the transcribers, although there are certainly cases where this could be justified.) The total number of operationally relevant speech acts in each transcript is given in the final column. The total number of operationally relevant speech acts in the eight transcripts is 879. There are altogether 25 crew members, including 8 captains, 8 first officers, and 5 flight engineers.

One use of Figure 24 is to identify the most loquacious speaker of each rank. The most loquacious captain and first officer are both in the Texas/Mena/73 transcript. The most loquacious flight engineer is in the United/Portland/78 transcript. This information is used in Section 9.3.4 to examine individual differences between speakers.

Another use of Figure 24 is to determine the frequency distribution of speech acts by speaker for each rank. These will show, for example, whether or not some few speakers are responsible for a majority of the speech acts in the sample. We would expect that each rank would show an approximately normal distribution of numbers of speech acts; this will increase our confidence that we have a random sample of speech acts. These distributions are presented as bar graphs in Figure 25. Here, the number of speakers producing between 1 and 10 speech acts is indicated by the leftmost bar, those producing between 11 and 20 by the next bar, and so on. It will be seen that the mean number of speech acts produced decreases strictly with rank, and that captains and first officers are closer together than any other two ranks. It will also be seen that for captains, who are the most experienced group of speakers, the frequency distribution is a reasonable approximation to a normal curve. For first officers, there is also a reasonable approximation. For flight engineers, there seems not to be a very good approximation, because the flight engineer in United/Portland/78 transcript produced twice as many speech acts as the next most loquacious flight engineer. For the other categories, there are too few speakers to be certain, but the distributions certainly appear to be reasonable approximations to normal curves.

It should be noted that the number of speech acts used for testing any particular hypothesis is generally less than that given in Figure 24. For example, in testing a hypothesis involving mitigation level, attention must be restricted to speech acts having a determinable mitigation level.

9.3 Representativeness of the Sample

We now discuss the **generalizability** of our results from the eight specific transcripts selected to the broader population of commercial aviation cockpit discourse. The results will generalize provided that the sample is representative. This subsection presents three arguments for the representativeness of our sample.

The first and most basic argument is that a sample is very likely to be representative if it is a random sample and is also sufficiently large; in fact, the probability that a random sample is not representative can be made as small as desired by making the sample large enough. For this reason, Section 9.3.2 gives arguments for believing that our sample is a random sample, and Section 9.3.3 argues that the sample is sufficiently large.

A second argument for representativeness, given in Section 9.3.4, is based on the fact that the sample can be successfully used as a standard of comparison for the behavior of crew members.

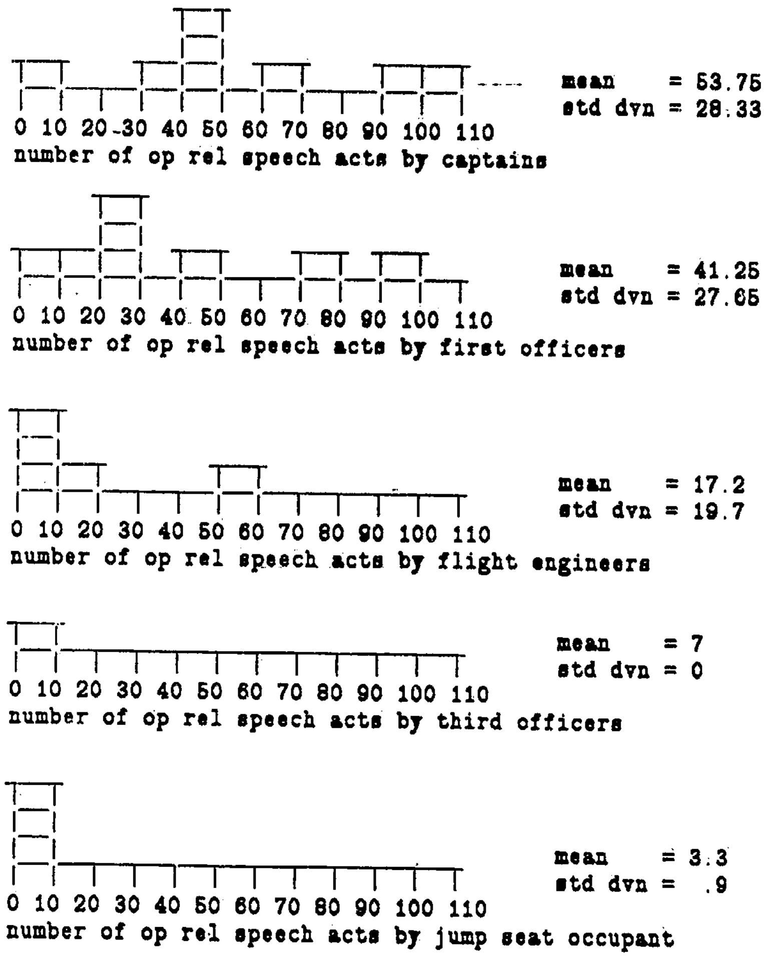


Figure 25: Operationally Relevant Speech Acts by Rank of Speaker

A third point, developed in Section 9.3.5, regards our use of a control subset of the sample for testing hypotheses originally formulated by examining a completely different subset of transcripts. This reduces the likelihood that the result obtained from testing a given hypothesis is due to some uncontrolled variable, different from the independent variable of the hypothesis in question.

Finally, Section 9.3.6 discusses of the status of these arguments. Briefly, they should not be regarded as conclusive, but rather as suggestive. The results of the statistical tests on research hypotheses in this study are clearly valid as *descriptive* statistics, that is, as statistical summaries of a particular sample. Moreover, if the arguments for generalizability are accepted, then the results can be given the usual *inferential* interpretation.

Of course, this study is limited by the origin of its data in accident transcripts, so that it is not clear exactly which aspects may generalize to non-accident transcripts. Consequently, it would be very interesting to study non-accident transcripts, either with data from simulator experiments, or better, with nonobtrusive data from non-accident flights.

9.3.1 Methodological Background

Because this report is a first study in a new area, we have chosen to discuss certain basic statistical issues in some detail, in order to clarify the assumptions and methods which serve as foundations for the study.

We first introduce a basic trichotomy of possible types of data collection, based on [Bowen & Weisberg 80]:

1. **Experimental** - conducted under laboratory conditions with manipulated independent variables.

2. **Sample** - a random subset of a given population collected in the field.

3. **Unobtrusive** - collected for reasons having nothing to do with the researcher, using nonreactive measures.

The data of this study clearly falls within the third category, and can also be argued to fall within the second (see below). In order to further discuss these categories, we introduce three particular issues concerning the **quality** of research. These issues are:

1. **Quality of Measurement**: Does a measurement procedure really give results that correspond to what the researcher wants to know? Three aspects of this issue are as follows:

   a. **Reliability** - Can the outcomes of the measurement procedure be reproduced tolerably well?

   b. **Validity** - Does the measurement procedure actually measure the construct of interest to the researcher?

   c. **Lack of Bias** - Does the measurement procedure systematically affect the resulting value?

2. **Control**: Are we sure that the observed results are not attributable to some other variables?

3. **Representation**: Do the results obtained generalize to the population as a whole?

We now compare the three modes of data collection with respect to the above criteria.

1. Experiments excel in control, but when social variables are involved, they can be very

weak on representation. Note that linguistic variables are especially sensitive to aspects of the data gathering situation.

2. Sampling excels in representation, and sample data can be obtained in more natural conditions than the lab; but sampling is weak on control.

3. Unobtrusive data cannot be affected by the conditions of measurement; the prime difficulty is that the measurements that the researcher really wants may be unavailable. Possible problems with control and representation imply that the population of interest, the sample involved, and the variables used, should all be carefully delineated.

Unobtrusive data are more valuable for studies of social variables because of the ubiquity of bias introduced by measurement. (This problem is an analog of the Heisenberg Uncertainty Principle. It has been stated for linguistics as the *Observer's Paradox*: "The aim of linguistic research ... must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation." [Labov 70].) Fortunately, unobtrusive data are available for the study of cockpit discourse, and are especially appropriate for the present research, which is primarily concerned with the role of social variables. Two such goals for this study are to identify potentially trainable linguistic phenomena, and to discover linguistic correlates or predictors for variables such as vigilance. A longer range goal is to develop criteria for the design of aviation procedures and equipment that involve the use of language.

9.3.2 Is the Dataset a Random Sample?

Underlying any use of statistical methodology is the basic question of whether or not the data used is really a random sample from a population. Our basic argument for the representativeness of the sample depends on this point, as does the applicability of the statistical testing reported in Section 9.5. Below, we give three different and mutually supporting arguments for believing that our dataset is a random sample. This belief is also reinforced by the homogeneity of the sample, as discussed in Section 9.3.4.

9.3.2.1. Statistical Independence of Transcript Selection Criteria

The most basic argument is that the criteria that were actually used for transcript selection are in fact *statistically independent* of the dependent measures used in the hypotheses. For example, it seems clear that whether or not a critical segment occurs in a transcript cannot effect the mitigation level of speech acts occurring in that transcript; the same argument can be made for all the other selection criteria and the other dependent variable: occurrence or non-occurrence of given speech act in a planning or reasoning discourse unit. (The criteria for transcript selection are given in Section 9.1.2 above.) Independence of these variables implies that the sample of speech acts in the chosen transcripts cannot have been biased by the transcript selection process.

9.3.2.2 Locality of Effects

The speech acts in our sample were not draw at random from a larger population of aviation speech acts, but rather were taken as they occurred, in sequential order within the selected transcripts. This raises the question of the possible effects of sequential dependencies. Language clearly does exhibit sequential dependencies at many levels. For example in English, when we see a *q*, we know that it will be followed by a *u*, and when we see *the*, we know that it will be followed by an adjective or a noun. The question is whether or not the effects of these sequential dependencies mean that we cannot obtain a random sample in this manner.

It is a general fact about language that although sequential dependencies do exist, their effects are largely confined to immediately adjacent units, and hence have little effect on the randomness of any reasonably large sample. To state this more precisely, the conditional probabilities $P\{f(n) \mid f(n-1),\ldots,f(n-k)\}$ of the $n^{\underline{th}}$ unit $f(n)$ given the previous k units, $f(n-1),\ldots,f(n-k)$, in general show very little dependence on units further than two or three earlier in the sequence. We call this the principle of **locality of effects**. Another way to state this principle is that "action at a distance" is very limited in language.

It must be assumed in this discussion that all the units involved are at the same linguistic level, for example, that they are all phonemes, or all morphemes, or all speech acts. For example, given a sentence containing a simple past tense main verb, we can make no prediction, or only a very weak prediction, about the form of the following sentence. However, if we also have the higher level information that the sentence forms part of a narrative, then we can make a much stronger prediction about the form of the following sentence -- that it too will probably have a past tense main verb. Of course, such higher level information is often available and valuable in doing linguistic analysis; but the restriction is reasonable for our hypotheses, which do not in fact involve variables on more than one linguistic level, and thus the argument is applicable.

9.3.2.3 Experience with Other Linguistic Data

There is a great deal of experience with random sampling of linguistic populations, for example with stylometric statistics, and it has been found emprically that selection procedures have surprisingly little effect for reasonably large samples; for example, [Herdan 66] speaks of the "remarkable fact of the stability of frequencies of ... linguistic forms." This stability has been observed in many different languages and historical periods for phonemic, lexical, morphological, syntactic and metrical levels of linguistic structure. The latter levels two present strong analogies with the discourse level structures with which the present study is primarily concerned. This argument for stability is further supported by the fact that the locality of effect principle holds for speech acts, just as it does for other linguistic forms.

9.3.3 Sample Size

Experience with statistical studies of other linguistic data suggests that samples of size one or two hundred units are generally adequate [Herdan 66], and smaller samples will of often do for phenomena that are not especially subtle. Thus (see Figure 45 in Section 9.6), there is only one hypothesis that might be in doubt onthe grounds of sample size, Hypothesis 8. However, as this hypothesis does not appear to be especially subtle, there seems to be no cause for more than raising a mild cautionary flag in connection with the result of testing this hypothesis.

In sociolinguistics, it is common practice to aggregate data from a number of different speakers from the same speech community. The experience of this research is that as long as attention is restricted to phenomena that really are characteristic of a speech community as a whole, there is little difficulty with individual differences, provided that the sample of linguistic units is large enough [Guy 80].

It might be thought that the data used in this study consists of utterances produced by too small a number of different speakers (25) to constitute a truly random sample. This is undoubtedly true at a sufficiently detailed level of analysis, where individual differences become a major interest. However, Section 9.3.4 argues that many linguistic phonemena of potential interest in the study of aviation safety are characteristic of the commercial air transport crew community as a whole, and are relatively independent of speaker (for native speakers of the same language).

9.3.4 The Sample is Homogeneous

Extensive catalogues of the frequencies of many different linguistic structures, from several different languages and historical periods, have been collected (see [Herdan 66]); these frequency distributions have been found to be so stable that it is possible to identify individuals who differ significantly from the average [Labov 70]. This research experience also suggests that our sample of 879 operationally relevant speech acts is certainly large enough.

Since we have aggregated data from a number of speakers, it might be questioned whether the sample is dominated by a few loquacious speakers who exhibit unusual linguistic behavior. To support the assertion that individual differences are relatively unimportant in this sample, compared to systematic differences arising out of the cockpit situation in which the language is produced, we may test whether or not a selected individual speaker's behavior differs significantly from that of his colleagues of the same rank in regard to some variable of interest. We have chosen the most important, and perhaps the most sensitive, measure used in the research reported here, namely degree of mitigation/aggravation. Comparing mitigation/aggravation of operationally relevant, non-checklist requests from the most loquacious captain (in the Texas/Mena/73 transcript) with the aggregation of all seven other captains yields the frequency data shown in Figure 26.

	mitigation level					
case	A	D	LM	HM	total	mean
Mena capt	0	63	13	1	77	.195
other capts	8	157	22	5	192	.125
sums	8	220	35	6	269	

Figure 26: Comparative Mitigation/Aggravation Frequencies for Captains

Using Student's t test with null a hypothesis of no difference in mean mitigation/aggravation score yields t=1.08 (df=267, p=.14). Using the χ^2 test yields χ^2=4.87 (df=3, p~.18). Thus, using either test, the null hypothesis must be accepted, and we conclude that there is no significant difference. (A rather detailed discussion of the applicability of these tests to the present data is given in Sections 9.4.3 and 9.4.4.)

	mitigation level					
case	A	D	LM	HM	total	mean
Mena FO	3	58	17	2	80	.225
other FOs	10	125	38	1	174	.172
sums	13	183	53	3	254	

Figure 27: Comparative Mitigation/Aggravation Frequencies for First Officers

The same thing can be done for first officers. Again, the most loquacious occurs in the Texas/Mena/73 transcript. Figure 27 shows the mitigation/aggravation frequencies for operationally relevant non-checklist requests to captains by first officers. This data yields t=.735 (df=252, p=.230) and χ^2=2.16 (df=3, p~.5) for the null hypotheses of no difference. Once again, the null hypothesis is rejected, and we conclude that there is no significanct difference between the mitigation/aggravation scores of this first officer and the aggregated score of the other seven first officers.

It seems less reasonable to do the same test for flight engineers, as there are far fewer speech acts involved. However, it does make sense to try pairwise comparisons between officers of the same rank. We have done a few of these at random, and many of them show no significant difference, although others do show a difference. In general, the differences picked out are confirmed through reference to the transcript and NTSB report, and this also supports the homogeneity of the sample.

The other way that a homogeneous sample can be used is to identify individuals whose behavior is significantly unusual. Let us now consider an example of this phenomenon. Figure 24 shows that the first officer in the Air Florida/Washington, D.C./82 transcript has approximately twice as many speech acts as his captain, whereas in the other seven transcripts, the captain has at least as many speech acts as his first officer. Testing the difference in mitigation level between this first officer and the seven others shows a significant difference: his speech acts are more mitigated. One may conjecture (and the press has done so) that he was so loquacious because he was nervous about the situation. However, he was not assertive about his concerns; on the contrary, he maintained a relatively high level of mitigation in his speech.

To summarize, we have shown that, for speakers of a given rank, the sample is not dominated by a few speakers with unusual linguistic behavior (athough the question is left open for flight engineers). More than this, we have given an instance where a significant difference between one speaker and the aggregated speech acts of the other speakers of the same rank corresponds to what appears to be significantly unusual behavior from this individual, in an unusual situation. (We have also found other instances of this phonemenon, not reported here.) This supports the view that the sample is sufficiently representative to serve as a meaningful standard of comparison for determining significant individual differences. We regard this as strongly suggestive evidence for the representativeness (and randomness) of the sample. For, if the sample were significantly nonrepresentative in regard to the dependent measures used in this study, then statistically significant individual differences from the sample as a whole would not always correspond to intuitively significant differences in behavior or situation. (Of course, this is not a rigorous statistical argument since it relies upon the judgement of the analysts.)

It would be interesting to perform similar studies for the other independent variable used in this study, namely the frequency of planning and explanation, but we have not yet done so.

9.3.5 Use of Control and Test Transcripts

We now discuss the division of transcripts into two groups. As stated in Section 9.1, this study is based on speech acts from eight transcripts of commercial air transport accidents. Two of these transcripts, chosen for the interest of their language and situation, were closely examined to seek hypotheses which either illuminate the basic structure of the transcripts, or else which have practical implications. We call these two transcripts, United/Portland/78 and Texas/Mena/73, the **hypothesis formulation group**. The remaining six transcripts were used to test the hypothesis; we call these transcripts the **test group**.

The six transcripts from the test group contain altogether 480 operationally relevant speech acts, while the two hypothesis formulation transcripts contain 399. Thus the eight transcripts from both groups contain a total of 879 operationally relevant speech acts. Each hypothesis selects, as a dataset for testing, a subset of the 399 speech acts of the hypothesis formulation group and a disjoint subset of the 480 speech acts of the test group. For example, the first hypothesis has as its dataset from any given transcript, all non-checklist operationally relevant requests having a defined mitigation level, where both speaker and addressee are crew members.

Each hypotheses is first tested on speech acts from the six transcripts of the test group. It is then tested on the speech acts from the two hypothesis formulation transcripts. Speech acts from these two groups are pooled when possible to yield a larger sample for a stronger test of the hypotheses. However, pooling is justified *only* if it is possible to avoid the methodological bias that results from testing hypotheses on the data from which they were formulated. For purposes of this study, the only case in which the two sets of speech acts cannot be pooled is that in which the hypothesis is accepted for data from the two hypothesis formulation transcripts, but is rejected for data from the six test transcripts. If the hypothesis is accepted for data from the six test transcripts and/or is rejected for data from the two hypothesis formulation transcripts, then the two datasets can be combined.

The purpose of this division is to reduce the probability that the obtained results are in actuality due to the effects of some uncontrolled variable.

9.3.6 Discussion

The results of the statistical tests performed in this study are clearly valid as descriptions of the properties of a particular sample. The arguments given earlier in this subsection support the view that this sample may be reasonably representative of the entire population of commercial air transport crew speech acts. We do not regard these arguments as either conclusive or definitive, but we find them fairly convincing, and in any case, interesting as an exploration of the assumptions required to support generalizability of the results.

The issue of representativeness could also be subjected to direct empirical study. The generalizability of our sample to the population of aviation speech acts as a whole could be studied by choosing at random a set of transcripts different from those used here, and then testing the most significant hypotheses on speech acts from those transcripts. (A similar study is reported in Section 9.3.4, showing that some parts of the present sample do not differ significantly from the whole. Of course, this is does not prove representativeness; but homogeneity of the present sample is suggestive evidence in favor of homogeneity, at the same level of granularity of analysis, of the entire population.)

9.4 Formulation of Hypotheses and Choice of Statistical Tests

This subsection precisely formulates the null hypothesis and dataset involved in each of the eight research hypotheses, and also discusses the statistical tests and level of significance used. The results of each test are given in Section 9.5 together with some discussion of the implications; these results are summarized in Section 9.6. The implications of the body of results as a whole are discussed in Section 11. The choice of hypotheses to be tested was influenced by the pioneering work of [Foushee & Manos 81].

9.4.1 Formulation of Null Hypotheses and Dataset Definitions

Eight research hypotheses have been chosen for testing on speech acts from aviation accident transcripts. These hypotheses concern the role in aviation discourse of the concepts and variables developed in this report. The eight hypotheses follow: first an informal statement of each research hypothesis is given in **boldface**; then a precise formulation of the null hypothesis actually used in the statistical test is given; also the subset of speech acts used as a dataset for the hypothesis is defined. (Section 9.3.5 discusses how the eight transcripts listed in Section 9.1.2 are divided into two subsets for testing each hypothesis.)

Each hypothesis is restricted to speech acts whose speaker and addressee are both crew members, because we are not studying air-to-ground communication, nor are we studying communication with flight attendants or passengers. They are restricted to operationally relevant speech acts because there is more linguistic variation in the non-operationally relevant portions of the text, and because non-operationally relevant speech acts are less important for our purpose. Checklist speech acts are excluded because checklist activity is highly stereotyped; in particular, these speech acts are almost always direct and almost never acknowledged. These restrictions apply to all eight research hypotheses and are not repeated for each one separately.

A requirement that does vary among hypotheses is the nature of well-definedness for the variables occurring in that hypothesis. For example, speech acts with unknown speaker cannot be used in testing hypotheses that involve speaker rank.

1. **Requests to superiors are more mitigated.** The null hypothesis is that the mean mitigation/aggravation score for requests to superiors equals the mean score for requests to subordinates. The mitigation/aggravation score is computed using weights -1 for aggravated, 0 for direct, 1 for low mitigation, and 2 for high mitigation (see Section 4 and also the discussion of condition (3) in Section 9.4.3); the same weights are used in each subsequent hypothesis involving the mitigation/aggravation scale.

2. **Requests are less mitigated in Crew Recognized Emergencies.** The null hypothesis here is that the mean mitigation/aggravation score for requests in CRE equals the mean mitigation/aggravation score for requests not in CRE.

3. **Requests are less mitigated in Crew Recognized Problems.** The null hypothesis is that the mean mitigation/aggravation score for requests in CRP equals the mean mitigation/aggravation score for requsts not in CRP.

4. **Subordinates plan and explain more often than superiors.** The null hypothesis is that the percentage of speech acts in explanation and planning discourse units produced by subordinates equals the percentage produced by superiors.

5. **Planning and reasoning are less common in Crew Recognized Emergencies.** The null hypothesis is that the percentage of speech acts that occur in planning and reasoning discourse units in CRE equals the percentage that occur in non-CRE.

6. **Planning and reasoning are more common in Crew Recognized Problems.** The null hypothesis is that the percentage of speech acts that occur in planning and reasoning discourse units in CRP equals the percentage in non-CRP.

7. **Topic-failed speech acts are more mitigated than topic-successful speech acts.** The null hypothesis is that the mean mitigation/aggravation score for speech acts whose topic has failed equals that for speech acts whose topic has succeeded.

8. **Unratified draft orders are more mitigated than ratified draft orders.** The null hypothesis is that the mean mitigation/aggravation score for draft orders that are not ratified equals the mean for draft orders that are ratified.

A number of other hypotheses were formulated in our second interim technical report [Structural Semantics 82]; however, it was found that these could not be tested with the present dataset, because the events involved, such as speech act misunderstanding, were found to be too rare.

9.4.2 Level of Significance

The reader who is not familiar with statistical research in linguistics and sociology should note that verifying hypotheses in these areas is in general more difficult than verifying hypotheses about physical science data, and that a .05 level-of significance is standard in the literature [Herdan 66]. We have adopted this convention, but it should be noted that a significance level of .03 would have sufficed for all the hypotheses actually accepted here.

9.4.3 Assumptions Underlying Use of the t Test

Only two statistics have been used for testing the hypotheses in this report: Student's t statistic and the χ^2 statistic. Both statistics are used for testing whether or not two samples differ significantly in regard to the values of some variable. The choice of statistic for testing a given hypothesis is determined by whether or not certain assumptions are satisfied by the data. It should be noted that modern statistical practice has found both of these statistics to be remarkably robust, so that only approximate satisfaction of their underlying assumptions is required [Bowen & Weisberg 80]. Whenever it is appropriate, Student's t statistic is preferable to the χ^2 statistic, because the t statistic is more powerful, that is, it will yield a more definitive decision on the same data.

According to the classical view (e.g., [Siegel 56]), appropriateness of the t statistic depends upon approximate satisfaction of four conditions:

(1) the dependent variable has a normal distribution for each of the two populations being compared;
(2) these distributions have equal variance;
(3) the two samples being compared are independent; and
(4) the dependent variable has values on an interval scale.

We will now discuss each of these assumptions in relation to the data involved in this study, and in the light of more modern views. Assumption (1) is usually valid for reasonably large samples, and in fact is satisfied by the mitigation scores examined below. Regarding assumption (2), we have computed the variances of each sample for all the hypotheses tested in this report, and have observed that they are approximately equal. (This could be tested using the F statistic, but we have not done so.)

The independency assumption (3) is more problematic because our units of analysis are speech acts rather than individuals. For some hypotheses, the speech acts in the samples compared are generated by different individuals, while for others they are generated by the same individuals in different situations. We have therefore used computational formulas for related- or single- sample (i.e., pooled variance) comparisons. (However, the outcomes should be virtually identical to those for independent sample test procedures.)

The role of assumption (4) is very controversial in the psychology literature, and many writers do not believe that it is necessary [Gaito 80]. Before discussing this issue in more detail, let us define four possible levels of scaling, following [Siegel 56]:

1. **Nominal Scale**: Arithmetically the weakest level of scaling, it is characterized by the use of values only as labels or classifications for objects, persons, characteristics, or events. The only admissible operation is testing equivalence of classified entities. For example, if numbers are assigned to discourse types (such as 1 for planning, 2 for reasoning, 3 for command and control, etc.), then it makes sense to ask whether two speech acts A1 and A2 are equivalent in the sense that they occur in the same kind of discourse type; it does not make sense to ask whether A1 is less than A2.

2. **Ordinal Scale**: Measures are ordinal when the values that are used to label entities can be ordered. For example, speakers in the cockpit have an an established rank, and the integers assigned to speakers (see Section 9.1.3) reflect this ordering; the lower the integer, the higher the rank. However, it does not make sense (in terms of what the numbers represent) to add two ranks, or to ask what is the average rank of a group of speakers.

3. **Interval Scale**: When a scale has the properties of an ordinal scale, and in addition it makes sense to measure and compare the *distance* between any two points on the scale, then we have a much stronger type of scaling, called **interval** scaling. The unit of measurement and the zero point are arbitrary for interval scales, in the sense that the value of any statistic (such as Student's t statistic) that is valid for interval scales will have *exactly the same* value for *any* choice of unit of measure and zero point. We argue later that the scale of mitigation/aggravation given in Section 4 may be of this type. The unit of measurement there was taken to be 1 and the zero point was taken to be "direct." Thus, the distance between "direct" and "high mitigation" is two units, and is thus equal to the distance between "aggravated" and "low mitigation." (Note that assigning the numerical values 1, 3, 5, and 7 to the four points on the scale, instead of the values -1, 0, 1 and 2 that were actually used, would make no difference in the obtained probability levels in testing the hypotheses that follow, because the t statistic will have exactly the same value.)

4. **Ratio Scale**: A scale that has the properties of an interval scale and in addition has a true zero point is a **ratio scale**. Mass or weight is an example of such a scale. The unit is still arbitrary (e.g., pounds or grams may be chosen), but an object of zero mass is still of zero mass whatever unit may be chosen. None of the measures used in this research are ratio scales.

We now argue that the mitigation/aggravation levels of speech acts approximate an interval scale, specifically a scale of just noticeable differences of mitigation/aggravation. If this argument is accepted, then assumption (4) is satisfied whenever the dependent variable is mitigation/aggravation score, and therefore the t test can be used for all hypotheses except 4, 5 and 6. To show that the intervals of the scale of mitigation/aggravation are "jnd's," trial studies were run using two scales having more levels of both mitigation and of aggravation, a first with three levels of each, and a second with one level of aggravation and three levels of mitigation; both had a single "direct" level. It was found that reliable coding could not be achieved using these finer scales. This suggests that the four level scale finally shown to be reliable (see Section 4) is a scale of "jnd's of mitigation/aggravation level." If this is the case, then the scale of mitigation/aggravation is an interval scale whose unit is one jnd of mitigation/aggravation. We do not regard this argument as entirely conclusive, because the earlier attempts at reliable scaling with more levels were not as rigorous as our final experiment, and there was no attempt to determine directly whether or not these levels are really jnd's. (It is also possible to test whether or not members of the aviation community perceive this scale to have equal distances between its levels; however, we have not done so.)

On the other hand, we would like to follow [Gaito 80] and others in claiming that use of the t test does not require satisfaction of the interval scale assumption[5] While the considerable successful experience with parametric statistics on non-interval data cited in the literature supports this claim, still we feel it necessary to justify the assignment of weights to mitigation levels that was used (-1 to aggravated, 0 to direct, etc.). Perhaps the above discussion of jnd's will serve as such a justification, even if it is not accepted that these levels consistute a true interval scale.

The reader who does not accept the above arguments may prefer to see the results of the χ^2 test for each hypothesis. These are given in Section 9.6, in a table summarizing the results of each test.

Student's t test uses as its null hypothesis that two distributions have the same mean. For the so-called "one tailed" test, to reject the null hypothesis is to assert that the means differ in a specified direction. (The "two tailed" t test asserts only that the two means are significantly different, without regard to the direction of difference; but only the one tailed test is used in this research.)

It might be noted that in general because of the relatively large size of our samples, we can make use of the normal approximation to the distribution of the t statistic. There is the only case where small sample statistics are needed; that is in testing Hypothesis 8 on the hypothesis formulation subset, that contains only 15 speech acts.

9.4.4 Assumptions Underlying Use of the χ^2 Test

The χ^2 test must be used for Hypotheses 4, 5 and 6 because the dependent variable used in these hypotheses takes the two values "planning or reasoning" and "not planning or reasoning." These two values do not form an interval scale (in fact, they do not even form an ordinal scale, but only a nominal-scale, because it makes no sense to ask whether "planning or reasoning" is greater than "not planning or reasoning"). Even if one accepts the use of parametric statistics on non-interval data, there still does not appear to be any sensible way to assign numbers to these two values, so it does not make sense to compute their means or standard deviations. Therefore the t test cannot be used, and we must use the χ^2 test.

There is no controversy about using the χ^2 statistic with measures that form only a nominal

---

[5] Gaito quotes Lord, "The numbers do not know where they come from."

scale, that is, a set of discrete categories. The only assumption that needs to be satisfied is that the samples of the two distributions are independent. There is no difficulty about this when the independent variable is rank, since the sets of speakers are then disjoint in the two groups being tested for difference; this justifies the use of this test for Hypothesis 4. For Hypotheses 5 and 6, the independent variables are CRE/non-CRE and CRP/non-CRP, respectively. We are unable to give a definitive justification for the applicability of the χ^2 test for these hypotheses, although we can give an argument that may be reasonably convincing: because of the relative stability of linguistic frequency distributions, the relatively large numbers of speech acts and speakers, and their relative independence of speaker[6], especially for such a close-knit community as commercial air transport crews (see Section 9.3.4), it may be expected that the average rate of planning or explanation (which is the dependent variable) over a number of individuals will also be stable.

The χ^2 test uses as its null hypothesis that two distributions are the same. To reject the null hypothesis is to conclude that the two distributions are in fact significantly different. Hypotheses 4, 5 and 6 each assert that two distributions differ in a specific way; in fact, each distribution is characterized by a single frequency, and these hypotheses each assert that that frequency is greater for one value of the independent variable than for another value. This is a stronger hypothesis than can be tested with the χ^2 statistic. However, if the two distributions do differ significantly, and if direct inspection shows that they actually differ in the correct direction, then the stronger hypothesis can also be accepted. The χ^2 test has actually been applied to every hypothesis; these results are reported in Section 9.6 below, in a table summarizing the results of statistical testing.

9.5 Results

The eight reseach hypotheses have two different types of implication. The first type of implication concerns the basic structure of language in the cockpit; verification of any hypothesis with this type of implication is a partial demonstration of the viability of the methodology developed in this report. All eight hypotheses assert relations between variables of linguistic structure, operational structure, and social structure. Linguistic structure variables include the discourse type, speech act type, and mitigation level of a given utterance. Operational structure variables include presence or absence of a Crew Recognized Emergency, a Crew Recognized Problem, and the operational relevance of a given speech act. The only social structure variable used in the present study is rank in the command hierarchy.

The second type of implication has a more applied direction, such as crew training. In particular, Hypotheses 7 and 8 have this type of implication. There are a number of reasons why it is more difficult to draw such implications. One is that the dataset consists only of accident transcripts, so that detailed information about system performance variables is not

---

[6] This argument is not circular, because the tests supporting the homogeneity of the sample in Section 9.3.4 use the t test, the justification of which has already been discussed in Section 9.4.3.

available, nor is there a control set of non-accident data. It is therefore impossible for this study to verify directly hypotheses about training, or about the relationship of linguistic variables to system performance variables. Moreover, it is difficult to identify and control for auxiliary variables that may interfere with the relationships of primary interest. A discussion of the overall significance of both types of results and of directions for future research is given in Section 11, and a summary of results is given in Section 9.6.

This section discusses the tests of the eight research hypotheses, each in a separate subsection. For each hypothesis, we indicate first the results from examining data from the six test transcripts, then the results of examining data from the two hypothesis formulation transcripts, and finally, provided the two groups can be combined, the results from all eight transcripts. In this discussion, the term "obtained level" is used for the probability level obtained for the experimental data assuming that the null hypothesis is true.

9.5.1 Requests to Superiors Are More Mitigated

This hypothesis represents the intuition that the speech of subordinates is more tentative and indirect than the speech of superiors. The hypothesis is important because it posits a direct effect of the basic social hierarchy on cockpit discourse. If this hypothesis is verified, and if it is also shown that more highly mitigated speech acts are more often misunderstood or ignored (as is strongly suggested by the acceptance of Hypotheses 7 and 8 below), then it should be worth testing whether training subordinates to use less mitigation would improve crew performance. Such a training hypothesis can not itself be tested with data from accident transcripts, but could be tested with simulator experiment data.

	mitigation level					
direction	A	D	LM	HM	total	mean
up	2	40	19	0	61	.279
down	9	57	9	2	77	.052
sums	11	97	28	2	138	

Figure 28: Test Group Mitigation/Aggravation Frequencies for Hypothesis 1

Frequency data for this hypothesis from the six test group transcripts are given in Figure 28. Because the hypothesis asserts that one mean is greater than another, it is tested with a one tailed Student's t test. The frequencies in Figure 28 yield t=2.38 (df=136 and p=.009), using the normal approximation, which is valid because of the large sample size. The hypothesis is therefore accepted, and we conclude that crew members indeed use more mitigation in making requests to superiors in the test transcript sample.

Testing the hypothesis with speech acts from the two hypothesis formulation transcripts yields a similar pattern of frequencies, but with an obtained probability of only .32. The hypothesis is therefore not supported by these data, perhaps because there are too few speech acts to achieve the desired significance level. However, because the hypothesis has been accepted on data from the test transcripts, the speech acts from the two groups can be combined. The pooled frequencies are shown in Figure 29. They yields t=2.01 (df=252, p=.022), so the hypothesis is accepted for the entire dataset. (See also the discussion of generalizability of results in Section 9.3.)

	mitigation level					
direction	A	D	LM	HM	total	mean
up	3	78	25	3	109	.257
down	13	108	19	5	145	.110
sums	16	186	44	8	254	

Figure 29: Total Mitigation/Aggravation Frequencies for Hypothesis 1

Note that only request speech acts were used in testing this hypothesis, and that requests occurring in checklists were excluded. The test was limited to requests because requests (which include orders, questions, draft orders and suggestions) are the speech acts of greatest practical importance for command and control discourse. This is because the request is the most characteristic speech act in command and control discourse, and also because the consequences of misunderstandings of requests are more direct and immediate than those of any other speech act. Requests within checklists were excluded because the highly stereotyped nature of checklists insures that virtually all requests will be direct and will exhibit little variability.

Since appropriateness of the parametric t test depends on homogeneity of variance, it is interesting to notice that in this dataset, the two distributions involved do indeed have approximately equal standard deviations. For speech acts from the six transcripts in the test group, the standard deviation of speech acts by subordinates is .516, while that of speech acts by superiors is .579. (Equality of variance could be tested with the F test, but we have not done so.)

9.5.2 Requests Are Less Mitigated in Crew Recognized Emergencies

This hypothesis reflects the intuition that when crew members know that they face an emergency situation, their speech is less tentative and indirect. It is based on the notion that in any utterance, the speaker is encoding both his understanding of the situation he is talking about (the propositional content) and his understanding of the relation between himself and his addressee. Mitigation level is a major linguistic means by which a speaker can indicate his understanding of this social relation. When the situation becomes urgent, we might expect the speaker to focus most of his attention on it, and thus less attention upon social relations.

Verification of this hypothesis would mean that indeed, crew members are able to vary their level of mitigation depending on their perception of the circumstances. This would mean that training crew members to use less mitigation in specified circumstances would not seem new or strange to them, because mitigation level is already something that they alter when aware that they are in an emergency situation. Under the assumption that what experienced crews do in emergency situations may be valuable, verification of this hypothesis would also lend some support to the hypothesis that training crews to speak more directly would improve their performance and thus reduce accidents (however, caution is advisable in drawing such a conclusion from the present dataset of accident transcripts).

	mitigation level					
condition	A	D	LM	HM	total	mean
CRE	4	15	0	0	19	-.211
non-CRE	8	109	30	2	149	.175
sums	12	124	30	2	168	

Figure 30: Test Group Mitigation/Aggravation Frequencies for Hypothesis 2

The frequencies obtained from the test transcripts for investigating this hypothesis are summarized in the Figure 30. These data yield t=3.05 (df=166, p=.001). and the hypothesis is therefore accepted. The obtained probability level for similar comparisons of speech acts in the hypothesis formulation group of transcripts is .026. It is therefore permissable to combine the two datasets, yielding the frequencies shown in Figure 31. Comparing mitigation levels during CRE and non-CRE for speech acts from all eight transcripts yielded t=3.46 (df=276, p=.0003). Hypothesis 2 is therefore very strongly supported.

	mitigation level					
condition	A	D	LM	HM	total	mean
CRE	6	32	1	0	39	-.128
non-CRE	11	178	43	7	239	.193
sums	17	210	44	7	278	

Figure 31: Total Mitigation/Aggravation Frequencies for Hypothesis 2

9.5.3 Requests are Less Mitigated in Crew Recognized Problems

This hypothesis corresponds to the intuition that crew members' speech is less tentative and indirect when they know they face a problem. Its significance is similar to that of the previous hypothesis. (Note that every CRE speech act is also a CRP speech act.)

condition	mitigation level				total	mean
	A	D	LM	HM		
CRP	9	48	10	0	67	.015
non-CRP	3	76	20	2	101	.218
sums	12	124	30	2	168	

Figure 32: Test Group Mitigation/Aggravation Frequencies for Hypothesis 3

The frequencies obtained from speech acts in the test group of transcripts are summarized in Figure 32, comparing CRP and non-CRP mitigations levels. These data give $t=2.34$ ($df=166$, $p=.010$). The hypothesis is therefore accepted for the test dataset. For the hypothesis formulation transcripts, the corresponding obtained probability level is .149. Combining the two groups produces the frequencies shown in Figure 33, for $t=1.79$ ($df=276$, $p=.047$). The hypothesis is therefore accepted for the dataset as a whole.

condition	mitigation level				total	mean
	A	D	LM	HM		
CRP	14	128	23	4	169	.101
non-CRP	3	82	21	3	109	.220
sums	17	210	44	7	278	

Figure 33: Total Mitigation/Aggravation Frequencies for Hypothesis 3

9.5.4 Subordinates Plan and Explain More Often

This research hypothesis probes, in an indirect way, the effects of social hierarchy on subordinates' contributions to explaining what is happening and to planning what should happen in the future. Rejection of this hypothesis would suggest that the social hierarchy might be having a detrimental effect on crew communications. As usual, the null hypothesis is the hypothesis of "no difference," in this case, that subordinates and superiors engage in equal amounts of planning and reasoning.

Discourse type frequencies for speech acts in the six test transcripts are summarized in Figure

	rank		
condition	sub	sup	total
pln/expl	25	38	63
n-pln/expl	204	213	417
sums	229	251	480

Figure 34: Test Group Rank Frequencies for Hypothesis 4

34. Statistical examination yielded $\chi^2=1.52$ and an obtained probability level somewhere between .10 and .20. Therefore the hypothesis is rejected with these data. A similar study of speech acts from the formulation transcripts gives $\chi^2=1.13$, for an unacceptable probability level between .20 and .30. It is therefore permissible to combine the two datasets. The pooled frequencies given in Figure 35 produce $\chi^2=2.97$, associated with a probability level a little more than .05. Observe that subordinates produce only 38% of the planning and explaination speech acts in this dataset, while superiors produce 62%; also observe that subordinates and superiors each produce about half of all speech acts in this dataset, but planning and explanation speech acts are only 9% of these speech acts. The obtained probability level means that observed frequencies as far from equal as these are would occur more than 5 percent of the time, if the null hypothesis of equal percentages were true. The null hypothesis therefore cannot be rejected on the pooled data, although it is close.

	rank		
condition	sub	sup	total
pln/expl	31	50	81
n-pln/expl	391	407	798
sums	422	457	879

Figure 35: Total Rank Frequencies for Hypothesis 4

Having rejected the research hypothesis, notice that the numbers in Figure 35 show that not only do subordinates *not* produce more plans and explanations than subordinates, but the *opposite* of the research hypothesis, namely that superiors produce more plans and explanations, is very nearly accepted. This outcome is interesting because modern management theory generally asserts that a group is more effective when subordinates contribute more than superiors. Moreover, informal examinations of accident transcripts have suggested to many observers that captains often behave in an autocratic manner that prevents subordinates from making appropriate contributions. Our results strongly suggest that it would be valuable to determine whether crew performance is improved by training subordinates to engage in more planning and explanation, and training captains to encourage this, at least in the condition of CRP but not CRE. It would also be important to determine if there are circumstances, such as CRE, in which it would be counterproductive to engage in more planning and explanation. Once again, it would be very interesting to compare the present results with results from data from normal flights.

A more careful analysis than is possible with the coding scheme used in this study could separate explanations produced in connection with plans from those produced in connection with draft orders, and it would be interesting to see if either subcategory of explanations is more frequently produced by subordinates. It would also be interesting to explore differences between planning and explanation in CRE and CRP (see the discussion of Hypotheses 4 and 5), and also to explore whether or not flight segment has any effect.

9.5.5 Planning and Explanation Are Less Common in Crew Recognized Emergencies

This hypothesis represents the intuition that when crew members are aware that they face an emergency, they do less planning and explaining, because an emergency calls for immediate action. Precise knowledge of the distribution of planning and explanation in accident transcripts is important because it may suggest circumstances in which crews should be trained to do more planning and explanation, or else less, when it proves to be counterproductive.

The speech act frequencies for this hypothesis in the test transcripts are summarized in Figure 36. The χ^2 statistic is used to test whether or not the proportion of planning and explanation speech acts occurring in CRE differs significantly from that in non-CRE. The data in Figure 36 yield χ^2=3.87 for an obtained probability level less than .05. The hypothesis is therefore (just barely) accepted at the .05 significance level.

	disc. type		
condition	Pl/E	n-Pl/E	total
CRE	1	67	68
non-CRE	51	495	546
sums	52	562	614

Figure 36: Test Group Discourse Type Frequencies for Hypothesis 5

The corresponding test for speech acts from the hypothesis formulation transcripts yields χ^2=7.03 ($p<.01$). Thus, it is permissible to combine the two datasets for Hypothesis 5. The combined frequencies appear in Figure 37 and yield χ^2=12.49 ($p<.001$); the hypothesis is therefore strongly supported on the pooled data. Further discussion of the implications of this result is included with that of the following hypothesis.

It should also be noted that because this study is based upon accident transcripts, it cannot be assumed that observed crew behavior in this data is necessarily optimal. It seems quite possible that the data used in this study are a combination of good and bad instances of cockpit planning and reasoning, and that testing the present hypothesis on data from normal flights would yield more definitive results.

	disc. type		
condition	Pl/E	n-Pl/E	total
CRE	1	102	103
non-CRE	127	809	936
sums	128	911	1039

Figure 37: Total Discourse Type Frequencies for Hypothesis 5

9.5.6 Planning and Explanation Are More Common in Crew Recognized Problems

This hypothesis corresponds to the intuition that crew members use more planning and explanation when they are aware that they face a problem. If verified, this hypothesis would strengthen our confidence in the relevance of the variables involved (discourse type and CRP), and would also confirm the value of training crews to plan and reason in problem situations.

	disc. type		
condition	Pl/E	n-Pl/E	total
CRP	45	23	68
non-CRP	184	362	546
sums	229	385	614

Figure 38: Test Group Discourse Type Frequencies for Hypothesis 6

The discourse type frequencies obtained from speech acts in the test transcripts are summarized in Figure 38. Testing the hypothesis yielded a χ^2=25.90, with an obtained probability level well beyond .001. The hypothesis is therefore very strongly confirmed in this dataset. The corresponding χ^2 value for discourse type frequencies from the hypothesis formulation transcripts is .27, for an obtained probability level of approximately .7. Frequencies by discourse type for speech acts from the combined group of eight transcripts are shown in Figure 39. These data yield χ^2=12.03, and an associated probability level again well below .001. The hypothesis is therefore strongly confirmed for the entire dataset.

	disc. type		
condition	P1/E	n-P1/E	total
CRP	79	24	103
non-CRP	548	388	936
sums	627	412	1039

Figure 39: Total Discourse Type Frequencies for Hypothesis 6

These results taken together with the findings relevant to Hypothesis 5 suggest that, perhaps contrary to expectation, more planning and reasoning occur when the crew believes that it is dealing with a problem, but not when it believes that it is dealing with an emergency. One explanation for this result is that by the time an emergency situation has developed, crew members may feel that it is too late to take the time to plan as a group, or to explain the reasons for taking specific actions. It is of course possible that more planning and explanation would be desirable in some emergency situations, but not in others. This suggests using simulator experiments to determine in which flight segments (if any) more planning and explanation produce better performance. In any case, these results make it clear that crews should plan as effectively as possible during CRP, because they not have time for planning during a subsequent emergency.

9.5.7 Topic Failed Speech Acts Are More Mitigated

This hypothesis and the next one attempt to probe the idea that excessive mitigation can have undesirable effects in the cockpit. Since the effect of mitigation on performance data (such as the probability of an accident) cannot be explored directly with the present data, we are forced to examine less direct connections.

This hypothesis represents the intuition that a new topic is less likely to be continued by its addressees if the speech act in which it is introduced is excessively mitigated. We count as topic failed any speech acts expressing a new topic not followed by a speech act having the same topic from another speaker. The frequencies relevant to this hypothesis using speech acts obtained from the six test transcripts are summarized in Figure 40.

	mitigation level					
condition	A	D	LM	HM	total	mean
topic fail	2	54	11	4	71	.239
topic succ	11	81	20	1	113	.097
sums	13	135	31	5	184	

Figure 40: Test Group Mitigation/Aggravation Frequencies for Hypothesis 7

A comparison of mitigation scores for the two topic conditions gives t=1.65 (df=182, p=.01), and thus this hypothesis is accepted. For comparisons based on the hypothesis formulation transcripts, t=2.23 (df=80, p=.013). Examining the combined dataset mitigation levels across topic conditions in all eight transcripts yields the frequencies shown in Figure 41. These data give t=2.493 (df=264, p=.0064). Therefore the hypothesis is accepted.

	mitigation level					
condition	A	D	LM	HM	total	mean
topic fail	2	69	21	6	98	.316
topic succ	14	121	30	3	168	.131
sums	16	190	51	9	266	

Figure 41: Total Mitigation/Aggravation Frequencies for Hypothesis 7

This result lends strong support to the intuition that excessive mitigation can have undesirable effects on crew performance. A number of NTSB reports have recommended assertiveness training for crew members to encourage effective participation by subordinates. Verification of the present hypothesis and the following one, demonstrate effects for one kind of lack of assertiveness. Moreover, this kind of lack of assertiveness is defined precisely enough to allow for both training and for the evaluation of training methods.

9.5.8 Unratified Draft Orders Are More Mitigated

This hypothesis attempts to test the intuition that when a crew member proposes a suggestion to the captain, the more indirect and tentative that suggestion is, the less likely the captain is to ratify it. The frequencies for ratified and unratified draft orders from the six test transcripts are given in Figure 42.

Statistical evaluation of the data in Figure 42 yields a t=2.927 (df=45, p=.002). The

	mitigation level					
condition	A	D	LM	HM	total	mean
not ratif	1	10	14	1	26	.577
ratified	1	17	3	0	21	.095
sums	2	27	17	1	47	

Figure 42: Test Group Mitigation/Aggravation Frequencies for Hypothesis 8

hypothesis is therefore accepted for speech acts from the test transcripts. For similarly classified speech acts from the hypothesis formulation transcripts, t=.589 (df=13). For less than 30 degrees of freedom, the normal approximation is not very accurate; we use instead a small sample t statistic table, which gives an obtained probability level of approximately .2. It is therefore permissible to combine the two groups, and frequencies for this dataset are given in Figure 43. The pooled data yields t=2.412 (df=60, p=.008). Thus, this hypothesis is strongly supported.

	mitigation level					
condition	A	D	LM	HM	total	mean
not ratif	2	11	17	4	34	.676
ratified	1	20	6	1	28	.250
sums	3	31	23	5	62	

Figure 43: Total Mitigation/Aggravation Frequencies for Hypothesis 8

Like Hypothesis 7, this hypothesis implies that excessive-mitigation can have undesirable effects on crew performance. In particular, this hypothesis focusses attention on the situation in which a subordinate makes a correct suggestion which is ignored. Training in linguistic directness should be valuable in correcting this kind of pattern.

9.6 Summary of Results

This subsection gives two figures showing first, the independent and dependent variables that are used in each hypothesis, and second, the results of testing each hypothesis.

Figure 44 shows the independent and dependent variables occurring, and which hypothesis uses each. (The two blanks suggest possibly interesting hypotheses that have not been tested in this study.)

	independent-variables				
dep vble	rank	CRE	CRP	topic failed	ratif
mitigatn	1	2	3	7	8
plan/expln	4	5	6		

Figure 44: Variables Used in Hypotheses

Figure 45 shows for each hypothesis: the size, N, of the dataset used to test it (in each case this includes speech acts from all 8 transcripts); the obtained t value (if any); the obtained χ^2 value; the number of degrees of freedom (for the χ^2 test); the obtained probability level for the t test; the obtained probability level for the χ^2 test; and the decision (whether or not the research hypothesis was accepted). The χ^2 values have not been given previously. The decisions obtained using the χ^2 test agree with those obtained using the t test, except in the case of Hypothesis 1. Although the χ^2 value is very close to that required for acceptance, a reader who remains doubtful about the applicability of the t test, may want to consider this hypothesis rejected.

Hypothesis	N	t	χ^2	df	P_t	P_χ	Decision
1	254	2.01	7.45	3	.022	.05+	Yes
2	278	3.46	12.81	3	.0003	<.01	Yes
3	278	1.79	4.70	3	.047	<.01	Yes
4	879		2.97	1		>.05	No
5	1039		12.49	1		<.001	Yes
6	1039		12.03	1		<.001	Yes
7	266	2.49	7.95	3	.0064	<.05	Yes
8	62	2.41	9.52	3	.008	.02+	Yes

Figure 45: Summary of Results

These results demonstrate that the linguistic study of CVR transcripts has produced results of interest for aviation safety. In particular, the results suggest the desirability of further research on training aircrews in linguistic behavior, and on linguistic measures of crew performance.

10 FURTHER RESEARCH

This section discusses both immediate directions for further research and also possible practical applications of the entire research program. The focus of the present study has been on basic research, the theoretical and methodological foundations necessary to apply linguistic methodology to the language of the cockpit. A number of hypotheses arising from this foundation have been formulated, tested, and verified, demonstrating, we believe, the correctness and potential value of the theory.

However, because the nature of data from CVR transcripts imposes serious restrictions on possible hypotheses, only a relatively few hypotheses have yet been tested. One problem is that each transcript represents a unique event; hence it is impossible to form hypotheses correlating linguistic patterns with specific types of events in the real world. Another problem is that in the absence of a video record, it is often difficult to tell what actions crew members took; hence, it is difficult to correlate linguistic patterns with their social effects. Both of these problems can be remedied by the use of data from flight simulators. And it is a major priority of this research program to apply the methodology developed in this report to data from flight simulators.

The success of the current research strongly indicates the value of linguitistic measures in future research and training. One value of such measures is their relative simplicity and low cost. Because we have shown that individual differences have a relatively small effect on some such measures, it is possible to compare such measures across crews, rather than being confined to successive research runs on the same crews. This simplifies the task of gathering simulator data, and also permits the study of actual flights performed by different crews. (At this point, the study of actual flights should focus on successfully completed flights, since this is the necessary comparison to the present study of flights ending in accidents.) Another value of such measures, both in simulator experiments and eventually in training is their sensitivity. We believe, and hope to test in later research, that these measures are more sensitive than behavioral measures, and will be able to indicate an earlier degradation of crew performance.

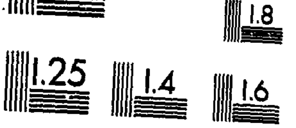
In the following subsections, we discuss some linguistic measures of crew performance which are suggested by the present research, and also some more speculative possibilities for improving air crew communication.

10.1 Degree of Command and Control Coherence

This subsection uses the methodology of the present report to define a linguistic variable that may be important in future studies, although it is not used in any of the hypotheses of this study. This variable grows directly out of the rules for speech act chains (in Section 6.2) and gives a social interpretation to the formal constraints on sequencing of those rules. Its value would lie in its correlation with performance or behavioral variables.

2 OF 2

5135 UN



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

10.1.1 The Notion of Degree of Command and Control Coherence

This definition attempts to capture the intuition that one can judge the degree to which a given sequence of utterances is well-integrated and tightly structured. Such a well-integrated sequence follows a request or report with an acknowledgement, support, challenge, or request. No requests or reports are left without acknowledgement or comment. Such a pattern allows a crew member to know that his utterance has been heard and attended to. In contrast, sequences in which reports and requests are followed by silence, by new topics, or by irrelevant material, do not allow a crew member to know whether his utterance has been accepted, rejected, or not received.

The discourse units present in segments with a high degree of command and control coherence are: **speech act chains**, which involve the transmission, acknowledgement, discussion and verbal fulfillment of orders; **plans**, which involve the discussion of possible future actions; and **explanation**, which involves diagnosing and agreeing upon an understanding of the current or expected state of affairs. The discourse units which we have found only in non-command and control coherent CVR discurse are **narratives**, including **pseudonarratives**, which in the cockpit tend not to be operationally relevant.

Figure 46 displays the major characteristics of high and low command and control coherent discourse.

High Command and Control Coherence	Low Command and Control Coherence
Continued propositional content; i.e. successive utterances refer to previous utterances.	Successive utterances are not connected to previous utterances
Acknowledgement is explicit	Acknowledgement is not used, or is inexplicit, i.e. an order is acknowledged by a nod, or by beginning to carry it out
Discourse units include speech act chains, plans and reasoning	Discourse units include narratives and pseudonarratives
Topic coherence is operationally relevant	Topic coherence is not operationally relevant

Figure 46: Characteristics of Command and Control Coherent Discourse

These factors mean that discourse with a high degree of command and control coherence makes crew interaction operationally relevant and explicit, characteristics which help to insure optimal crew coordination and resource management.

10.1.2 Topic Coherence

As discussed in Section 8.2, topical coherence may or may not be operationally relevant. But operationally relevant topic coherence is a factor in computing the degree of command and control coherence. Consider (57), which shows topic coherence both with and without operational relevance.

```
(57a) CAM-2  What's all this, lights in the fields?
```
Operationally relevant to the question of visibility

```
(57b) CAM-2  What the # are they, chicken farms?
```
Possibly operationally relevant to the question of location

```
(57c) CAM-1  Yeah
```
Operationally relevant as an acknowledgement

```
(57d) CAM-2  God Almighty
```
Neutral to the question of operational relevance

```
(57e) CAM-2  They're planning on growing a few eggs, aint they
```
Not operationally relevant

```
                                        (Texas/Mena/73; 8:40:0)
```

Thus, in computing the degree of command and control coherence for this segment, the last two utterances would not be counted, since they they are not operationally relevant.

10.1.3 Computation of Command and control Coherence

For a segment of text of a given length, the degree of command and control coherence is computed using the following formula:

```
                                    Command and Control Utterances
Command and Control Coherence  =  --------------------------------
                                     Total Number of Utterances
```

This is the simplest possible formula for this computation. Later work on this variable may show that a more complex computation is necessary.

A command and control utterance is one which forms part of a valid speech act chain, as given by the command and control grammar; this may include segments of planing or reasoning. A non-command and control utterance is one which is part of any other discourse unit, or which is isolated and does not form a part of any larger unit. There are several points to be made about this definition.

1. We exclude single utterances from command and control coherence. This means that an order which is immediately complied with still does not count as command and control coherent. The reason for this is that such non-verbalized compliance places a demand on the speaker to look at the the addressee to see if his order has been received and acted upon. Such a demand on visual attention is probably non-optimal resource management, because considerable visual attention may be already demanded by the task at hand.

2. The definition, and the grammar, exclude sequences of the form Report Report, since the operational relevance of the second report is either not present, or not made explicit. An example would be

```
(58a)  CAM-2  We dont want to get too far up the #### it
              gets hilly.
(58b)  CAM-1  Yeah stars are shining
                                      (Texas/Mena/73, 17:02)
```

3. The formula is purely formal; it does not exclude sequences that have the form of a valid speech act chain but which are not operationally relevant. (59) is an example of this sort constructed by the analysts.

```
(59a) 6  Captain?
(59b) 1  Yes Carol?
(59c) 6  Did you want me to check the name of that
          restaurant for you?
(59d) 1  Yes please
(59e) 6  OK I'll get it
```

We consider that this chain is indeed operationally relevant but relevant to a goal other than that of flying the airplane. Further, we conjecture that maintenance of the form of command and control discourse for a non-operationally relevant matter can still strengthen the habit of using that form in operationally relevant situations, and hence has a beneficial effect.

4. This variable can be computed for text segments of any length. The segment could be an entire transcript, a specified time period, or a segment defined by any linguistic or behavioral variable, such as CRE, physiological indicators, etc.

10.1.4 Relation to Previous Work and Potential Use

This variable can be seen as an extension of the finding of [Foushee & Manos 81] that use of a greater number of the proper form of commands and acknowledgements is correlated with mission success. By defining the linguistic form of proper command and control sequences, we are able to make this finding more sensitive, and hence we hope more useful. We expect that command and control coherence will function as a linguistic correlate of resource management, attention, and vigilance. Thus, it should be valuable in studying these factors, particularly since it may deteriorate earlier than behavioral or physiological indicators.

10.2 Linguistic Measures and Flight Phase

Another valuable direction for research would be to investigate the relation of the linguistic variables of the present study to flight phase -- taxi, takeoff, climb, cruise, approach, and land. It is possible that such factors as rate of planning and explanation in Crew Recognized Emergency vary according to the flight phase in which the CRE falls, since the flight phase would determine, to some extent, the amount of time available for planning and explanation. Other variables might be similarly sensitive to flight phase. Research into this relation would be valuable in refining the current hypotheses, and thus making them more precise in their application to training.

10.3 Other Linguistic Variables

The variable discussed in the previous subsection may be viewed as a model for how linguistic variables of interest may be formulated and correlated with problems of crew coordination and resource management. Other variables of this kind which are suggested by the present project include: rate of planning and reasoning in Crew Recognized Problem and Crew Recognized Emergency situations, number of Requests with high prior spectra of interpretation, use of explanation in constructing false hypotheses about the nature of a problem situation, rate of request-report-acknowledgement triples (an easily computable subset of command and control coherent discourse), relation of profanity to topic success, etc. These variables should be easily testable on flight simulator data, in which there is sufficient repetition of the situations of interest. We also expect that further variables will be suggested by this data.

10.4 Approaches to Training

As we have already noted, further work must be done to move from the current theoretical and methodological framework to a body of validated test results, which can serve as a solid foundation for training recommendations and other forms of application. However, even at this preliminary stage, we would like to suggest some implications for application which have been suggested by this reseach.

One method for training would be to use films or video tapes illustrating the effects of certain patterns of communication on crew coordination and decision making. Examples could be shown of excessively mitigated or ambiguous requests and suggestions, of excessive attention to one aspect of a problem, to the neglect of the entire situation, of ignoring subordinates reports or challenges, and of the entire crew's construction of a false hypothesis. This approach could be combined with an approach which involves the insertion of peer commentary into tapes of actual flight simulations [Frankel & Beckman 82].

Becoming somewhat more speculative, it might be possible to design new speech acts having formal command and control status, in order to address particular communication problems. For example, a formal **challenge** speech act, perhaps termed **a note**, might be created, which would be addressed by a subordinate to the captain, and which the captain would be legally obligated to acknowledge as such. (Of course the captain need not *ratify* the content of the note, but need only *acknowledge* that he had received it.) The use of such a formal speech act would prevent the captain's misunderstanding the crew member's intention to challenge. We expect that such a device would be difficult for crew members to use in an explicit way, but that it could be used more easily as part of an "off record" strategy. Just the possibility of such a device being used could have beneficial effects, even if it were very rarely used.

Another speculative application for the approaches discussed in this report is the developement of linguistic countermeasures for fatigue. It might be, for example, that some linguistic patterns were more conducive to vigilance and alertness than others. Or it might be that certain patterns were diagnostics of low alertness, and could be used by the crew as such.

Moving futher into the future, cockpit automation may well proceed to the point where it is desirable to have complex verbal output from the system to the crew, including reports, acknowledgements, plans, and explanations. The latter would be particularly important for promoting effective crew utilization of on-board diagnostic systems, as experience with similar systems for medical diagnosis has shown [Swartout 81]. In order to integrate such verbal readouts of system functions with crew routines, it would be helpful if the same discourse forms were used by both the crew and the system, particularly in the case of the very complex structures used in planning and explanation. This would also be true for visual CRT readouts. Work on medical expert systems has already shown that it is extremely important to match the form of the system's output to a form easily assimilated and assessed by humans. It will be even more important in situations where the information must be used in a real time operational setting, particularly in an emergency situation.

11 CONCLUSIONS

Based on the work reported above, it may be concluded that we now have available a methodology for the detailed analysis of cockpit discourse that can be applied to improving aviation safety. For example, the methodology can be used to formulate and evaluate hypotheses about the behavior of air-crews during such language-intensive activities as planning and decision making. This methodology has been used to formulate a number of linguistic variables that might serve as measures for various aspects of air crew performance, such as vigilance and crew coordination. The methodology has also been used to formulate a number of training suggestions for air crew language use that can be tested to see if they improve performance.

In support of this methodology, the statistical hypotheses tested in Section 9, while far from comprehensive, provide convincing evidence that the variables we have isolated are reliable and valid, and have powerful relationships with one another and with the general structure of cockpit activity; moroever, there is suggestive evidence that they may have powerful relationships with crew and system performance levels. In particular, the important role of **mitigation** in cockpit communication has been clearly demonstrated by showing its correlation with a number of basic structural and decision making properties such as rank, topic failure, and draft order ratification.

It should be noted that there are two levels of interpetation for this research. The first is the descriptive level, demonstrating relations within the dataset. There is no question that the results of this study can be given this interpretation. The second level of interpretation is inferential, generalizing from this dataset to all aviation accidents. Because statistically rigorous reserach on natural data at the discourse level is quite new, there may be some questions about the validity of this interpretation. This issue is discussed in some detail in Section 9.

Perhaps more important, in the long run, than the validation of any specific training hypothesis, is the basic understanding of the structure of crew coordination and resource management that is emerging from the discourse level analysis of cockpit language. This discourse level structure should correlate both with crew mangagement level objectives and with system level variables. It should therefore serve as a basis for automating aspects of aviation that involve communication, as well as for evolving and evaluating other research directions.

The following two subsections detail what we believe to have been the major contributions of the work described in this report.

11.1 General and Basic Contributions

1. A classification of the discourse types that occur in aviation discourse. These are: command and control chain, including the subtype of checklist; planning; explanation; and narrative and pseudo-narrative.

2. A theory of the structure of command and control chains that includes a determination of its relationships to planning and explanation, as well as its basic speech acts which are request, report, acknowledgement and declaration.

3. A general theory of the structure of discourse; this theory involves analyzing a given discourse unit as a sequence of transformations that construct an underlying tree structure representing the structure of the discourse, i.e., a hierarchical classification of the discourse parts and their relationships.

4. A scale of **mitigation** levels for speech acts occurring in aviation discourse. This scale ranges from "highly mitigated" to "aggravated" and has "direct" as its zero point. An experimental validation of this scale was conducted with six subjects who were commercial flight personnel judging selected utterances from accident transcripts.

5. A theory of speech act misinterpretations, having as its central notions the **prior** and **posterior spectra** of a speech act.

6. A theory of draft orders (suggestions for action that have not yet been ratified by the captain) and how they come to be ratified has been developed, based on the theories of planning, explanation, and command and control discourse.

7. A collection of variables has been isolated that summarize many important characteristics of the speech acts that occur in cockpit discourse.

8. A basic method and set of computational tools has been developed for testing statistical hypotheses concerned with speech acts and discourse structure. The tools include LISP programs for checking the consistency of coded data sets, for extracting relevant data from them, and for performing the necessary statistical calculations.

11.2 Applied and Specific Contributions

This subsection describes what we believe are the most important specific contributions of this research to aviation safety. It should be remembered that these contributions are necessarily rather limited at this time, because of the restriction of our data to accident transcripts. It should be possible to go much further in the directions indicated here when the data set includes both systems data and non-accident data. Consequently, many of these contributions are in fact suggestions for further research based on the results of the present work.

1. It has been shown that the average mitigation level of requests by subordinates is significantly higher than that of requests by superiors. It has not been shown that this asymmetry contributes to the misinterpretation of suggestions and commands in the cockpit, but would be important to test this hypothesis, simply because it would probably not be difficult to train subordinate crew members to use less mitigated language, or (as the NTSB puts it) to be more assertive.

2. It has been shown that there are significant regional differences in the interpretation of mitigation. This may be another factor contributing to the misinterpretation of speech acts in the cockpit; further research would be valuable since it would not be difficult to train crew members to a better understanding of these regional differences.

3. It has been shown that requests are less mitigated during a Crew Recognized Problem, and are still less mitigated during a Crew Recognized Emergency. This suggests that crew members should not find it strange or abnormal to be trained to use less mitigation, since variation of mitigation lvelis something that they already do under certain conditions. It also suggests that assertiveness training would actually be reinforcing a tendency that already appears under problem and emergency conditions.

4. It has been shown that superiors produce a higher proportion of explanation or planning speech acts than subordinates. The optimal ratio is not clear; it would be important to investigate this. It seems likely that this ratio would be a good indicator of degree of authority delegated by a given captain to his crew.

5. It has been shown that planning and explanation are much more common during crew recognized problems, and that they are *less* common during crew recognized emergencies. This suggests further research to discover whether training crew members to engage in more planning and reasoning under real emergency conditions would improve performance.

6. It has been shown that more mitigated speech acts introducing a new topic, are less likely to have their topic become the subject of further conversation. This demonstrates the importance of crew members not using mitigated language when introducing operationally significant topics. Because this also is presumably behavior for which crew members can be trained, it would be interesting to explore both the basic linguistic phenomena further, and to test whether or not such training can improve any objective performance measures.

7. It has been shown, with a very high level of significance, that on the average, draft orders that do not get ratified are more mitigated than those that do get ratified. The implications of this result are very similar to those of the previous result, but concern the ratification of subordinates' suggestions rather than the success of their topics.

8. The research reported here suggests that a number of other linguistic variables should be investigated for correlation with objective system and crew performance variables. These variables include: degree of command and control coherence, as defined in Section 10.2; the rate of request-report-acknowledge triples; the rate of planning and reasoning; and the rate of simple acknowledgements. A number of other such variables have been suggested at various places in the text. In certain cases, it might be less costly to use a reliable linguistic variable as an indicator of some objective performance measure than to measure it directly. In other cases, important training implications might be discovered.

9. Finally, the research program initiated in this report should have many applications to the design of aviation procedures and equipment that involve communication. This possibility of application arises from the clear demonstration that air crew discourse involves definite linguistic structures, and that these structures correspond in specific ways to the operational structure of the flight. This means that there are only certain times when is natural for certain kinds of communications to occur, and that there are natural forms for each kind of communication. For example, a piece of equipment in the cockpit that produced complex verbal information about the status of the flight plan would probably not be useful unless it produced this information at the right time and in the right form. This implies that its designers should understand the structure of plans and explanations in aviation discourse, and build this structure into the equipment.

We believe it would be worthwhile to investigate a number of different discourse settings using the methodology described in this report. For example, it should be possible to study the language used in space flights, in helicopter flights, in submarines, and in controlling nuclear reactors; this could lead to improved training methods, linguistic measures of the quality of crew coordination, and design criteria for equipment and procedures that involve language.

Bibliography

[Austin 62] Austin, W.
How to Do Things with Words.
Clarendon Press, Oxford, 1962.

[Bowen & Weisberg 80]
Bowen, B. D. and Weisberg, H. F.
An Introduction to Data Analysis.
Freeman, 1980.

[Brown and Levinson 79]
Brown, Penelope and Levinson,Stephen.
Universals in Language: Politeness Phenomena.
In Esther N. Goody (editor), *Questions and Politeness: Strategies in Social Interaction*, . Cambridge University Press, 1979.

[Chomsky 65] Chomsky, Noam.
Aspects of the Theory of Syntax.
MIT Press, 1965.

[Foushee & Manos 81]
Foushee, H. Clayton and Manos, Karen L.
Information Transfer within the Cockpit: Problems in Intracockpit Communications.
Technical Report, NASA Ames Research Center, 1981.
in NASA Technical Paper 1895, Information Transfer Problems in the Aviation System, edited by C. E. Billings and E. S. Cheaney.

[Frankel & Beckman 82]
Frankel, Richard and Beckman, Howard.
IMPACT: An Interaction-Based Method for Preserving and Analyzing Clinical Transactions.
In Pettigrew, L. (editor), *Explorations in Provider and Patient Interactions*, . Irvington Publishers, 1982.

[Gaito 80] Gaito, John.
Measurement Scales and Statistics: Resurgence of an Old Misconception.
Psychological Bulletin 87(3):564-567, 1980.

[Gazdar 79] Gazdar, Gerald.
Pragmatics: Implicature, Presupposition and Logical Form.
Academic Press, 1979.

[Goffman 67] Goffman Erving.
Interaction Ritual: Essays on Face and Face Behavior.
Anchor Books, 1967.

[Goguen 69] Goguen, Joseph A.
The Logic of Inexact Concepts.
Synthese 19:325-373, 1968-1969.

[Goguen, Linde & Weiner 81]
Goguen, Joseph, Weiner, James and Linde, Charlotte.
Reasoning and Natural Explanation.
Technical Report, Structural Semantics, 1981.
submitted for publication.

[Gordon & Lakoff 71]
Gordon, David and Lakoff, George.
Conversational Postulates.
Papers from the Regional Meeting, Chicago Linguistics Society , 1971.

[Grosz 77] Grosz, Barbara J.
The Representation and Use of Focus in Dialogue Understanding.
Technical Report, SRI International, 1977.
Artificial Intelligence Center Technical Note 151.

[Guy 80] Guy, Gregory.
Variation in the Group and the Individual: The Case of Final Stop Deletion.
In William Labov (editor), *Locating Language in Time and Space*, . Academic Press, 1980.

[Herdan 66] Herdan, Gustav.
The Advanced Theory of Langauge as Choice and Chance.
Springer-Verlag, 1966.

[Keenan & Schieffelin 75]
Keenan, Eleanor and Schiefflin, Bambi.
Topic as a Discourse Notion.
In Li, Charles N. (editor), *Subject and Topic*, . Academic Press, 1975.

[Labov 70] Labov, William.
The Study of Language in its Social Context.
Studium Generale 23:30-87, 1970.

[Labov 72] Labov, William.
The Transformation of Experience into Narrative Syntax.
In *Language in the Inner City*, . University of Pennsylvania Press, Philadelphia, 1972.

[Labov & Fanshel 77]
Labov, William and Fanshel, David.
Therapeutic Discourse; Psychotherapy as Conversation.
Academic Press, 1977.

[Linde 74] Linde, Charlotte.
The Linguistic Encoding of Spatial Information.
PhD thesis, Columbia University, 1974.

[Linde & Goguen 78]
Linde, Charlotte and Goguen, Joseph.
The Structure of Planning Discourse.
Journal of Social and Biological Structures 1:219-251, 1978.

[Linde & Labov 75]
Linde, Charlotte and Labov, William.
Spatial Networks as a Site for the Study of Language and Thought.
Language 51, 1975.

[Matisoff 79] Matisoff, James.
Blessings, Curses, Hopes and Fears: Psycho-ostensive Expressions in Yiddish.
Institute for the Study of Human Issues, 1979.

[Murphy 80] Murphy, Miles.
Analysis of Eighty-four Commercial Aviation Incidents: Implications for a Resource Management Approach to Crew Training.
In *Proceedings, Annual Reliability and Maintainability Symposium*, . IEEE, 1980.

[Newell & Simon 72]
Newell, Alan and Simon, Herbert .A.
Human Problem Solving.
Prentice Hall, 1972.

[Polanyi 79] Polanyi, Livia.
So What's the Point?
Semiotica , 1979.

[Ruffell Smith 79]
Ruffell Smith, H.P.
A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance and Decisions.
Technical Report, NASA Ames Research Center, 1979.

[Schegloff & Sachs 73]
Schegloff, Emanuel and Sachs, Harvey.
Opening up Closings.
Semiotics 8(4), 1973.

[Searle 69] Searle, J. R.
Speech Acts.
Cambridge University Press, 1969.

[Searle 71] Searle, John.
What Is a Speech Act?
In Searle, John (editor), *The Philosophy of Language*, . Oxford University Press, 1971.

[Searle 79] Searle, John.
Expression and Meaning.
Cambridge University Press, 1979.

[Siegel 56] Siegel, S.
Nonparametric Statistics for the Behavioral Sciences.
McGraw-Hill, 1956.

[Structural Semantics 82]
Goguen, Joseph and Linde, Charlotte.
Linguistic Methodology for the Analysis of Aviation Accidents: Second Interim Technical Report.
Technical Report, Structural Semantics, 1982.

[Swartout 81] Swartout, William.
Producing Explanations and Justifications of Expert Consulting Programs.
Technical Report, MIT Laboratory for Computer Science, 1981.

[Weiner 79] Weiner, James L.
The Structure of Natural Explanation: Theory and Application.
PhD thesis, UCLA, Computer Science Department, 1979.
also published as Report SP-4035, Systen Development Corporation, Santa Monica, CA, October.

[Weiner 80] Weiner, J.
BLAH: a System which Explains its Reasoning.
Artificial Intelligence 15:19-48, 1980.

[Zadeh 65] Zadeh, Lotfi.
Fuzzy Sets.
Information and Control 8:338-353, 1965.

[Zadeh 77] Zadeh, Lotfi.
Fuzzy Sets as a Basis for a Theory of Possibility.
Technical Report, Electronics Research Laboratory, University of California at Berkeley, 1977.

I. Summaries of Eleven Transcripts

The following summaries are all official NTSB abstracts, except numbers 7 and 10, which were prepared by Structural Semantics from ALPA reports.

1. United/Portland/79
About 1815 Pacific standard time on December 28, 1978, United Airlines Inc., Flight 173 crashed into a wooded populated area of suburban Portland Oregon, during an approach to Portland International Airport. The aircraft had delayed southeast of the airport at a low altitude for about 1 hour while the flightcrew coped with a landing gear malfunction and prepared the passengers for the possibility of a landing gear failure upon landing. The plane crashed about 6 nmi southeast of the airport. The aircraft was destroyed; there was no fire. Of the 181 passengers and 8 crewmembers aboard, 8 passengers, the flight engineer and a flight attendant were killed and 21 passengers and 2 crewmembers were injured seriously.

The National Transportation Safety Board determined that the probable cause of the accident was the failure of the captain to monitor properly the aircraft's fuel state and to properly respond to the low fuel state and the crewmembers' advisories regarding fuel state. This resulted in fuel exhaustion to all engines. His inattention resulted from preoccupation with a landing gear malfunction and preparations for a possble emergency landing.

Contributing to the accident was the failure of the other two flight crewmembers either to fully comprehend the criticality of the fuel state or to successfully communicate their concern to the captain.

2. Eastern/Miami/72
An Eastern Air Lines Lockheed L-1011 crashed at 2342 eastern standard time, December 29, 1972, 18.7 miles west-northwest of Miami International Airport, Miami, Florida. The aircraft was destroyed. Of the 163 passengers and 13 crewmembes aboard, 94 passengers and 5 crewmembers received fatal injuries. Two survivors died later as a result of their injuries.

Following a missed approach because of a suspected nose gear malfunction, the aircraft climbed to a 2,000 feet mean sea level and proceeded on a westerly heading. The three flight crewmembers and a jumpseat occupant became engrossed in the malfunction.

The National Transportation Safety Board determines that the probable cause of this accident was the failure of the flightcrew to monitor the flight instruments during the final 4 minutes of flight, and to detect an unexpected descent soon enough to prevent impact with the ground. Preoccupation with a malfunction of the nose landing gear position indicating system distracted the crew's attention from the instruments and allowed the descent to go unnoticed.

As a result of the investigation of this accident, the Safety Board has made recommendations to the Administrator of the Federal Aviation Administration.

3. Northwest Orient/Thiells/74
About 1926 e.s.t. on December 1, 1974, Northwest Airlines Flight 6231, a Boeing 727-251, crashed about 3.2 nmi west of Thiells, New York. The accident occurred abou 12 minutes after the flight had departed John F. Kennedy International Airport, Jamaica New York, and while on a ferry flight to Buffalo, New York. Three crewmembers, the only persons aboard the aircraft, died in the crash. The aircraft was destroyed.

The aircraft stalled at 24,800 feet m.s.l. and entered an uncontrolled spiralling descent into the ground. Throughout the stall and descent, the flightcrew did not recognize the actual condition of the aircraft, and did not take the correct measures necessary to return the aircraft to level flight. Near 3,500 feel m.s.l, a large portion of the left horizontal stabilizer separated from the aircraft, which made control of the aircraft impossible.

The National Transportation Safety Board determines that the probable cause of this accident was the loss of control of the aircraft because the flightcrew failed to recognize and correct the aircraft's high-angle-of-attack, low-speed stall and its descending spiral. The stall was precipitated by the flightcrew's improper reaction to erroneous airspeed and Mach indications which had resulted from a blockage of the pitot heads by atmospheric icing. Contrary to standard operational procedures, the flightcrew had not activated the pitot head heaters.

4. Allegheny/Rochester/78
About 1750 e.d.t., July 9, 1978, Allegheny Airlines Inc., Flight 453, a British Aerospace Corporation BAC 1-11, overran the departure end of runway 28 at the Monroe County Airport, Rochester, New York, after completing a precision approach and landing in visual flight conditions. After the aircraft overran the end of the runway, it crossed a drainage dich and came to rest 728 ft past the end of the runway threshold. Although the aircraft was damaged substantially when it hit the drainage ditch, thehere was no fire. There were 73 passengers and a crew of 4 on board; one passenger was injured seriously.

The landing aircraft passed over the runway threshold at 184 KIAS -- kns above the reference speed -- and landed nose wheel first at a point about 2,540 ft down the 5,500-ft runway at a speed of about 163 KIAS -- 40 to 45 kns above the normal touchdown speed. A go-around was not attempted.

The National Transportation Safety Board determines that the probable cause of the accident was the captain's complete lack of awareness of airspeed, vertical speed, and aircraft performance throughout an ILS approach and landing in visual meteorological conditions which resulted in his landing the aircraft at an excessively high speed and with insufficient runway remaining for stopping the aircraft, but with sufficient aircraft performance capability to reject the landing well after touchdown. Contributing to the accident was the first officer's failure to provide required callouts which might have alerted the captain to the airspeed and sink rate deviations. The Safety Board was unable to determine the reason for the captain's lack of awareness or the first officer's failure to provide required callouts.

5. World/Cold Bay/73
About 0542 Alaska daylight time on September 8, 1973, World Airways Inc., Flight 802, a DC-8-63F, crashed into Mt. Dutton, near King Cove, Alaska. The six occupants -- three crewmembers and three nonrevenue company employees -- were killed. The aircraft was destroyed by impact and fire.

Flight 802 was a Military Airlift Command contract cargo flight from Travis AFB, california, to Clark AFB, Philippine Republic, with intermediate stops at Cold Bay, Alaska, and Yokota AFB, Japan. It was cleared for an approach 125 miles east of the Cold Bay Aiport. The flight reported that it was leaving 31,000 feet; this was Flight 802's last recorded transmission. The aircraft crashed at the 3,500-foot level of Mt. Dutton, approximately 15.5 miles east of the airport.

The National Transportation Safety Board determines that the probable cause of the accident was the captain's deviation from approved instrument approach procedures. As a result of the deviation, the flight descended into an area of unreliable navigation signals and obstructing terrain.

6. Texas International/Mena/73
At 2052, September 27, 1973, a Texas International Airlines, Inc., CV-600, N94230, crashed in the Ouachita Mountain Range, Arkansas. The accident occurred 80 nautical miles north-northwest of Texarkana and 8.5 nautical miles north-northwest of Mena, Arkansas. Eight passengers and three crewmembers were killed, and the aircraft was destroyed. The aircraft was making a round trip flight from Dallas, Texas, to Memphis, Tennessee, with intermediate stops at Texarkana, El Dorado, and Pine Bluff, Arkansas. The accident occurred during the westbound flight from El Dorado to Texarkana. The flight was conducted at night under visual flight rules. A cold front with associated thunderstorms and instrument meteorological conditions existed between El Dorado and Texarkana. The crew deviated about 100 nautical miles north of the direct course to their destination and attempted to operate the aircraft visually in instrument meteorological conditions. No radio transmissions were made by the crew after takeoff. The aircraft was found at 1730 c.d.t., on September 30, 1973.

The National Transportation Safety Board determines that the probable cause of the accident was the captain's attempt to operate the flight under visual flight rules in night instrument conditions, without using all the navigational aids and information available to him; and his deviation from the preplanned route, without adequate prior information. The carrier did not monitor and control adequately the actions of the flightcrew or the progress of the flight.

7. Pan Am/Den Pasar/74
At 1562 Greenwich Mean Time on April 22, 1974, a Pan Am Boeing 707 on route from Hong Kong to Sydney crashed into a steep hillside 37 miles north of Den Pasar International Airport, Indonesia. The eleven crewmembers and ninety-six passengers were killed and the aircraft was destroyed.

According to the Aircraft Accident Report prepared by the Directorate General of Air Communications in Indonesia, the probable cause of the accident was "the premature execution of a right hand turn to join the 263 degrees outbound track which was based on the indication given by only one of the ADF's." The ALPA investigator felt that there was no indication of a decision to make a premature turn and instead that the accident was caused by a number of smaller contributing factors including erroneous instruments and the apparent non-utilization of a number of available navaids.

8. Air Florida/Washington, D.C./82
On January 13, 1982, Air Florida Flight 90, a Boeing 737-222 (N62AF), was a scheduled flight to Fort Lauderdale, Florida, from Washington National Airport, Washington D.C. There were 74 passengers, including 3 infants and 5 crewmembers on board. The flight's scheduled departure time was delayed about 1 hour 45 minutes due to a moderate to heavy snowfall which necessitated the temporary closing of the airport.

Following takeoff from runway 36, which was made with snow and/or ice adhering to the aircraft, the aircraft at 1601 e.s.t. crashed into the barrier wall of the northbound span of the 14th Street Bridge, which connects the District of Columbia with Arlington County, Virginia, and plunged into the ice-covered Potomac River. It came to rest on the west side of the bridge, 0.75 nmi from the departure end of runway 36. Four passengers and one crewmember survived the crash.

When the aircraft hit the bridge, it struck seven occupied vehicles and then tore away a section of the bridge barrier wall and bridge railing. Four persons in the vehicles were killed; four were injured.

The National Transportation Safety Board determines that the probable cause of this accident was the flightcrew's failure to use engine anti-ice during ground operation and takeoff, their decision to take off with snow/ice on the airfoil surfaces of the aircraft, and the captain's failure to reject takeoff during the early stages when his attention was called to anomalous engine instrument readings. Contributing to the accident were the prolonged ground delay between deicing and the receipt of ATC takeoff clearance during which the airplane was exposed to continual precipitation, the known inherent pitchup characteristics of the 737 aircraft when the leading edge is contaminated with even small amounts of snow or ice, and the limited experience of the flightcrew in jet transport winter operations.

9. Southern/New Hope/77
At 1619 e.s.t. April 4, 1977, a Southern Airways, Inc., DC-9, Flight 242, crashed in New Hope, Georgia. After losing both engines in flight, it attmpted an emergency landing on a highway. Of the 85 persons aboard Flight 242, 62 were killed, 22 were seriously injured, and 1 was slightly injured. Eight persons on the ground were killed and one person was seriously injured; one person died about 1 month later.

Flight 242 entered a severe thunderstorm between 17,000 feet and 14,000 feet near Rome

Georgia, en route from Huntsville to Atlanta. Both engines were damaged and all thrust was lost. The engines could not be restarted and the flightcrew was forced to make an emergency landing.

The National Transportation Safety Board determines that the probable cause of this accident was the total and unique loss of thrust from both engines while the aircraft was penetrating an area of severe thunderstorms. The loss of thrust was caused by the ingestion of massive amounts of water and hail which in combination with thrust lever movement induced severe stalling in and major damage to the engine compressors.

Major conributing factors included the failure of the company's dispatching system to provide the flightcrew with up-to-date severe weather information pertaining to the aircraft's intended route of flight, the capain's reliance on airborne weather radar for penetration of thunderstorm areas, and limitations in the Federal Aviation Administration's air traffic control system which precluded the timely dissemination of real-time hazardous weather information to the flightcrew.

10. PSA/San Diego/78

About 0901:47, September 25, 1978, Pacific Southwest Airline, Inc., Flight 182, a Boeing 727-214, and a Gibbs Flite Center, Inc., Cessna 172, collided in midair about 3 nautical miles northeast of Lindbergh Field, San Diego, California. Both aircraft crashed in a residential area. One hundred and thirty-seven persons, including those on both aircraft were killed; 7 persons on the ground were killed; and 9 persons on the ground were injured. Twenty-two dwellings were damaged or destroyed. The weather was clear, and the visibility was 10 miles.

The Cessna was climbing on a northeast heading and was in radio contact with the San Diego approach control. Flight 182 was on a visual approach to runway 27. Its flightcrew had reported sighting the Cessna and was cleared by the approach controller to maintain visual separation and to contact the Lindbergh tower. Upon contacting the tower, Flight 182 was again advised of the Cessna's position. The flightcrew did not have the Cessna in sight. They thought they had passed it and continued their approach. The aircraft collided near 2,600 ft m.s.l.

The National Transportation Safety Board determines that the probable cause of the accident was the failure of the flightcrew of Flight 182 to comply with the provisions of a maintain-visual-separation clearance, including the requirement to inform the controller when they no longer had the other aircraft in sight.

Contributing to the accident were the air traffic control procedures in effect which authorized the controllers to use visual separation procedures to separate two aircraft on potentially conflicting tracks when the capability was available to provide either lateral or vertical radar separation to either aircraft.

11. Pan Am, KLM/Teneriffe/77

At 1706 Greenwich Mean Time on March 27, 1977, a KLM Boeing 747 crashed into a Pan Am

Boeing 747, on a runway at Los Rodeos airport, Teneriffe. The KLM Flight from Amsterdam to Las Palmas had been rerouted to Teneriffe, as had the Pan Am flight from New York to Las Palmas, because of the terrorist bombing of the airport. Five hundred and eighty people were killed. There was extensive damage to both aircraft.

The probable cause of the accident as determined by ALPA was the KLM pilot's false hypothesis that the ruway was clear for takeoff. A number of short-term and long-term factors may have contributed to this hypothesis including inadequate visual information and ambiguous or misleading aural information. In addition, information transfer was degraded due to the varying terminology and accents of the flight crews and the controllers.

II. Index and Glossary

This appendix provides definitions for much of the technical terminology, notation and abbreviations used in this report. Exceptions include the following: some particularly well known terms from linguistics, psychology, statistics, and aviation; notations defined and used only wihin the scope of a small portion of the report; and abbreviations and terms whose meaning involves large parts of theories are provided as reminders rather than definitions. Where appropriate, citations to the literature are provided. The parenthesized number refers to the section of this report giving the definition.

Act -- Category in command and control speech act grammar including both physical actions and speech acts. (6.2.1)

Acknowledgement -- Indication that the speaker has heard some report, or that he will perform the action indicated by a request. (3.4)

Ack -- Abbreviation for acknowledgement. (3.4)

Aggravation -- Linguistic strategy which increases the liklihood of an utterance giving offense. (4.1)

ASRS -- Abbreviation for Aviation Safety Report System.

Assertive -- Speech act which commits the speaker (in varying degrees) to the truth of the expressed proposition [Searle 79]. (3.4)

ATC -- Abbreviation for Air Traffic Control.

CAM-1,2,3,4,5,6,7 -- Utterance by captain, copilot, flight engineer, third officer, jumpseat occupant, head flight attendent or flight attendent, respectively, recorded by Cockpit Area Microphone.

Chain -- Sequence of speech acts having the same propositional content. Or, in command and control speech act grammar, a node type which is the top level subordinator of such a sequence. (6.2.2)

Command and Control -- Perspective involving a strict hierarchy of authority in which the giving of commands, reports, acknowledgements, and declarations has a formal and legal status.

Command and Control Coherence -- Variable indicating for any given segment of text, the degree to which it is well-integrated and tightly structured. (10.1)

Command and Control Speech Act Chain -- Sequence of command and control speech acts which all have the same topic. (6.2)

Commissive -- Speech act which commits the speaker to some future course of action [Searle 79]. (3.4)

CRE -- Abbreviation for crew recognized emergency.

Crew Recognized Emergency -- Condition in which the entire crew attends to the situation which led directly to the accident. (5.1)

Crew Recognized Problem -- Situation recognized by the crew as potentially dangerous and not a normal part of flight operations. (5.2)

CRP -- Abbreviation for crew recognized problems.

Critical Segment -- Segment of transcript containing observable degradation or failure of crew coordination which is actually or potentially critical to the completion of the flight. (9.1.2)

CRT -- Abbreviation for cathode ray tube, i.e., video screen for computer display.

Declaration -- Speech act which, if successfully performed, brings about a correspondence between the propositional content and reality [Searle 79]. (3.4)

Directive -- Speech act whih attempts (to some degree) to get the hearer to do something [Searle 79]. (3.4)

Discourse Success -- Of a topic, continuation of the topic in a way that is not operationally relevant. Contrasts with operational success. (8.2)

Discourse Type -- Theory of the structure of a class of discourse units. (6.1)

Discourse Unit -- Segment of talk longer than a single sentence, produced by one or more speakers, with socially recognizable initial and final boundaries, and an internal structure which can be formally described. (6.1)

Draft Order -- Suggested action which may or may not come to have the social force of a command. (7.3)

Dynamic Planning -- Planning which occurs under conditions of changing information (as in the cockpit situation). Contrasts with static planning. (7.2.3)

EXOR -- Exclusive or. (3.2.1)

Explanation -- Discourse unit consisting of a proposition to be demonstrated and a structure of supporting reasons, often with multiple embedded relationships of subordination. (7.2)

Expl -- Abbreviation for explanation.

Expressive -- Class of speech act which expresses a psychological state about a state of affairs specified in the propositional content [Searle 79]. (3.4)

Face -- Public self-image which every community member wants to claim for himself [Goffman 67]. (4.1)

Felicity Condition -- Conditions which must be satisfied in order for a speech act to be properly, i.e., felicitously uttered [Searle 69]. (3.2.2)

Focus -- Presumed focus of attention of the participants in a given discourse. (6.1.1)

Illocutionary Force -- Speaker's intention for the social force of a speech act; that is, what he wishes to accomplish with his utterance [Searle 69]. (3.3.1)

Indirect Speech Act -- Speech act which accomplishes its social force indirectly, that is, which does not mark its social force by its syntactic form or by the specific words it uses. (3.2.2)

Mitigation -- Linguistic strategy which expresses a given propositional content in such a way as to avoid giving offense. (4.1)

Negative Face -- The basic claim to territories, personal reserves, rights to non-distraction -- i.e., to freedom of action and freedom from imposition [Brown and Levinson 79]. (4.1)

Negative Politeness -- Attempts by the speaker to minimize the degree of trespass to the addressee's autonomy [Brown and Levinson 79]. (4.1)

NTSB -- Abbreviation for National Transportation Safety Board.

Off Record Strategy -- Politeness strategies in which the speaker avoids being held accountable for what he intends to convey [Brown and Levinson 79]. (4.1)

Operational Relevance -- Directly involved with successful mission completion. (5.3)

Plan -- Discourse type consisting of the statement of a goal and subordinated actions for achieving it [Linde & Goguen 78]. (7.2)

Positive Face -- The positive consistent self-image or "personality" (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants [Brown and Levinson 79]. (4.1)

Positive Politenssess -- Attempts to minimize the distance between speaker and addressee, so that the speaker's and addressee's desires appear to be the same [Brown and Levinson 79]. (4.1)

Posterior Force -- Social force of a speech act as interpreted by its addressee; determined by making use of the response it actually received in its context. (3.3.2)

Posterior Spectrum -- Range of interpretations of social force and their relative possibility values, as judged by an analyst on the basis of the addressee's response to the speech act. (3.3.2)

Preparatory Condition -- A felicity condition for speech acts, covering what must be satisfied before the act is made; for example, to give an order, a speaker must has appropriate authority over the addressee, and the addressee must have the ability to perform the action [Searle 69]. (3.2)

Prior Force -- Social force of a speech act before it receives a response from its addressee, determined by its linguistic form, the previous context, the identity of its speaker and intended addressee, and the shared information a available to them. (3.3.2)

Prior Spectrum -- Fuzzy set of prior forces; spectrum of possible interpretations of the speech act. (3.3.2)

Projection -- Reports about future states of the world. (3.5)

Propositional Content -- Proposition about the world, which depending on the social force, may be assesserted, requested, denied, etc. by a speech act [Searle 69]. (3.2.1)

Psycho-ostensive -- Non-operationally relevant report of the speaker's psychological state [Matisoff 79]. (3.4)

RDO- 1,2,... -- Utterance by the designated crewmember taken from transcription of radio transmission.

Rank -- The official command and control authority of a participant.

Ratification -- The process by which a draft order or plan acquires the social force of an order. (7.3)

Request -- Speech act type which includes orders, requests, suggestions, and questions. (3.4)

Req -- Abbreviation for request.

Report -- Speech act type which indicates some state of the world. Includes support and challenge. (3.4)

Rep -- Abbreviation for report.

Scale of Mitigation/Aggravation -- See Mitigation/Aggravation Scale.

Social Force -- The effect which a speech act has in the world. (3.1)

Speech Act -- 1. An utterance which directly performs some action in the world [Austin 62]. 2. Category in command and control grammar including reports, requests, acknowledgements, and declarations. (3.2)

Spact -- Abbreviation for speech act.

Speech-Act Chain -- A sequence of speech acts, each of which builds on the previous one so as to preserve the major propositional content. (6.2)

Speech Act Chart -- A graphic device for displaying selected features of speech acts as a function of time, including relevant aspects of propositional content, type of speech act, and speaker. (3.5)

Static Planning -- Planning in a situation in which the information available to the group is static during the period of interaction. (7.2.3)

Subordinator -- Portion of text or node in a tree indicating the specific relationship of subordination holding between two pieces of text. (6.1)

Topic -- The propositional content of an utterance; informally, "what the speaker is talking about." (8.1)

Topic Failure -- Situation in which some speaker introduces a new topic and no other speaker follows it with an utterance having the same topic. (8.2)

Topic Success -- Situation in which some speaker introduces a new topic, and some other speaker follows it with an utterance having the same topic. (8.2)

Transformation -- Internal structure of the planning and explanation discorse types, representing the real-time effects of proposals by members to add, delete, or modify plan or explanation parts [Linde & Goguen 78], [Goguen, Linde & Weiner 81]. (6.1.1)

Tree -- Hierarchical representation of planning or explanation discourse structure showing relations of logical subordination [Linde & Goguen 78], [Goguen, Linde & Weiner 81]. (6.1.1)

* -- In transcript excerpts, indicates the omission of untranscribable material.

\# -- In transcript excerpts, indicates the omission of "non-pertinent" material, in general, obscenity or profanity.