Checklist Interruption and Resumption:
A Linguistic Study

C. Linde
J. Goguen
Structural Semantics

Checklist Interruption and Resumption: A Linguistic Study

Charlotte Linde and Joseph Goguen
Structural Semantics
P.O. Box 707, Palo Alto CA 94302

Abstract

This study forms part of a project investigating the relationships among the formal structure of aviation procedures, the ways in which the crew members are taught to execute them, and the ways in which they are actually performed in flight. Specifically this report examines the interactions between the performance of checklists and interruptions, considering both interruptions by radio communications and by other crew members. The data consist of 14 crews' performance of a full mission simulation of a commercial Boeing 707 flight. The results show that good crews have a higher ratio of checklist speech acts to all speech acts within the span of the performance of the checklist. Further, it is not number of interruptions but length of interruptions which is associated with crew quality. Use of explicit holds is also associated with high crew quality.

Table of Contents

List of Figures

List of Tables

1 Introduction

This study is part of a larger project investigating the relationships among the formal structure of aircraft procedures, the ways in which crew members are taught to execute them, and the ways in which they are actually performed in flight. Specifically, this report examines the interaction between the performance of checklists and interruptions, considering both interruptions by radio communications and by other crew members. This study is an exploratory one, intended to show that a study of checklist interruption can be a valuable tool to suggest possible changes in aircrew training. In order to do this, we have developed a number of linguistic and interactional variables with which to investigate checklists. These can be used to study the ways in which aircrews actually perform checklists, and to demonstrate ways in which training and practice do not match. It should be noted that in situations of mismatch, it is not necessarily the case that further training is necessary. It may be that the actual practice of crew members is preferrable to the actions recommended by the training process. This is an empirically testable question, and should not be decided *a priori*.

1.1 Motivation

Prior research at NASA Ames Research Center indicated that patterns of communication among crewmembers in the cockpit are a significant factor in air carrier accidents: (Ruffell-Smith, 1979) (Foushee and Manos, 1981), (Goguen and Linde, 1983), (Murphy, 1980), and (Murphy *et al*, 1984). Therefore, additional research on aviation communication patterns could provide the basis for changes in training crew members, and formulating aviation procedures. The current research investigates specific patterns of communication involved in checklists which may contribute to such accidents.

1.2 Choice of Checklists

Of all cockpit communications patterns, formally specified procedures may represent the most important way that flightcrews accomplish the communications and actions necessary for mission completion. In fact, procedures specify not only what actions are to be taken, but also the ordering of these actions, and the communications required among crew members to coordinate them. Moreover, since procedures are well defined and highly constrained, they provide a focal point for studying how crews actually interact and to what extent they follow specified forms.

The procedures involved in commercial aviation include checklists, briefings, and callouts.

Of this array of procedures, we focus on checklists. Checklists may be defined as a specified list of actions to be performed in a challenge and response manner. For normal checklists, the actions are checks of activities already performed, rather than immediate performances of those actions. Checklists have been chosen for study, for the following reasons:

1. **Complexity.** Of all procedures, checklists involve the most complex interactions among crew members, and are the most closely related to crew coordination. Hence a study of how they are actually handled by crew members is of the greatest interest.

2. **Frequency.** There are more types and instances of checklists than any other type of procedure. The existence of multiple instances, both within and across crews increases the comparabilty of the data.

3. **Documentation.** The Aircraft Operating Manuals of the two airlines consulted for this project offer the most complete specification of checklists.

4. **Relation to previous work.** Checklist structure is directly related to previous work on planning (Linde and Goguen, 1978, Goguen and Linde, 1983, Goguen, Linde and Murphy, 1984). This provides a convenient theoretical framework within which to carry out an analysis, as well as direct continuity with previous work.

5. **Relevance to crew coordination.** Progress in understanding checklists is relevant to understanding crew coordination, since scheduling and executing checklists involves resource managment, crew communication, and scheduling.

1.3 Choice of Interactional Aspects

There are a number of questions involved in checklist performance. One important issue is the relationship between normative statements of how checklists should be performed, and the ways in which crew members *do* perform them. Such a focus leads to the following kinds of question: What is the accuracy of actual checklist performance? That is, do crew members carry out the checklists in the form prescribed by the Aircraft Operations Manual of their airline? The discovery of problems in this area could lead to recommendations that crews be more extensively trained in following the prescribed form or that the form be changed, if it were shown to cause problems. Another issue concerns the accuracy of responses and the reasons for inaccuracy. When crew members make a response to a checklist query, is that response correct? Since human error of this type is always to be expected, a discovery that crew members were making a sizeable number of these errors would lead to recommendations that training place a greater emphasis on crew concept, so that a second crew member always checks the accuracy of such responses. These are extremely important questions, which are currently under study by other NASA projects. Therefore the current study focuses on a third issue: The types and effects of interactions

between checklist performance and other activities. As discussed in Section 2, the normative statement of how checklists should be performed is that crew members should never allow checklist performance to be interrupted, whether by interaction with Air Traffic Control, or by other cockpit concerns, except in the case of actual emergency situations, which then require additional checklists, embedded within the normal checklist. (See the appendix for a discussion of this type of embedding.) This rule is given both in Aircraft Operations Manuals, and in the course of crew training. Although the rule is stated strongly, with little provision for exceptions, in practice, we find frequent interruptions of both types. This report defines a number of interruption types and measures of interruption, and investigates the degree of compliance with the rules of checklist performance, and the relationship between noncompliance and crew safety performance.

2 Interactive Aspects of Checklist Performance

This section discusses the normative form of checklist performance, and the types of interrruptions which may occur in its course.

2.1 Normative Checklist Performance

Resource management training suggests that checklists should be performed without interruption. Checklists should not be scheduled until other pressing crew concerns have been dealt with, so that the checklist performance can proceed without interruption. This is, indeed, a major aspect of resource management for the pilot flying. Furthermore, if there is a radio transmission to the crew, it is to be ignored until the completion of the checklist. If the pilot flying chooses to respond, the correct procedure is to place an explit hold on the checklist, by saying, "Hold it at *name of checklist item.*"

These procedures for handling checklist performance are given by the Aircraft Operating Manual of a major commercial airline, hereafter referred to as Airline A. (We quote this manual rather than the Aviation Training Institute manual used in the simulator experiment, since the ATI manual does not give an explicit statement of how checklists are to be performed.) Note that the same rules apply for most or all other commercial airlines.

> The pilot flying will request the remaining checklists at the proper time. The flight engineer is not to initiate a checklist, but he should remind the pilot flying if he feels the request for it is overdue.

> A checklist normally should not be started until sufficient time and attention can be devoted to its expeditious completion.

Do not skip items. If the captain elects not to accomplish an item on the checklist at that time, he will say, "Hold the checklist at the _____." When the captain says "Continue the checklist," the reading of the checklist will continue just as though there had been no interruption.

The particular equipment used by the various airlines will dictate how the hold is implemented. Possibilities include cards, manual pages, mechanical scrolls, and computer displays, each of which have different mechanisms for place holding.

2.2 Self Interruption

One possible source of interruption is the crew itself. That is, crew members may choose to interrupt their performance of the checklist in order to deal with other cockpit concerns, to discuss information just gathered from the radio, or to discuss checklist items in an informal manner. Section 4.1 discusses the correlation of such interruptions with poor safety performance. We would expect to find such an effect, since interruption of a checklist places a greater load on memory. Crew members must deal with the interruption, while remembering both the fact that they were performing a check list, and also the identity of the last checklist item they had completed. Furthermore, some member of the crew must make the decision to attempt a resumption of the checklist.

As we have seen, the normative procedure for dealing with interruption is to place an explicit hold on the checklist if the crew must deal with any cockpit concern other than the checklist. If this is done, the load on memory may not be as great, since placing such a hold makes a social acknowledgement of the fact that the checklist has been interrupted. On the other hand, social acknowledgement of the hold may actually dilute the responsibility for resumption, since each crewmember may believe that another crewmember will take the responsibility of resuming the checklist.

In an earlier stage of this project, we studied the formal structure of checklists, and found that they are tree structured plans. In this system, a formal hold on a checklist constitutes a POP marker, that is , an indication that the focus of attention for the discussion has moved out of the plan, with an associated marker of where to return to. (See the appendix for a fuller discussion of this structure.)

2.3 Radio Interruption

Another source of interruption is radio transmission, either to the crew, or to or from another aircraft. The crewmembers have two choices in this situation. They may continue their performance of the checklist during the radio transmission. In this case, we may assume either that the crewmembers are not listening to the radio, or that they are attempting to attend to both the radio and the checklist. This may temporarily increase their workload, and may increase the likelihood of error in one or both tasks. Or, they may interrupt their performance of the checklist in order to listen to the radio. Such an interruption may be either unmarked, or marked with an explicit hold. We also note that an interruption to listen to or respond to the radio is often followed by conversation between crew members of the information gathered from the radio, thus extending the period of the interruption.

3 Choice of Data

The data for this study comes from a full mission simulation of a commercial Boeing 707 flight. (See (Murphy *et al*, 1984) for a description of this project.) This section describes the scenario of this simulation, and the choice of data from the scenario.

3.1 The Scenario

The simulation scenario represented a flight from Tucson to Phoenix, continuing to Los Angeles. Each crew flew the scenario once, with no prior knowledge of the problem that was to be introduced. The scenario was designed to produce a series of overlapping problems. First, the crew was given a hold at the Peris intersection in order to burn fuel reserves. After a hold at Los Angeles International Airport (LAX), an indication that the nose gear was not down and locked forced a missed approach on the landing. This situation was exacerbated because it occurred at a time when the entire Los Angeles basin, including possible alternate airports (e.g. Ontario (ONT) and Long Beach (LGB)) was experiencing low and deteriorating ceilings due to incoming coastal fog. After the missed approach, the crews performed the gear check procedure to determine that the gear was down and pinned; this procedure permitted them to assume that the panel light indication was faulty. The crews were then required to choose an alternate airport; deteriorating weather suggested that only Palmdale and Ontario might be open. The decision was not a clear one: while Palmdale had better weather than Ontario, it is not equipped with commercial passenger handling facilities.

While on the ground at Phoenix (PHX), the crew was given weather information indicating some degradation at LAX. During the latter part of the cruise to Los Angeles, they were given direct information and other cues about the further deterioration of the ceiling and runway visual range at LAX. These cues included the hold at Peris due to traffic backup, and a conversation between ATC and another aircraft about its missed approach at LAX and its return for a second attempt. By requesting weather conditions at possible alternates, the crews could determine that conditions at other coastal airports (such as Long Beach (LGB)) were similar to LAX, while Ontario (ONT), which is located inland from LAX was behind LAX in weather and visibility deterioration, and that Palmdale (PMD), located over a mountain range out of the Los Angeles basin, had good weather with clear visibility.

3.2 The Simulator and Crew

The scenario was flown in a Boeing 720B flight training simulator, a late version of the Boeing 707, which was leased from Aviation Training Institute. A current, professional air traffic controller was used in the simulation. Crew members in the simulation were paid volunteers, whose experience represented a wide range of airline of origin and recency or currency on B-707 linde operations. Some were current on the B-707, others had recent B-707 experience but were currently flying other jet aircraft in line operations, and some were retired. Crew composition ranged from one crew in which all members were retired from line flight, to one crew which was currently flying the B-707 as an intact crew. All crewmembers received six hours of classroom differences training and four to eight hours of simulator differences training. The number of hours of simulator differences training which a crewmember received was based on recency. Subjects were formed into crews before the simulator training and were instructed in coordinated procedures during this training.

3.3 Choice of Data for the Present Analysis

All cockpit communications during the second part of the scenario, the flight from Phoenix to Los Angeles, were transcribed. However only 14 of the 16 flights were used, because the other two transcriptions were not available at the time of the linguistic coding.

In order to study the widest range of checklist execution, checklists from both normal and problem segments of the flight from Phoenix to Los Angeles were sampled, and checklists performed both on the ground and in the air were included, since workload demands appear to differ radically in these conditions. Checklists which were performed explicitly by all the crews were selected, since many crews omitted explicit verbal performance of one or more checklists. Table 1 shows the checklists performed by each crew. The three checklists

chosen for analysis were the afterstart checklist, which is performed while still on the ground, the instrument approach checklist, and the prelanding checklist.

Note that some checklists were performed more than once, since the scenario forced a go-around at LAX. Also note that many checklists were omitted, even by crews given high ratings in safety performance. Contrary to expectations, there appears to be no relation between crew quality and number of checklists performed.

Checklists Performed	Crew Number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	16
Before Start	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Starting Engines	1	1	1	1	1	1	1	1	1	0	1	1	0	1
Afterstart	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Waveoff	1	1	1	0	1	1	1	1	1	0	1	1	1	1
Before Takeoff	1	1	1	1	1	1	1	0	0	1	1	1	1	0
Climb	1	1	1	0	2	1	1	?	1	1	1	1	2	0
Inrange	1	1	1	1	1	2	1	2	1	0	2	3	2	1
Instrument Approach	2	2	2	1	3	2	2	1	?	2	3	3	1	2
Prelanding	1	2	1	1	2	1	2	3	?	1	3	3	1	1
Afterlanding	0	0	0	1	1	1	0	1	0	1	1	1	1	0
Parking	1	0	1	1	0	0	1	1	?	0	1	1	1	0
AC Terminating	1	0	0	0	0	0	0	0	0	0	1	1	0	0
Totals	12	11	11	10	13	12	12	13	9	8	16	18	13	7

Table 1: Number of Checklists Performed by Each Crew

The checklists used in the flights were those provided by Aviation Training Institute, the site of the simulator. Because the subjects came from a number of different airlines, the forms most familiar to them could not be used. All crewmembers received preflight training to familiarize them with this form.

There were a number of instances in which crews began a checklist, but did not finish it. These cases were not coded, since it is impossible to determine exactly the exact point at which the checklist was abandoned. Hence, it was impossible to determine the total number of speech acts. Table 2 shows the number of checklists performed by each crew, and indicates the variation between crews.

CREW	UNCOMPLETED CHECKLIST	CREW RANK
1	Instrument Approach	2
6	Instrument Approach Instrument Approach Prelanding	12
9	Instrument Approach Prelanding	8
10	Instrument Approach	11
13	Instrument Approach	10

Table 2: Uncompleted Checklists, by Crew

4 Variables

Several methods were devised to measure the interaction of checklist performance with other activities: the continuity ratio of checklists, the treatment of interruptions, and the treatment of resumptions.

4.1 Continuity Ratio

The performance of a checklist defines a **checklist span**: the period of time between the first call for the checklist and the conclusion of the checklist. Within the checklist span, we may consider the **continuity ratio**: the ratio of checklist speech acts to the total number of speech acts within the checklist span. Checklist speech acts are only those utterances which form part of the challenge/response checklist form specified by the Aircraft Operating Manual. They do not include either discussion about when or whether the checklist should be performed or speech acts which discuss checklist topics in a form not given by the Aircraft Operating Manual, which are viewed as interruptions to the checklist.

4.2 Interruptions and Resumptions

As already discussed, there may be one or more interruptions within a given checklist, either by the radio, or by the crew themselves. For the codable checklists there are the same number of interruptions and resumptions.

4.3 Identity of Crew Member Requesting Resumption

Once a checklist is interrupted, it may be resumed by any crew member. Airline procedures do not specify who is required to resume it. The identity of the crew member requesting resumption is of great interest, since it shows who is taking responsibility for this aspect of resource management. We have coded the identity of the crew member requesting the resumption: Pilot Flying, Pilot Not Flying, or Second Officer.

4.4 Radio Interruptions

There are four possible ways in which a radio transmission may interrupt checklist performance. These are:

1. **Radio to Crew Overlap.** This variable represents the number of speech acts in a radio transmission to the crew which overlap the performance of the checklist. (That is, the crew continues the performance of the checklist while ATC is addressing them. A simultaneous utterance of more than one word is coded as an overlap.)

2. **Radio to Other Overlap.** This variable represents the number of speech acts in a radio transmission to or from any other aircraft which overlap the performance of the checklist. (That is, the crew continues the performance of the checklist during radio traffic between ATC and other aircraft.)

3. **Radio to Crew Interrupt.** This variable represents the number of speech acts in a radio transmission to the crew which interrupts the performance of the checklist. (That is, the crew interrupts the performance of the checklist to listen to a radio transmission directed to them by ATC.)

4. **Radio to Other Interrupt.** This variable represents the number of speech acts in a radio transmission to or from any other aircraft which interrupt the performance of the checklist. (That is, the crew interrupts the performance of the checklist during radio traffic between ATC and other aircraft.)

Note that although it is logically possible to have a crew intiated radio transmission which interrupts the performance of the checklist, in practice we find no such examples.

4.5 Safety Performance Variable

In order to understand the effects of variation in checklist performance, performance on one or more activities must be measured to distinguish effective from ineffective crews. Unfortunately, such performance variables are currently not available in reliable form.

Other analyses performed on this data (Murphy and Awe, 1985), provide a number of

variables derived from a study in which six professionally active retired captains rated videotapes of the simulated flights on a number of crew coordination and decision-making variables. The raters were six retired captains, all of whom are currently employed as analysts or researchers by the NASA/Aviation Safety Reporting System. The raters recieved three 2-hour training sessions, which included an explanation of the rating system and the problems of rating, trial rating runs, and discussion of the trials.

Of the variables included in the peer rating study, the one that seems most appropriate for the present study is safety performance. It is the closest of the measures to a single objective (or near-objective) measure of the quality of overall crew performance. The safety performance variable reflects the raters' assessments of the risk involved in a given crew's solution to the major scenario problem. The judgment of the level of safety performance includes the choice of airport for landing was made, the amount of fuel on board at landing, and the altitudes reached during below minimum approaches when the runway was not visible.

There are two ways in which such a measure of crew performance can be assessed. The first is to validate it against objective measures of performance such as error rates, fuel on landing, and so forth. Such comparisons have not yet been undertaken. The second is to examine correlations among the set of peer rating variables, to learn how highly intercorrelated the safety performance judgments were among the raters and whether they are associated with other ratings of crew performance. These relationships have been examined by (Murphy and Awe, 1985). Inter-rater reliability for safety performance judgments is extremely high ($r=.99$), indicating that it is appropriate to sum and average the ratings on this dimension. Ratings on other dimensions are also highly reliable. However, an examination of intercorrelations among mean ratings reveals that judged safety performance is significantly associated ($df=14$, $p<.05$) with only two other dimensions.[1] Moreover without objective validation, the extent to which these associations are affected by artifacts in the peer review method is unclear.

Given the problems with ratings of safety performance, we must treat it as a variable of unknown quality whose validity at present is uncertain. It is used here for exploratory purposes because no other, more fully validated, measure is available at this time.

---

[1] While (Murphy and Awe, 1985) report all intercorrelations between rating variables to be significant, their findings assume 90 degrees of freedom. However assessing relationships between these variables for the 16 crews requires first combining the ratings from the 6 judges on each; the resulting degrees of freedom for each obtained correlation is 14. This reanalysis yields only two significant correlations between Safety Performance and any of the other variables; these are Decision Quality and Decision Efficiency.

5 Results

This section describes the testing of a number of hypotheses using the checklist performance variables discussed in Section 4.

5.1 Continuity Ratio and Checklist Performance

The hypothesis is that good crews have a higher continuity ratio in their checklist performance. This hypothesis tests the explicit instruction that checklists take priority over all other cockpit acitivities. If this instruction is correct, it should mean that indeed the better crews have a higher continuity ratio, because they permit fewer and shorter interruptions of their checklist performance.

5.1.1 Computation of Continuity Ratio

In order to test this hypothesis, we must determine the **continuity ratio** for each crew. The continuity ratio is computed as follows:

$$\frac{\text{Number of Checklist Speech Acts in Checklist Span}}{\text{Number of All Speech Acts in Checklist Span}}$$

Note that in computing this ratio, we have used only those checklists in our sample performed while in the air (the Instrument Approach and Prelanding checklists), since the checklist performed while on the ground (the Afterstart checklist) was interrupted rarely and there was very little variance between crews. Checklist speech acts include only those speech acts which accomplish the checklist in the challenge/response manner prescribed by the checklist; however, variations in the wording of the challenge or response are not considered. The total of all speech acts in the checklist span includes checklist speech acts, meta-checklist speech acts, nonstandard checklist speech acts, radio speech acts by the crew, by ATC to the crew and to other aircraft, and by other aircraft, and discussions by the crew of nonchecklist topics. (It would be interesting and valuable to study the patterning of each of these types separately, but the present data set is not large enough to support this.)

As mentioned in Section 2.1, a checklist may contain an explicit hold. If such an explicit hold is used, the speech acts between the explicit hold and the resumption of the checklist are not counted as part of the total of the checklist span. The reason for this is that an explicit hold may be considered to be a formal suspension of a checklist, rather than an

interruption of it.[2]

A number of checklists contain material at the end that is to be performed silently by the flight engineer. When he has concluded these silent items, he announces that the checklist is complete. We have not counted as part of the checklist span speech acts coming between the last checklist item to be performed aloud in challenge/response fashion and the explicit announcement of the end of the checklist, since no further attention to the cheklist is necessary for other crew members. We have, however, counted it as part of the checklist span if the flight engineer read this material aloud.

5.1.2 Testing the Hypothesis

The safety performance variable discussed in Section 4.5 ranks crews on a seven point scale. The ranking on this scale was used to divide the 14 crews into the top and bottom seven. Table 2 shows the continuity ratios for the 14 crews.

	TOP CREWS			BOTTOM CREWS	
Crew	Safety Perf. Mean	Interruption Ratio	Crew	Safety Perf. Mean	Interruption Ratio
1	6.0	.75	5	1.17	.49
2	5.33	.61	6	1.5	.58
3	3.67	.52	8	1.83	.47
4	5.33	.33	10	1.83	.12
7	4.67	.78	11	2.67	.51
9	2.83	.67	12	2.67	.45
16	6.33	.52	13	2.33	.53

Table 3: Continuity Ratios for Top and Bottom Groups of Crews

The Mann-Whitney U statistic is used to test whether the differences between the two groups were significant. This test yields U=10, p=.036. The hypothesis is therefore accepted. In fact, the actual probability might have been higher, since the exigencies of the coding system eliminated some of the lowest continuity ratios by eliminating those checklists which could not be coded because they were interrupted and not resumed. As table 2 shows, with one exeption, crews which have uncompleted checklists have scores which fall in the lower half of the safety performance scale .

---

[2]This interpretation is supported both by the the Aircraft Operations Manual discussion of checklist performance cited in Section 2.1, and by discussion with Captain Jack Raabe, a retired Pan American Airlines check pilot now at the Batelle Institute.

5.2 Comparison of Number of Interruptions with Interruption Ratio

It might be argued that the total number of interruptions is the crucial factor in checklist peformance, rather than the interruption ratio. That is, three interruptions of five speech acts each may be as bad or worse than interruption one of fifteen speech acts. We can test this by computing the ratio of the number of interruptions to the number of checklists performed. (This ratio provides a necessary normalization, since not all crews peformed the same number of checklists.)

As in the previous hypothesis, the safety performance variable is used to divide the crews into the top and bottom seven. Table 3 shows the ratios for the 14 crews.

| | TOP CREWS | | | BOTTOM CREWS | |
Crew	Safety Perf. Mean	Interruption Ratio	Crew	Safety Perf. Mean	Interruption Ratio
1	6.0	--	5	1.17	.67
2	5.33	.75	6	1.5	.67
3	3.67	1.0	8	1.83	.25
4	5.33	.33	10	1.83	2.0
7	4.67	.20	11	2.67	.20
9	2.83	--	12	2.67	.25
6	6.33	.20	13	2.33	.50

Table 4: Ratio of Number of Interruptions to Number of Checklists Performed

The Mann-Whitney U statistic is again used to compare the two groups. This test yields $U=19$, $p=.5$. The hypothesis is therefore rejected, and we conclude that number of interruptions is not a factor which differentiates good and poor crews.

5.3 Identity of Crew Member Resuming Checklist

When a checklist is interrupted, someone must resume it. The crew member calling for the resumption may be the pilot flying, the pilot not flying, or the second officer. Effective resource management would dictate that the pilot flying should call for the resumption, since he is the one who should remember that the checklist was suspended, and must attend to the the fact that the checklist has been suspended, and to determining a suitable time to resume it. Table 4 shows the ratio of resumptions by the pilot flying to all resumptions in interrupted checklists.

Crew	Safety Perf. Mean	Resumption by Pilot Flying	Resumption by Pilot Not Flying	Resumption by Flight Eng.	Resumption Ratio for Pilot Flying
1	6.0	--	--	--	--
2	5.33	2	0	2	.5
3	3.67	0	0	3	0
4	5.33	0	1	0	0
7	4.67	1	0	0	1
9	2.83	--	--	--	--
16	6.33	0	0	1	0
5	1.17	2	0	2	.5
6	1.5	0	0	3	0
8	1.83	1	0	0	1
10	1.83	0	0	2	0
11	2.67	0	1	0	0
12	2.67	0	0	1	0
13	2.33	1	0	1	.5
Total		7	2	15	

Table 5: Ratio of Resumptions by Pilot Flying to All Resumptions

Using the Mann-Whitney U test to compare resumptions in good and poor crews, as defined by the safety performance measure, we obtain U=16.5, p >1.0. We see that there is no relation between which crew member resumed a checklist and safety performance, although such a relation would be predicted by the fact that the Aircraft Operations Manual prescribes that resumptions should be by the pilot flying. Note, however, that the lack of signifcance may be the result of the number of tied ranks in the data. It is interesting, however, to note that 63% of the resumptions were by the flight engineer, for both good and poor crews, contrary to the instruction of the Aircraft Operations Manual.

5.4 Explicit Holds of Checklists

Correct checklist procedures require that the crew place an explicit hold at the next item to be performed when a checklist is interrupted. That is, the pilot flying should say something like "Let's hold it at [item]." Examples of explicit holds were surprisingly rare, and so the number of instances is too small to permit statistical testing. However, note that the only two crews which did use explicit holds are crew 16 (Rank = 1) and crew 2 (Rank = 3). (Crew 1 (Rank = 2) has no interruptions, and hence no occasion for explicit holds.) This distribution suggests that in spite of training which specifies the use of explicit holds, only

the best crews followed this instruction. (It might be argued that this is a phenomenon particular to the simulator situation. However, (Linde and Goguen, 1987) showed that in at least some aspects, simulator crews appear to be on good behavior and more attentive to proper procedure than they are during actual flight.)

5.5 Overlaps Versus Interrupts

We may now consider the distribution of radio interrupts and radio overlaps. As discussed in Section 2.3, a radio transmission may occur while the crew is performing a checklist. In this case, the crew may either continue with the checklist, thus treating the transmission as an overlap, or they may suspend the checklist, treating it as an interrupt. As mentioned in Section 2.1, the recommendation is that it be treated as an overlap; that is, the crew should ignore the transmission, even if it is directed to them, until completing the checklist. Table 5 shows the numbers of interrupts and overlaps for each crew.

Crew	Number of Overlaps	Number of Interrupts
1	2	0
2	1	0
3	1	1
4	2	2
5	3	5
6	3	1
7	1	1
8	1	0
9	0	0
10	8	1
11	2	0
12	4	0
13	3	2
16	6	2
Total	37	15

Table 6: Number of Overlaps and Interrupts for Each Crew

Given the size of the data set, it is not possible to distinguish between the treatment of overlaps and interrupts by good and bad crews. However, it is interesting to note that 28% of the instances are treated by the crews as interrupts, which is contrary to training and to normative policy on checklist preformance.

6 Discussion and Conclusions

Our results demonstrate that good crews have high checklist continuity ratios. Furthermore, it is not the *number* of interruptions, but the *length* of interruptions that is associated with crew quality. This seems reasonable, since crew members cannot fully control the number of interruptions, but they can exercise some control over how long the interruptions last. That is, a crew has control over its own interruptions of a checklist, but not over radio transmissions. However, crew members *can* control whether they discuss a radio transmission or whether they immediately return to the checklist. Likewise, when some other matter requires immediate attention, the crew has control over whether they place an explicit hold on the checklist, or whether they interrupt the checklist to discuss that matter, without making an explicit decision to hold and then return to the checklist. This difference between the number and length of interruptions suggests that a greater burden may be placed on the memory by one long interruption than by several short ones.

The data suggested that only the best crews used explicit holds to suspend the checklist. An explicit hold is a linguistic device which changes the social status of an interruption. If the explicit hold is used, the crew members have a linguistic acknowledgement that the pilot flying has turned from the checklist to some other matter of concern, but intends to return. Without an explicit hold, crew members are not certain of whether the checklist has been interrupted or abandoned. In terms of the formal model of checklists given in the appendix, a checklist may be seen as a tree-structured plan. Execution of such a plan involves a movement of the crewmembers' focus of attention as they proceed through the checklist. An explicit hold corresponds to a POP marker, an explicit indication that the focus of attention has moved from some point within the tree, indicating some item on the checklist, to the top node of the tree, indicating a shift in the focus of attention to something other than the checklist. Other studies of traversal of large-scale linguistic structures have shown that such large, non-sequential movements require explicit linguistic marking (Linde and Goguen, 1978). In the case of checklists, the formal theory provides a description of the relation between the explicit marking and the rest of the checklist.

It is now reasonable ask what value this line of investigation may have. There are several possible types of application. One immediate goal is to develop measures for investigation of checklist performance which go beyond simple questions of accuracy of response. Such measures have been developed by this study, and can be applied easily to other types of aviation data.

The second potential application for this research is for crew training. At present, crew training in checklist execution is focussed on doing the checklists precisely "by the book." However, a review of interruptions and resumptions that occur in practice suggests situations in which doing it by the book is not possible. This study found a number of instances in which almost all crews did *not* do it by the book. These include ignoring the instructions to pay no attention to radio transmissions while performing a checklist and to give checklist performance priority over all other cockpit concerns.

These findings suggest that further investigation is needed to determine whether such instructions are, in fact, correct and should be stressed more during training, or whether they do not lead to optimum performance, and should be modified. Furthermore, the instruction to mark an interruption with a formal and explicit hold was very rarely carried out. Again, this suggests the need for further research to determine whether the instruction is justified, and if so, how to develop more effective training procedures.

Another application of the measures developed by this study is the development of a simple test of the nature and overall quality of crew interaction. That is, certain of the variables proposed above, such as the average length of interruptions, may indicate the nature and quality of crew interaction and coordination. Other variables, such as whether radio transmissions overlap or interrupt checklist performance can serve as a measure of crew attention to outside information. This interpretation of these variables can be checked by correlations with variables of crew performance derived from a peer review method, from objective studies of performance errors, or from studies using focus group techniques and interviews to obtain the jugements of experienced flight crew members. If correlations are found between these various types of measure, and the measures of checklist performance, this could lead to a measure of crew interaction quality which which would be simpler and less costly than large-scale linguistic or psychological studies.

I. Appendix: The Formal Structure of Checklists

Research has shown that the formal structure of checklists is best studied by viewing checklists as *plans*, in the precise formal sense described in (Linde and Goguen, 1978, Goguen and Linde, 1983). Just as it is possible to write a formal grammar to describe the syntax of sentences of a given language, it is possible to specify a formal grammar for larger units such as plans, or checklists. This appendix reviews the theory of discourse analysis required for such description, gives the grammar of plans, and discusses the modification of this grammar necessary to describe checklists.

I.1 Discourse Unit and Discourse Type

The larger units of language that are appropriate for the study of aviation communication are called discourse units (see (Goguen and Linde, 1983, Goguen, Linde and Murphy, 1984)). A **discourse unit** is a segment of spoken language, longer than a single sentence, having initial and final boundaries that are socially recognizable, and having a formally definable internal structure. (This definition generalizes the criteria given by (Labov, 1972) for the narrative of personal experience.) A **discourse type** is a class of discourse units having the same internal structure. Discourse types that have been studied include the narrative (Labov, 1972), the spatial desciption (Linde, 1974, Linde and Labov, 1975), the joke (Sachs, 1974), small group planning (Linde and Goguen, 1978), explanation (Goguen, Weiner and Linde, 1983), and the command and control speech act chain (Goguen and Linde, 1983). There are a number of points to be made about these definitions:

1. **Level of Unit.** In the linguistic hierarchy, the discourse unit is immediately above the sentence, and hence is composed of sentences.

2. **Socially Recognized Boundaries.** The discourse unit has boundaries which are recognized as such by the participants in the conversation. These boundaries are often recognized through their stereotyped form; for example, **They lived happily ever after.** as the end of a fairy tale, **It seems there was a ...** as the beginning of a joke, **And that was it.** as the end of a narrative. Or they may be recognized as encoding a certain type of semantic information; for example, an abstract of a story, summarizing its point, can serve as an initial boundary.

3. **Formally Definable Internal Structure.** Labov has given an account of the structure of narrative which is, in effect, a phrase structure grammar (Labov, 1972). Plans and reasoning have been described using transformational grammars in which the transformations mirror the real-time additions, deletions, and modifications made by

speakers (Goguen, Weiner and Linde, 1983, Linde and Goguen, 1978). Such a grammar defines a discourse type as the class of discourse units whose internal structure is consistent with those transformational rules.

It has been found that the most important discourse types in the study of crew communication are planning, reasoning, and the command and control speech act chain. Instances of narrative and pseudonarrative are also present in cockpit communications, but they are used only in non-operationally relevant ways. Only planning has been found to be directly relevant to the study of checklists, since checklists may be described as a specific type of plan.

I.2 Theory of Planning

Small group planning plays a basic role in aviation discourse. Planning may be viewed as a linguistic and interactional activity carried on by a group of people, rather than as an individual mental activity carried on by a single person. The linguistic study of small group planning (Linde and Goguen, 1978) has shown that the language used to accomplish planning is a discourse type: 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 acceptance or rejection of it. And it has a precise internal structure, consisting of members' proposals to add new subplans, and to modify or replace parts of plans previously proposed by others.

Formally, the internal structure of a planning discourse unit is described as a sequence of **transformations** on the plan being formed by the group. These transformations represent the real-time effects of proposals by members to add, delete, or modify plan parts. (Note that beginning a plan, by stating a goal, is also a transformation, in this case, a transformation of addition.) The relations of logical subordination that hold among the plan parts are represented by a **tree** structure. Figures I-1 and I-2 show a plan from the 1978 United Airlines accident near Portland, Oregon (NTSB, 1979). Its major goal, stated by the first officer, is to **call out the equipment**, and his plan for this is to **have the company call**. This PLAN/GOAL relationship is indicated in Figure I-1. In Figure I-2, the Captain *replaces* the First Officer's plan with a plan to **call dispatch in San Francisco**. In Figure I-3, he *adds* a node indicating that **maintenance down there will handle it that way**.

The order of application of transformations is the same as the order of production of clauses

```
CAM-2   He's going to have              PLAN/GOAL
        the company call               /    \
        out the equipment?            /      \
                                  have the    call out
                                company call   the equipment
```
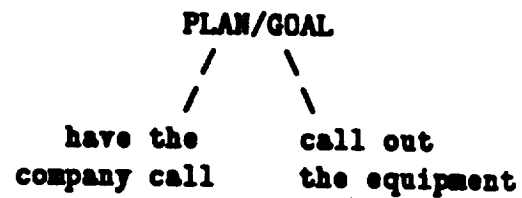
Figure I-1: A GOAL/PLAN Node

```
CAM-1   We'll call dispatch             GOAL/PLAN
        in San Francisco               /    \
                                      /      \
                                 call out     ACTOR/SAY/TO
                                the equipment   /   |   \
                                               /    |    \
                                              we   call  dispatch
                                                         in San
                                                         Francisco
```
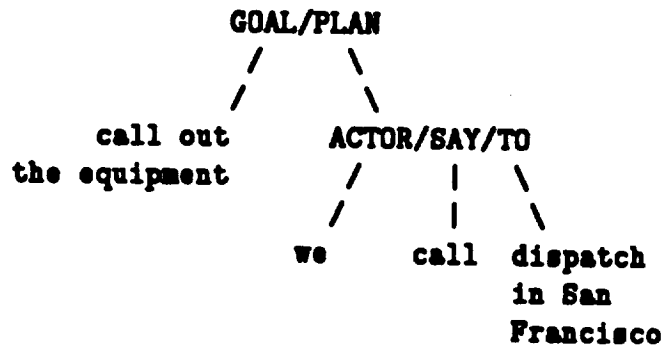
Figure I-2: Addition of an ACTOR/SAY/TO Node

```
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
```
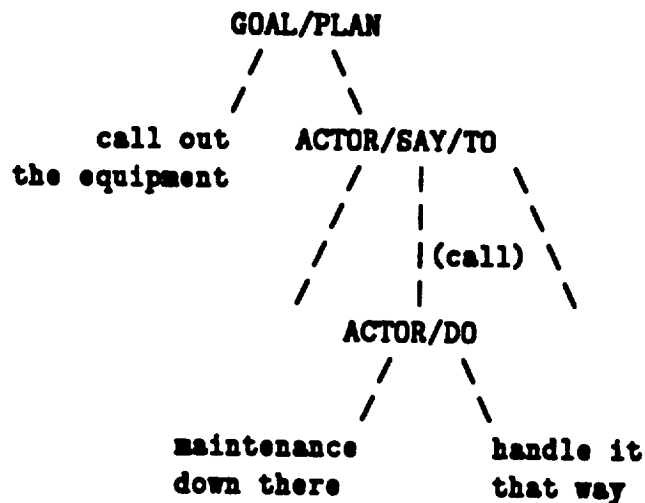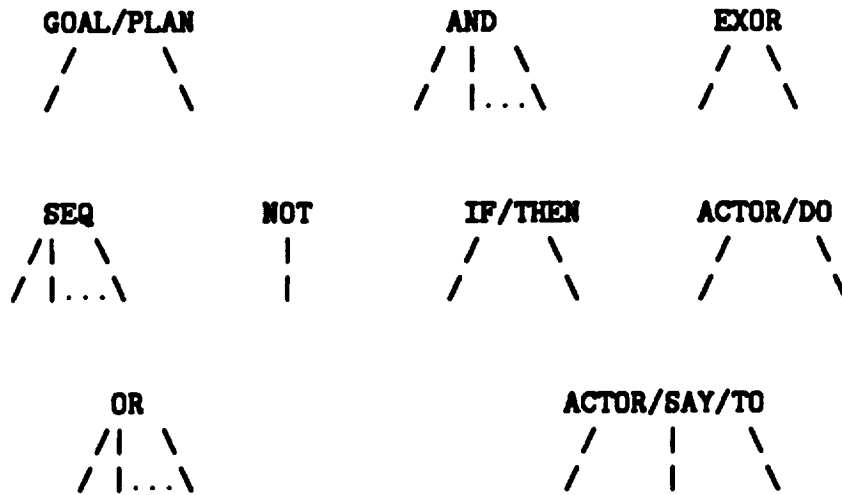
Figure I-3: Addition of an ACTOR/DO Node

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. However, the present research on procedures requires only simple addition transformations; see (Linde and Goguen, 1978) for a full list of transformations.

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 of plans; IF/THEN, for a conditional plan or goal; and ACTOR/DO, with 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 instead are filled in with language produced by the speakers. Note that the parts of compound nodes may be freely permuted, depending on the order in the text; thus, we find GOAL/PLAN and PLAN/GOAL, IF/THEN and THEN/IF, etc. See Figure I-4 for a display of all the subordinators found in previous research on planning.

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


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


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

Figure I-4: Subordinators Found in Planning

I.2.1 Structural and Interactional Properties of Checklists

We now consider the checklists in the Aircraft Operations Manual of Airline A. There are altogether 28 such checklists, 2 Normal (each with 6 named subprocedures), 12 Abnormal, and 13 Emergency. One measure of their complexity is the number of nodes in the plan tree describing them. Among the checklists having explicit Challenge/Response graphical structure, this ranges in complexity from RUNAWAY TRIM, the simplest checklist with just 3 nodes, to ELECTRICAL SMOKE OR FIRE, by far the most complex checklist of Airline A, with a total of 88 nodes. To indicate typical structures for these checklists, Figures I-5 and I-6 show respectively the plan trees of two of the Emergency Checklists, the Phase I (memory items) of APU FIRE and (all of) ENGINE OVERHEAT.

The structure shown in Figure I-5, a SEQuence of Challenge/Response pairs, is particularly characteristic of checklists. A Challenge/Response pair is indicated by a CH/R node; the question of how to interpet these nodes in terms of the primitive node types given in the
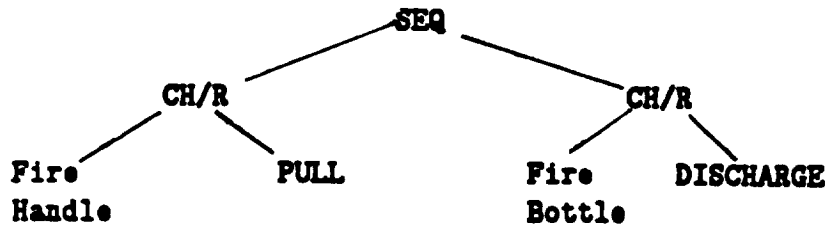
Figure I-5:   APU FIRE Phase I Checklist

theory of planning is discussed in Section 4.2.1. For reference, we include the text for this checklist:

> **Fire Handle**
>   .... **PULL Fire Bottle**.................... **DISCHARGE**

The reader may have noticed that these plan trees do not a GOAL/PLAN root node; this is because no explicit goal is indicated in the specification of this procedure in the AOM. The default goal can be taken to be accomplishing a safe landing.

Having illustrated the most typical structures of checklists, a complete checklist is now analyzed. Figure I-6 shows the complete structure of the simplest Emergency Checklist, ENGINE OVERHEAT. Only the initial CH/R node is in Phase I; the rest represents Phase II checklist items. This checklist contains a Condition/Action node (indicated COND/ACTION); this node type is common in Emergency/Abnormal checklists. The Aircraft Operations Manual analyzed here uses a special notation for conditional actions. For example, the first COND/ACTION node of Figure I-6 appears in the manual, with its Condition in **boldface**, as follows:

> ● **If overheat light goes out,**
>   operate engine at reduced thrust.

Finally, the GOTO node in Figure I-6 indicates that if this point in the tree is reached in executing the plan, then the designated checklist, ENGINE FIRE, should be executed.
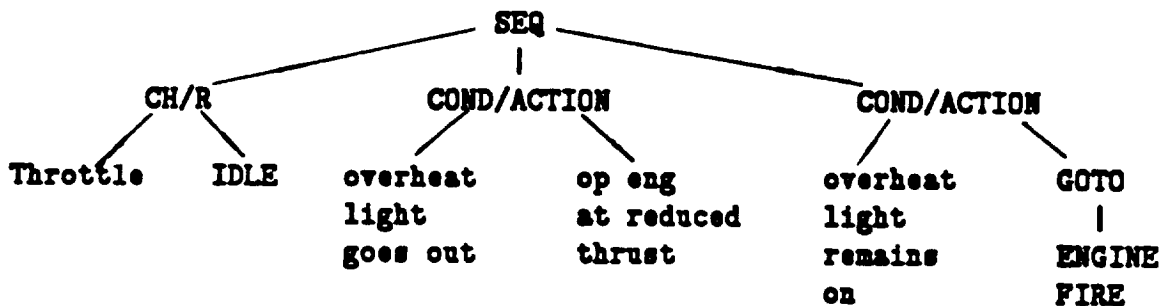

Figure I-6:   ENGINE OVERHEAT Checklist

I.2.2 Focus of Attention

In addition to the subordinators which are used to build up the tree, the description also requires a mechanism to describe the focus of attention of the participants. We use a **pointer**, which is a formal marker of position in the plan tree corresponding to the participants' sense of "what we are doing now." (Note that this pointer is part of the abstract description of the process of planning. It may or may not have an equivalent in the physical world, such as a crewmember's finger moving down the checklist, or a cursor on a terminal display.) In general, the next transformation will apply at the node marked by the pointer's current location.
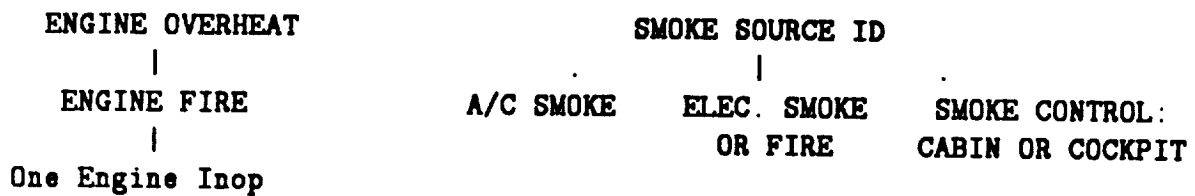
Pointers may move as part of the application of some other transformation, or they may be moved by a transformation whose only effect is the shift of focus of attention. All the subordinators described in Section I.2 have corresponding addition transformations, with the effect of moving the pointer to the newly added node in the plan tree. This new node becomes the current focus of attention, and the next transformation will apply at that node, unless the focus, and thus the pointer, shifts to some previously added node. A transformation which moves a pointer upward in the tree without adding a node to the tree is called a *POP* transformation.

In ordinary discourse types, POPs may be indicated by such linguistic markers as "so," "well," "anyway," "OK", and summaries of the preceding text. In checklist performance, the normative indicator of a POP marker is a linguistic indication of explicit hold. This indicates both that the social focus of attention has been moved to a task indicated by another node, and that a return to the original task is intended.
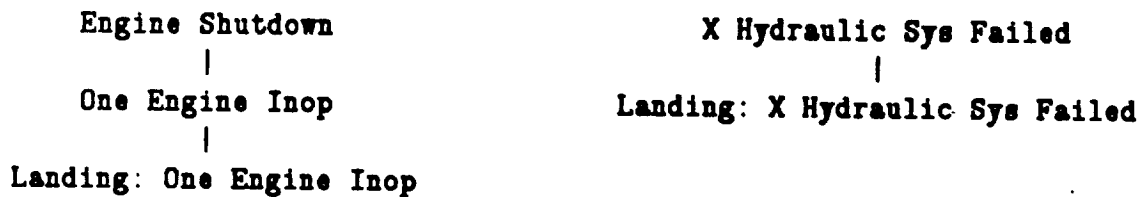
I.2.3 Dependencies among Checklists

We now turn to the question of dependencies among checklists. One checklist **depends** upon another if the first checklist requires the performance of the second when some specified condition holds. This is formally indicated by a GOTO from the first checklist to the second checklist, an instruction which requires the crew to go to the indicated checklist. Such plan GOTOs are indicated in Figures I-7 and I-8. Theoretically, at least three forms of dependency are possible: either the crewmembers interrupt their performance of the first checklist in order to perform the second, and then return to it, or they continue with the second checklist and do not return to the first, or or they are referred from an abnormal checklist to a normal checklist with some actions of that checklist edited. The data we have examined contain only the second case, which we may refer to as embedding.

Normal checklists do not have a complex dependency structure. They do not embed other normal checklists, although the response to a challenge in a normal checklist may be a report of the performance of an individual checklist. Emergency/abnormal checklists may embed other emergency checklists under certain designated conditions, or may unconditionally require the performance of another emergency/abnormal checklist at a later time. Finally in the most complex case, an abnormal checklist may refer to a normal checklist with some actions of that checklist edited. The abnormal checklist which results from the edit may be considered to depend on the altered normal checklist. (Altered items in the edited normal checklist are indicated in the AOM of Airline A by shading, so that graphically, the normal checklist, the abnormal checklist, and the relationship between them can be seen simultaneously.) Figures I-7 and I-8 show the formal structure of such dependencies.

```
ENGINE OVERHEAT                    SMOKE SOURCE ID
     |                        .          |            .
ENGINE FIRE           A/C SMOKE     ELEC. SMOKE     SMOKE CONTROL:
     |                               OR FIRE        CABIN OR COCKPIT
One Engine Inop
```

Figure I-7: Dependencies among Emergency Checklists

```
Engine Shutdown                      X Hydraulic Sys Failed
     |                                         |
One Engine Inop               Landing: X Hydraulic Sys Failed
     |
Landing: One Engine Inop
```

Note: X may be: "A"; "B"; or "A" and "B". Then
Landing: "A" and "B" Hydraulic Sys Failed is just
the Normal Landing Checklist.

Figure I-8: Dependencies among Abnormal Checklists

It may be worth noting that some checklists explicitly indicate that they are never initial checklists, that is, they can only be executed following execution of some prior checklist. For example, the SMOKE CONTROL: CABIN OR COCKPIT emergency checklist begins with

Use Smoke Source Identification checklist first

I.2.4 Summary and Discussion

Having considered the structure of checklists as plans, and added certain features to cover plans that are executed in real tine, we may now ask whether checklists form a subcategory of plans that is distinguishable by some particular structural properties. The answer is that essentially, they do not. Checklists exhibit all the structural diversity of plans; there is no plan structure found in the analysis of small group planning that is not also found in checklists. Checklists do tend to have a preponderance of certain characteristic structures, especially SEQ, COND/ACTION and CH/R nodes; EXOR nodes are rather rare (there are only two in our data). However, this is a statistical, not a structural property.

The most important new node types, CH/R and COND/ACTION nodes, do not represent additions to the basic theory of planning, since they have been analyzed in terms of the basic plan node types, and thus they serve as abbreviations for complex interactional structures that particularly common in aviation procedures. Another point of agreement between checklists and plans is the possible embedding of explanations within them. We have several examples of this in the Expanded Checklist Section of the AOM. Perhaps the major structural difference of the checklists in our data from plans in other bodies of data that we have studied (such as political planning) is the frequent omission of an explicit statement of a GOAL. Another difference is that some checklists have explicit dependencies upon others. This phenomenon is implicit in our theory of planning, but we had not found any examples, since our previous small group planning data involved spontaneously generated plans, whereas checklists are pre-set plans, to which the crew may refer as needed. Thus we conclude that checklists can be formally described as plans, and that such a description can be used as a framework to guide the empirical investigation of checklist performance.

References

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.

Goguen, Joseph and Linde, Charlotte. *Linguistic Methodology for the Analysis of Aviation Accidents*. Technical Report, Structural Semantics, 1983. NASA Contractor Report 3741, Ames Research Center.

Goguen, Joseph, Linde, Charlotte and Murphy, Miles. Crew Communication as a Factor in Aviation Accidents. In Hartzell, E. James, and Hart, Sandra (Eds.), *Papers from the 20th Annual Conference on Manual Control*, NASA Ames Research Center, 1984.

Goguen, Joseph, Weiner, James and Linde, Charlotte. Reasoning and Natural Explanation. *International Journal of Man-Machine Studies*, 1983, *19*, 521-559.

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

Linde, Charlotte. *The Linguistic Encoding of Spatial Information*. PhD thesis, Columbia University, 1974.

Linde, Charlotte and Goguen, Joseph. The Structure of Planning Discourse. *Journal of Social and Biological Structures*, 1978, *1*, 219-251.

Linde, Charlotte, and Goguen, Joseph. *Linguistic Measures for Evaluating Flight Simulations*. Technical Report, Structural Semantics, 1987. NASA Contractor Report, Ames Research Center.

Linde, Charlotte and Labov, William. Spatial Networks as a Site for the Study of Language and Thought. *Language*, 1975, *51*, 924-939.

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.

Murphy, M., Randle, R., Tanner, T., Frankel, R., Goguen, J., and Linde, C. A full mission simulator study of aircrew performance: The measurement of crew coordination factors and their relation to flight task performance. In Hartzell, E. James, and Hart, Sandra (Eds.), *Papers from the 20th Annual Conference on Manual Control*, NASA Ames Research Center, 1984.

Murphy, Miles, and Awe, Cynthia. Aircrew Coordination and Decisionmaking: Peer Ratings of Video Tapes Made During a Full Mission Simulation. In *Proceedings of the 21st Annual Conference on Manual Control*, NASA, 1985.

National Transportation Safety Board. *Aircraft Accident Report -- United Airlines, Inc., McDonnell-Douglas DC-8-61, N8082U, Portland*. Technical Report, National Transportation Safety Board, 1979.

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.

Sachs, Harvey. An Analysis of the Course of a Joke's Telling in Conversation. In Bauman, R. and Sherzer, J. (Ed.), *Explorations in the Ethnography of Speaking,* Cambridge University Press, 1974.

1. Report No. NASA CR-177460	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Checklist interruption and resumption: A Linguistic study		5. Report Date June 1987
		6. Performing Organization Code
7. Author(s) Charlotte Linde & Joseph Goguen		8. Performing Organization Report No.
9. Performing Organization Name and Address Structural Semantics P. O. Box 707 Palo Alto, CA		10. Work Unit No. 505-67-51
		11. Contract or Grant No. NAS2-11052
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D. C. 20546		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code

15. Supplementary Notes

Point of contact: Sandra G. Hart, Ames Research Center, M/S 239-3 Moffett Field, CA 94035, (415) 694 6072 or FTS 464 6072

16. Abstract

This study forms part of a project investigating the relationships among the formal structure of aviation procedures, the ways in which the crew members are taught to execute them, and the ways in which they are actually performed in flight. Specifically this report examines the interactions between the performance of checklists and interruptions, considering both interruptions by radio communications and by other crew members. The data consist of 14 crews' performance of a full mission simulation of a commercial Boeing 707 flight. The results show that good crews have a higher ratio of checklist speech acts to all speech acts within the span of the performance of the checklist. Further, it is not number of interruptions but length of interruptions which is associated with crew quality. Use of explicit holds is also associated with crew quality.

17. Key Words (Suggested by Author(s)) Communications Checklists Cockpit procedures Transport operations	18. Distribution Statement Unclassified – Unlimited Subject category – 03		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 31	22. Price*