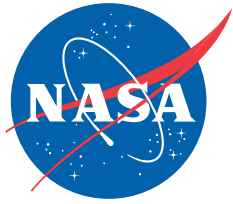# What Happened, and Why: Toward an Understanding of Human Error Based on Automated Analyses of Incident Reports—Vol. II

Thomas A. Ferryman and Christian Posse
Battelle/Pacific Northwest National Laboratory
Richland, Washington

Loren J. Rosenthal
Battelle Memorial Institute
Mountain View, California

Ashok N. Srivastava, Ph.D. and Irving C. Statler, Ph.D.
Ames Research Center
Moffett Field, California

December 2006

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the Lead Center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA Access Help Desk at (301) 621-0134

- Telephone the NASA Access Help Desk at (301) 621-0390

- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320
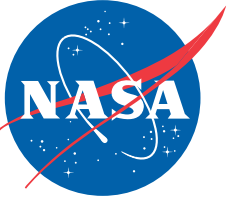
What Happened, and Why: Toward an Understanding of Human Error Based on Automated Analyses of Incident Reports—Vol. II

Thomas A. Ferryman and Christian Posse
Battelle/Pacific Northwest National Laboratory
Richland, Washington

Loren J. Rosenthal
Battelle Memorial Institute
Mountain View, California

Ashok N. Srivastava, Ph.D. and Irving C. Statler, Ph.D.
Ames Research Center
Moffett Field, California

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

December 2006

Available from:

Table of Contents

List of Figures

1 BACKGROUND

This is a report on a continuing study of automated analyses of experiential textual reports to gain insight into the causal factors of human errors in aviation operations. The intent of this research is to better understand the quantitative and qualitative attributes of an aviation incident, and to identify the respective contributions of their interaction to incident occurrence.

NASA's Aviation Safety Program (AvSP), initiated in 2000 and ended in 2005, developed technologies that could, if implemented, reduce the aircraft accident rate by a factor of five within ten years and by a factor of ten within twenty years. One of the AvSP projects, the Aviation System Monitoring and Modeling (ASMM) project, addressed the need to provide decision makers with the tools for identifying and correcting the predisposing conditions that could lead to accidents. Much depends on being able to determine how complex systems have failed and how human behavior influenced such outcome failures.

In the approach to the study reported here, the focus is on uncovering and understanding those precursor conditions that elevate the probability of downstream human errors and that, in turn, may contribute to aviation safety incidents or accidents. A goal is to assist the aviation safety analyst to understand how these systemic features shape human behavior so as to know how to improve the performance of the system. Information extracted from quantitative data sources helps the domain expert understand the objective aspects of *what* happened, and from data sources such as incident reports to understand the subjective aspects of *why* the incident occurred.

The experiential account of the incident reporter is the best available source of information about why an incident happened. Volume I (Maille et al. 2006) of this report describes the exploration of a first-generation process for searching large databases of aviation accident or incident textual reports, and analyzing them for *what* happened as well as for *why* (the causal factors of human behavior). The studies reported here in Volume II were based on the theoretical foundation and experiments described in Volume I. While the reader is encouraged to review that publication, the discussions and results of Volume I are sufficiently summarized to allow this report to stand on its own.

Mining the databases of experiential accounts of incidents poses several challenges. The current process that relies heavily on humans reading the reports is labor-intensive and requires high-priced domain expertise. Further, analyses of incident reports require not only experts with knowledge of aviation operations to understand what happened, but often also experts in human factors to explain why the reported event happened. The process of extracting information from large databases of incident reports requires new automated analytical capabilities to help the experts mine these rich and complex sources for insight into the causal, contributing, and aggravating factors of an event.

A pragmatic approach to these challenges needs to start with a model that captures the underlying structure of an incident report with which to guide the automated analyses. Volume I described such a conceptual model and an approach to using it in automated analyses of textual data sources.

Reporters of incidents usually describe situations they have encountered during flight operations having safety implications as stories about what happened, the involvement and behavior of people

involved, and important features that can help the analyst understand why these events occurred. All the information in the report can be associated with a description of a state of the reporter's world, or with the characterization of an event that contributes to a transition from one state to the next. Operational personnel consider that an incident occurs if, for a period of time, the state of their world is considered as unsafe, compromised, or anomalous. The development of the 'Incident Model,' based on a sequence of states and transitions during the evolution of an incident, as a generic model of any anecdotal report of an incident in aviation operations is described in Volume I. A set of parameters that describe each state and a set of parameters that describe each transition are the descriptors of each step in the chronology of an incident. In Volume I, a simplified subset of the Incident Model called the 'Scenario' (fig. 1) was introduced and defined as:

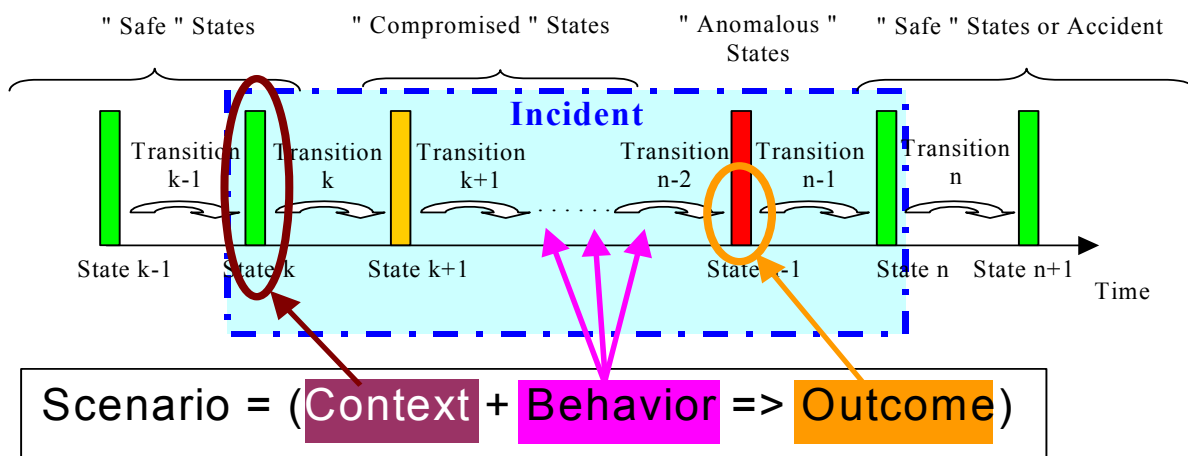**SCENARIO = {CONTEXT + BEHAVIOR ➔ OUTCOME}**

Figure 1. The relation of scenario to the incident model.

The Scenario is a simplified description of an incident in which:

- The 'Context' fits the exact description of the situation in the last safe state.[1]

- The 'Behavior' contains all the problematic events that occur during the transition from the last safe state to the anomalous state.

- The 'Outcome' describes why the resultant state is considered as anomalous. When the Outcome is an anomalous or unsafe state, the confluence of the factors of the last safe state and the factors influencing the Behavior of the transition is identified as a precursor.

This representation is a significant simplification of reality, but it offers a pragmatic approach for guiding the clustering of similar incidents in this first-generation automated system.

The Scenario is defined by the subsets of parameters that describe the Context, the Behavior, and the Outcome of the incident model that are specific to the "story" of a particular incident report. A typical aviation safety incident report includes a set of attributes (often in fixed fields of the

---

[1] The term "state" as used in this report means the state of the <u>entire</u> system relevant to the reporter's world.

2

reporting form) and the values of those attributes (entered by the reporter) plus the reporter's narrative of the incident. The fixed fields of the forms contain a good deal of structured objective information relating to the Context and Outcome of a reported safety event (i.e., *what* happened), but very little structured information relating to the Behaviors of the people and automation that contributed to the events (i.e., *why* it happened).

An experiment described in Volume I confirmed that automated tools could reliably cluster incident reports and provide an adequate description of the Context and the Outcome of a Scenario on the basis of the structured objective information. A further experiment described in Volume I demonstrated that there are statistically significant relationships in typical incident reports between the objective parameters describing Contextual Factors and the objective parameters describing anomalous Outcomes. More importantly, a review of the results by domain experts confirmed that these were also operationally significant relationships.

All of this was preparatory to pursuing the objective of automatically defining the *why*. The experiments described in Volume I demonstrated that automated techniques could adequately describe *what* happened, and that the *what* could be used as a basis of similarity for clustering incident reports. There remained the challenge to identify the causal factors of the Behavior that produced the transition from the last safe state to the unwanted Outcome—the *why*—in the Scenario model of figure 1 for each such cluster. This causal information must be extracted from the experiential narrative of the reporter of the incident. Consequently, in the development of the second stage of filtering to be described in this report, the understanding of *why* the incident happened relies on exploitation of the free text and the extraction of subjective parameters.[2]

The questions that this challenge posed were

Is there a conceptual paradigm that will provide a reliable explanation of the discriminating factors that constitute the Behavior entailed in incidents from large aviation databases?

Can this description be used to "tune" automated analyses that will extract useful information about Behavior in a set of similar incidents?

2   INITIAL METHODS AND APPROACHES

It was desirable to minimize the domain for analyzing why a Scenario happened so as to maximize the possibility of success with this first-generation process. This was achieved in two ways: first, domain knowledge was used to generate rules to maximize the information extracted automatically from the objective parameters about what happened, and second, a simplified model of Behavior was used as a basis for guiding the automated analysis to understand why.

---

[2] The experiential narratives of incident reports such as those in the ASRS are rich sources of information regarding the behaviors of pilots, air traffic controllers, other persons, and automated agents during the course of the reported events. However, the unstructured nature of these data creates an analytical challenge.

The simplified model of Behavior used to "guide" the automated system in a plausible direction relied on knowledge of human behavior. A perspective emerging from scientific literature is that the occasional errors made by pilots and other skilled experts occur in a somewhat random fashion, so that human-factors scientists speak of factors influencing the probability of errors rather than *causing* errors. Also, an accident or an incident is, most often, the consequence of a complex interplay of multiple factors, combining in ways driven in large degree by chance. Multiple factors, not all of which can be determined and measured, interact to produce a human error in a given instance. There is an implicit concept here of the strength or dominance of causal factors, so that it becomes important to identify the strongest causal factors while admitting that weaker factors and interactions may well play a role. It is in this sense that the term causal factors is used throughout this report.

2.1 Situation Awareness

At this exploratory stage of the research, the omission of many plausible (albeit rare) factors that are known to influence human behavior (such as physiological and psychomotor factors) was acceptable if an ability to aid in the identification of a few important common ones could be demonstrated. Furthermore, from the perspective of this study, references to the causal factors of human error mean the systemic features (both latent and proximate) that cause one or more of the human operators of the system to be unable to predict correctly the consequences of his/her/their action(s). This is associated with an inadequate awareness of the state of his/her/their world. So, in this initial attempt to cope with this complex problem, it was proposed that the behavioral failure responsible for transitioning from the Context of the last safe state to a compromised or anomalous state of the Outcome in aviation Scenarios is always associated with a loss of "Situation Awareness (SA)." This approach is based on the substantial body of literature reporting on a variety of perspectives of SA and its role in human behavior. (See, for example, Endsley 1988, Gronlund et al. 1998, Durso and Gronlund 1999, Shively et al. 1997, and Sohn and Doane 2000, Hartel et al. 1991, Endsley 1995a, Jones & Endsley 1996, and Gibson et al. 1997.)

Further, it was proposed that the discriminating and constructive factors of loss of SA are failures to

> **Detect.** Detection is the act of discovering, discerning, or capturing attention as this is related to the existence, presence, or fact of an event.
>
> **Recognize.** Recognition is the act of relating a detected event to a class or type of event that has been perceived before.
>
> **Interpret.** Interpretation is the act of relating a specific event type to a network of actual and possible events of various other types.
>
> **Comprehend.** Comprehension is the act of perceiving the significance of an event.
>
> **Predict.** Prediction is the act of forecasting what will happen in the near future.

(See, for example, Endsley 2000a, Endsley 2000b, and Shively, R. J., et al. 1997) Detection, Recognition, Interpretation, Comprehension, and Prediction (DRICP) were seen as useful concepts for "tuning" the automated analyses of this second stage of filtering.

Endsley had earlier proposed a three-level taxonomic structure for classifying and describing errors in SA that correspond to the five components of SA proposed above for use in this study (Endsley 1994, 1995a, & 1995b). Detection and Recognition are necessary for Endsley's Level 1 SA (Endsley 1996, 2000a, & 2000b). Interpretation and Comprehension are necessary for Level 2 SA (Endsley 1996, 2000a, & 2000b). Prediction is the primary roll in her highest level (Level 3) of SA. The relationship between these two frameworks of Jones and Endsley and DRICP is presented in table A.

It was suggested that, if it was found to be impossible to discriminate to the five levels of detail of DRICP, the analyses would be adapted to Endsley's three-level taxonomy of perception, comprehension, and projection. In any case, Endsley's lower-level descriptions of each of her three levels would help to develop representative concepts, words, or phrases that a reporter of an incident might use to indicate the components of loss of SA. The framework shown in table A helped structure the creation of the concepts of behavior that the automated data mining would pursue.

TABLE A. RELATIONSHIPS BETWEEN DRICP AND JONES & ENDSLEY'S
LEVELS OF LOSS OF SA

		Scenario Model Behavior Descriptors				
		Detection	Recognition	Interpretation	Comprehension	Prediction
Jones and Endsley	**Level 1: Fail to perceive information or misperception of information**	++	++			
	Data not available	++	++			
	Hard to discriminate or detect data	++	++			
	Failure to monitor or observe data	++	++			
	Misperception of data			++	++	
	Memory loss			++	++	
	Level 2: Improper integration or comprehension of information			++	++	
	Lack of or incomplete mental model			++	++	
	Use of incorrect mental model			++	++	
	Over-reliance on default values			++	++	
	Other			?	?	
	Level 3: Incorrect projection of future actions of the system					++
	Lack of or incomplete mental model					++
	Over-projection of current trends					++
	Other					++

However, before the development of this second stage of filtering was undertaken, a small experiment was conducted to see whether aviation-safety subject-matter experts would agree with the losses of SA that were identified automatically.

2.2 The Workshop

The procedure for the analysis, described in Volume I, was based on the expectation of being able to identify words and phrases with which to discriminate and reliably find the specific aspect(s) of loss of SA (at least to the granularity of Jones and Endsley's three levels) that correlated statistically with the Context and the anomalous Outcome of each safety-related Scenario. The proposal was to seek relationships between those aspects of loss of SA for a specific Outcome with factors of the Context that had already been found to exist in each Scenario to identify the factors associated with *why* the event occurred.

Before proceeding with these analyses, a workshop was convened of representatives from human factors, aviation operations, ASRS, computational linguistics, human performance modeling, and data mining to develop representative concepts, words, or phrases that a reporter of an incident might use to indicate the components of SA. (The workshop participants are identified in Appendix A.) In multiple sessions, small groups read the same subset of ASRS reports to relate them to each of Jones and Endsley's three levels of SA shown in table A. The participants were divided among the sessions, with intent to assess the strength of agreement on the identification of the level(s) of the loss of SA portrayed in each report and to capture illustrative phrases for each level that could be used to guide automated mining of the unstructured text.

The objective was not realized. However, the process was enlightening. The groups had difficulty making sharp distinctions among the concepts of loss of SA even at the granularity of Jones and Endsley's three levels of SA. Therefore, they were unable to come to agreement on illustrative phrases in the ASRS reports that were indicative of each level. However, retrospective analysis of the discussions during these sessions revealed a fundamental aspect of the nature of these reports.

During these discussions, the workshop team always seemed to go back to contextual factors as they searched for clues to Jones and Endsley's categories of loss of SA in the sample set of ASRS reports. Invariably, the workshop team found that the reporters spoke of the environment surrounding the incident. In fact, it seemed that reporters found it hard to refer to categories of loss of SA without linking these to various contextual factors, and further analysis from a human-factors perspective pointed to good psychological reasons for this. Since an individual (in this case, the reporter) is a constant from his or her own point of view, it is difficult for that individual to attribute the cause of an unusual event (incident) to his or her own behavior. The individual has had the experience of many, many other situations where he or she was also present, and which were normal, not anomalous. It is natural to attribute an unusual event to something unusual in the environment, hence to a contextual factor rather than a (personal) human factor such as loss of SA. This could explain the rather high incidence of statements about distractions, interruptions, and high workload, and the very low incidence of reference to any psychological factors associated with loss of SA in the sample subset of ASRS reports used for the workshop.

A subsequent review of a larger number of ASRS reports with this new perspective reinforced this interpretation of the results of the workshop. It was found that, frequently, reporters provide "advice" in their reports to the ASRS, which implies the reporter's perception of causal factors and potential interventions. The advice is often associated with the words 'should' or 'ought,' as in the following few exemplary excerpts taken from ASRS reports:

> *I FEEL THAT THESE DEER SHOULD BE EITHER RELOCATED OR INSTALL A FENCE AROUND THE ARPT.*
>
> *THIS APCH SHOULD BE OTS IF SUCH A RESTR IS REQUIRED, AND WINDS ALOFT ARE FROM THE N.*
>
> *I FEEL THAT THERE SHOULD HAVE BEEN A HOLD SHORT LINE BEFORE RWY 1.*
>
> *A SE TXWY SHOULD BE INSTALLED.*
>
> *THE INCORRECT PAPERWORK SHOULD HAVE BEEN DISCOVERED WELL BEFORE DEPARTING, BUT PRESSURE TO KEEP THINGS MOVING PUT US IN A 'GO' MODE.*
>
> *I THINK THE SIGN SHOULD BE TO THE L (S) OF TXWY G AND PARALLEL TO RWY 35L/17R.*
>
> *I FREQUENTLY HAVE TO ASK FOR WIND INFO FOR TKOFS AND LNDGS WHICH SHOULD ALWAYS BE GIVEN.*
>
> *BECAUSE OF TIME PRESSURE TO GET FLT OUT, THE REQUIREMENT TO CHK OIL QUANTITY WAS OVERLOOKED AND NO ONE FROM MAINT SHOWED UP TO CHK. AN ENTRY SHOULD HAVE BEEN MADE AND SIGNED OFF IN LOGBOOK BUT WAS NOT.*
>
> *PERHAPS A TAXI INSTRUCTION OF A DIFFERENT TYPE, IN REGARDS TO THE TXWY MERGER, OUGHT TO BE GIVEN SOME CONSIDERATION.*
>
> *IT OUGHT TO HAVE A WARNING HORN AT 300 FT FROM ALT.*
>
> *TWR OUGHT NOT TO SEQUENCE RELATIVELY FAST ACFT SUCH AS MY M20R BEHIND TRAINERS.*
>
> *WE OUGHT TO GO TO SCHOOL ON THIS TO PREVENT PROBS THAT COULD RESULT AT NIGHT OR IN THE WX.*

Notice that in these examples and many others (though not all others), the causal factors and recommended interventions are linked to contextual factors rather than behavioral factors. (Behavioral examples can be found in association with "I should," "I should have," "we should," or "we should have": *"WE SHOULD HAVE TAKEN HIM OUT OF PLT'S SEAT EARLIER."*)

It was decided that this argument was valid, and the fact that contextual factors are going to be closely connected in the experiential report with behavior (i.e., loss of SA) should be accepted. Moreover, this is consistent with the original objective (as stated in Volume I (Maille et al. 2006)) and, in particular, at Step 3 of figure 9 of that report); namely, to identify the objective contextual factors related to failure modes of SA. The realizations gained from the workshop caused a change in approach to focus on fully capturing the context as it is reported as influencing the reporter's behavior rather than on the cognitive failure modes of SA.

3   A MODEL FOR BEHAVIOR

3.1   Contextual Shaping Factors

In Volume I, what was called the "full and complete" set of parameters to describe any pilot's aviation safety incident report was identified. Of these, those parameters that were objective, categorical, and measurable, and were claimed to be adequate to describe *what* happened were identified. These objective parameters that could adequately describe the Context and the Outcome of a Scenario are, for the most part, contained within the fixed fields of a typical incident report. An understanding of *why* the incident occurred entailed the remaining parameters (that were arbitrarily labeled "subjective"), which would have to be extracted from the free narrative part of the report. It was postulated in Volume I that these parameters could be related to the failure modes of SA and that they would correlate with the existing contextual factors to explain the *why* of the failures. However, the experience of the workshop was convincing evidence that the reporters were already describing these contextual factors and a search for the cognitive failure modes per se could be bypassed. These contextual factors are the ones that the experiential report says influenced the reporter's behavior and, to avoid confusion with the objective parameters of the Context (from the fixed fields), these factors derived from the narrative are referred to as "Shaping Factors." This term has been used in a number of reports, such as Sasou, K., & Reason, J. 1999.[3]

To test the merits of this approach, the "full and complete" set of parameters was reviewed to identify a set of Shaping Factors that could be used to guide the automated analysis in the next experiment. In Volume I, a codification form was described that had been designed to update the codification of ASRS reports called the X-Form. After several years of experience entering reports into the ASRS database and conducting retrospective searches, the X-Form was developed by experienced ASRS analysts in collaboration with human-factors research and aviation operational personnel to improve the descriptions of factors that influence human performance in aviation operations. The X-Form was selected for a subset of parameters to use as the "Shaping Factors" in an initial experiment to evaluate a capability to identify these automatically from the free narrative of a set of incident reports. Table B is the set of fourteen Shaping Factors with brief definitions and exemplary expressions taken from incident reports that were used to evaluate an ability to identify factors such as these automatically from the free narratives of incident reports (Posse et al. 2004).

---

[3] In the literature (e.g., Swain 1983), the "Context" has been called the external Performance Shaping Factors (PSF), and what has been labeled in this report as the "Shaping Factors" have been called the internal PSF.

TABLE B. SHAPING FACTORS

Factor	Definition	Example
Attitude	Any indication of unprofessional or antagonistic attitude by a controller or flight crew member, e.g., complacency or 'get-homeitis' (in a hurry to get home).	*"I believe a contributing factor was complacency flying a very familiar approach, also it was our last leg get-thereitis."*
Communication Environment	Interferences with communications in the cockpit such as noise, auditory interference, radio frequency congestion, or language barrier.	*"We were unable to hear because traffic alert and collision avoidance systems were very loud."*
Duty Cycle	A strong indication of an unusual working period e.g., a long day, flying very late at night, exceeding duty time regulations, having short and inadequate rest periods.	*"Flight had previously been delayed and we had minimum rest period coming up, less than 9 hours."*
Familiarity	Any indication of a lack of factual knowledge, such as new to or unfamiliar with company, airport, or aircraft.	*"Both pilots were unfamiliar with the area."*
Illusion	Illusions include bright lights that cause something to blend in, black hole, white out, or sloping terrain.	*"I was flying and was experiencing a black hole effect."*
Physical Environment	Unusual physical conditions that could impair flying or make things difficult, such as unusually hot or cold temperatures inside the cockpit, cluttered workspace, visual interference, bad weather, or turbulence.	*"This occurred because of the intense glare of the sun."*
Physical Factors	Pilot ailment that could impair flying or make things more difficult, such as being tired, fatigued, drugged, incapacitated, influenced by alcohol, suffering from vertigo, illness, dizziness, hypoxia, nausea, loss of sight, or loss of hearing.	*"I allowed fatigue and stress to cloud my judgment."*
Preoccupation	A preoccupation, distraction, or division of attention that creates a deficit in performance, such as being preoccupied, busy (doing something else), or distracted.	*"My attention was divided inappropriately."*
Pressure	Psychological pressure, such as feeling intimidated, pressured, pressed for time, or being low on fuel.	*"I felt rushed to complete the checklist in time."*
Proficiency	A general deficit in capabilities, such as inexperience, lack of training, not qualified, not current, or lack of proficiency.	*"The biggest safety factor here is the lack of adequate training in the newer autopilot system."*
Resource Deficiency	Absence, insufficient number, or poor quality of a resource, such as overworked or unavailable controller, insufficient or out-of-date chart, equipment malfunction, inoperative, deferred, or missing equipment.	*"Later I learned the minimum equipment list was wrong."*
Taskload	Indicators of a heavy workload or many tasks at once, such as short-handed crew.	*"Due to high workload, I forgot to switch to tower."*
Unexpected	Something sudden and surprising that is not expected.	*"Had we known of him prior to takeoff we would have made adjustments."*
Other	Anything else that could be a shaper, such as shift change, passenger discomfort, or disorientation.	*"This happened during shift change." "Ill passenger on board."*

These Shaping Factors are not mutually exclusive and some may be more difficult to discriminate than others. It is not suggested that these are the "full and complete" set of contextual factors that can influence human performance. At this stage, they were simply selected to test the concept for automated data mining on a subset of typical incident reports.

A series of brainstorming sessions involving aviation safety experts, human factors experts, English specialists, and data analysts from non-aviation backgrounds generated a large number of seed *keywords* (i.e., words that consistently and reliably indicate the presence of a particular shaping function*), simple expressions* (i.e., strings that systematically indicate the presence of particular shaping functions), and *template expressions* (i.e., complex expressions associated with a particular shaping function that may occur in many different variants in the narrative) for each shaping function. (Examples of each for the shaping factor of "Pilot Fatigue" are presented later in this report.)

The next step was to test the ability to use these definitions and exemplary phrases to discriminate the Shaping Factors from the free narratives of incident reports pertaining to each Scenario. However, it is not sufficient to find such expressions in the narratives, because the context in which they are used must be taken into account. Most importantly, this means checking for the presence of negative modifiers in close proximity to a given expression that changes the meaning of the expression. For example, the expressions like "I have not had an exhausting day," "If I had felt I was tired I would...," "…fatigue was not an issue," indicate that fatigue was, in fact, not a significant factor.

Consequently, it was recognized that the traditional keyword-based and bag-of-words approaches that worked so well on the fixed fields for analyzing the Context and the Outcome of each Scenario would be of limited value for analyzing the Shaping Factors of Behavior. More complex pattern extraction methods were required, and an understanding of the nature of the narratives to be analyzed was needed in order to identify the most appropriate available methods for this task.

3.2   The Narratives of Incident Reports

The narratives of incident reports are, generally, highly informal and replete with the acronyms, idioms, multiple spellings, misspellings, and ambiguous abbreviations that are specific to each domain and even to tasks. The informal reporting setting tends to produce poor grammar, spelling variants, very domain- or task-specific expressions, as well as jargon. (This statement is strongly supported in the study reported by Van Delden and Gomez (Feb., 2004).) Stream of consciousness permeates the reports, which may exhibit feelings of anger, guilt, or defense. Styles and even languages vary dramatically, from telegraphic to very detailed, and depend on whether the narrator is a pilot, an air-traffic controller, a flight attendant, a mechanic, or ground personnel. The following ASRS report is representative of incident reports and illustrates some of these characteristics:

> *"I RETURNED FROM A LCL TRNING FLT. PER TWR, WE LANDED ON RWY 4R AT MDW. SHORTLY AFTER OUR T/D, THE ACFT EXPERIENCED SEVERE NOSE SHIMMY VIBRATION. WE SLOWED DOWN AND TURNED LEFT AS PER TWR INSTRUCTIONS OFF OF RWY 4R. WE SWITCHED TO GND AND THEY TOLD US*

TO TAXI TO THE N RAMP. WE PROCEEDED TO DO SO. FIVE SECS LATER, GND ASKED US HOW DO WE HEAR. WE TOLD THEM LOUD AND CLEAR. AFTER 15 SECS GND TOLD US TO CALL THE TWR. WE DID AND TWR TOLD US WE XED AN ACTIVE RWY (4L) W/O THEIR PERMISSION. TWR SAID AS WE WERE LNDG HE TOLD US TO TURN LEFT AND HOLD SHORT OF RWY 4L AND REMAIN WITH HIM. WE OBVIOUSLY DID NOT HEAR HIM DUE TO THE EXTREME NOISE CAUSED BY THE NOSE SHIMMY. GND SHOULD NOT HAVE TOLD US TO TAXI TO THE N RAMP. TWR TOLD US ON THE PHONE THAT THERE WAS LNDG TFC ROLLING ON RWY 4L AT THE TIME WE XED RWY 4L...."

Consequently, the first step toward automated analyses of such narratives entailed language normalization and knowledge infusion from aviation-domain experts. Normalization produces readable and correct English text, while knowledge infusion elicits the peculiar expressions used by personnel involved in aviation operations. A way was needed to standardize the language of the culture of flight operations before automated analysis of this sort of text could be considered.

3.3 Textual Preparation

3.3.1 PLADS

PLADS is a tool developed by Battelle's Pacific Northwest Division for this project to standardize the language of unstructured text so as to facilitate its reliable automated analysis. PLADS uses a combination of standard English and domain-specific concepts to improve text-mining effectiveness. PLADS is used as a pre-processor in conjunction with other text analysis tools, whether they entail statistical (i.e., "bag-of-words") or Natural Language Processing (NLP) tools, to facilitate analysis of free text. PLADS is composed of software (Java, Matlab, and Perl) and lexicons. The development of the lexicons when PLADS is adapted to a new domain for the first time requires an expert in PLADS working with an expert in the reporting language of that domain.

PLADS is an acronym of the names of the following five stages of filtering performed on each report prior to its automated analysis:

- **P**hrases identified and concatenated. Identify phrases in the unstructured text by statistical means by identifying 2-, 3-, 4-, 5-word strings that occur more often than one would expect based solely on the individual word frequency. Then concatenate the phrase into what would be identified as a single word to subsequent software: e.g., ClassCAirspace, UnitedStatesOfAmerica.

- **L**eave some words unprocessed.

- **A**ugment some words to make the meaning more useful for computer analysis. Some words may be abbreviations for instruments and/or concepts with make/model/series, or numeric values of selected concepts for example:
  - "B-757-300" might be augmented with the word "airplane."
  - "FL28" ("FL26," "FL30") means "flight level at approximately 28,000 (26,000, 30,000) feet." Augmenting with "FlightLevel" enables subsequent software to identify these 3 (and others) as related to a flight-level concept, leaving the refinements of which specific flight level to finer grain analysis.

- "24L," "24R," "25L," "25R" all relate to runways. Augmenting with the word "runway" enables the software to capture that concept.

- Proper names are often augmented with the more general concept, e.g., "Dallas" augmented with "city."

- Airport abbreviations are often augmented with the word "airport," e.g., "LAX," "ORD," "DFW."

- **D**elete some words to simplify the analysis. These are often called "stop" words. Examples include: "the," "a," "an." Sometimes numbers are dropped out.

- **S**ubstitute some words for others. Often there are many ways to express the same concept. This includes synonyms, abbreviations, jargon, and slang. For example "pilot" might be substituted for these words: "pilots," "co-pilot," "captain," "co-captain," "left seater," "PIC," "Pilot-in-Charge," "plt," and "plts." Standard abbreviations can be checked and full meanings substituted. Spelled-out numbers may be replaced by the numeral.

Following is the example of the ASRS report shown in the previous section after it has been processed through PLADS (which does not do a perfect job as, for example, "TURNING FLIGHT" should be "TRAINING FLIGHT"):

> *I RETURNED FROM A LOCAL TURNING FLIGHT. PER TOWER, WE LANDED ON RUNWAY 4R AT MDW. SHORTLY AFTER OUR TOUCHDOWN, THE AIRCRAFT EXPERIENCED SEVERE NOSE SHIMMY VIBRATION. WE SLOWED DOWN AND TURNED LEFT AS PER TOWER INSTRUCTIONS OFF OF RUNWAY 4R. WE SWITCHED TO GROUND AND THEY TOLD US TO TAXI TO THE NORTH RAMP. WE PROCEEDED TO DO SO. FIVE SECONDS LATER, GROUND ASKED US HOW DO WE HEAR. WE TOLD THEM LOUD AND CLEAR. AFTER 15 SECONDS GROUND TOLD US TO CALL THE TOWER. WE DID AND TOWER TOLD US WE CROSSED AN ACTIVE RUNWAY WITHOUT THEIR PERMISSION. TOWER SAID AS WE WERE LANDING HE TOLD US TO TURN LEFT AND HOLD SHORT OF RUNWAY 4L AND REMAIN WITH HIM. WE OBVIOUSLY DID NOT HEAR HIM DUE TO THE EXTREME NOISE CAUSED BY THE NOSE SHIMMY. GROUND SHOULD NOT HAVE TOLD US TO TAXI TO THE NORTH RAMP. TOWER TOLD US ON THE PHONE THAT THERE WAS LANDING TRAFFIC ROLLING ON RUNWAY 4L AT THE TIME WE CROSSED RUNWAY 4L… .*

3.4   Analysis of Free Narratives

The use of PLADS (even with its imperfections) greatly improves the potential value of any subsequent automated text analysis because it reduces the domain to be analyzed. The previous work to capture the Context and Outcome of incident reports was based on the fixed-field data for which statistical (i.e., "bag of words") tools worked very well. However, something more in Natural Language Processing would be needed to capture the information about "Shaping Factors" that the reporter conveys in the free text.

Figure 2 diagrams the state of the art of Natural Language Processing showing the stages of expression detection with examples of the expressions that would be identified with a particular concept (in this example, the concept of "familiarity").[4] Capabilities for extracting meaning are extremely limited, and even syntactic parsing or event extraction push the state of the art of Natural Language Processing (NLP). Analysis of aviation-incident Scenarios requires some of both.

Example: "Familiarity"

mature

String Matching
"Unfamiliar", "Unacquainted"

Similar Expressions
"This setting was * new to me"
(* = "somewhat", "quite", "entirely"
 but • "not", or "not quite")

NLP Depth

Part-of-Speech

Entity Extraction
"The regulations had * changed * since … ."

Syntactic Parsing
"I had not performed this procedure for years… ."

Event Extraction

Event Co-reference
"I [did something] … we went too high because on this plane the [tool] is more sensitive than on the 737 (suggesting that the pilot is familiar with the B737, but not the plane he/she is flying right now)."

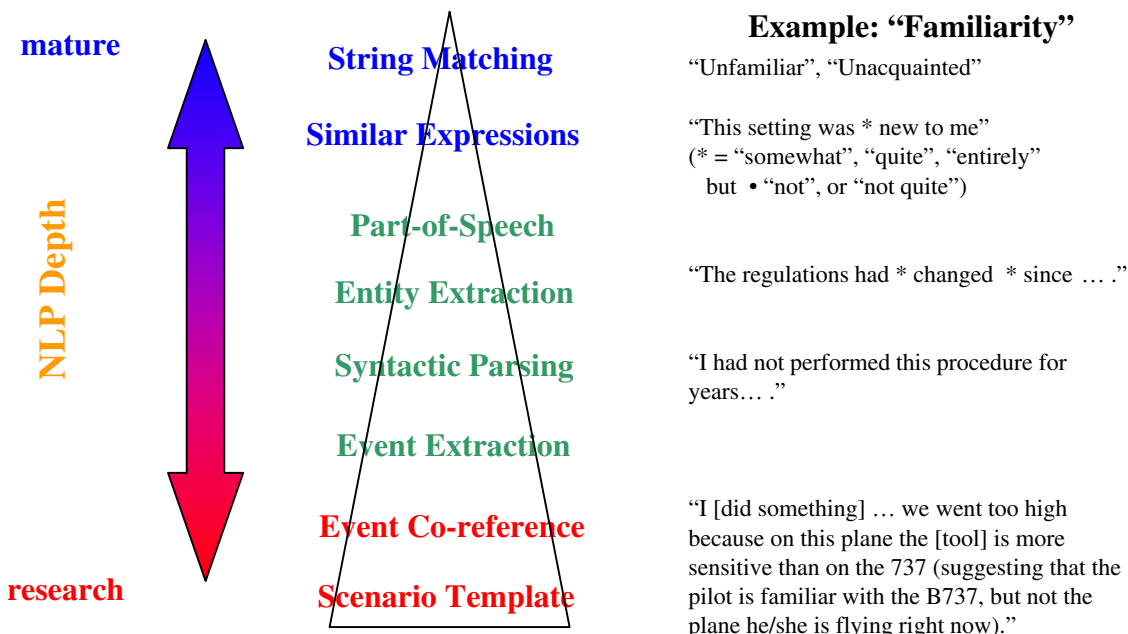**research**

Scenario Template

Figure 2. Spectrum of concept expressions.

After an investigation into available tools, the General Architecture for Text Engineering (GATE) tool was selected to capture NLP of the incident-report narratives after they had been processed through PLADS.

3.4.1   GATE

GATE (General Architecture for Text Engineering), a software tool created by the University of Sheffield (see http://nlp.shef.ac.uk/.), is one of the most widely used human language processing systems in the world. GATE has had great success at TREC (Text REtrieval Conference) series co-sponsored by the National Institute of Standards and Technology, the Information Technology Laboratory's Retrieval Group of the Information Access Division, and the Advanced R&D Activity of the DOD in head-to-head competition with numerous other techniques.

GATE comprises an architecture, framework, and graphical development environment to identify evidence of specific concepts contained in unstructured text. The concepts may be vaguely defined, and phrased in a way as to require subtle insight to identify their existence. (See Manning and Schültze 2000 and Cunningham et al. 2002). GATE provides framework for applying customizable tools for data mining. A GATE gazetteer is a list of expressions compiled into finite state machines

---

[4] Figure 2 was internally generated, but it is consistent with the state of the art presented in all of the current literature on analysis of free text.

that can match text tokens. Gazetteers tend to produce more false positives, but less false negatives because exact matches are determined out of context. Therefore, gazetteers are a good place to start encoding simple shaper expressions to control the false negative rate. GATE includes a pattern specification language called <u>J</u>ava <u>A</u>nnotation <u>P</u>atterns <u>E</u>ngine (JAPE), which describes patterns to match and annotations to be created as a result. This means that the aviation-incident reports can be tagged with respect to the shaping factors, where the tags correspond to the annotations generated by JAPE. JAPE rules can be refined very rapidly without having to rewrite large pieces of code and can easily handle matching exceptions. The tools available within GATE enable specific concepts to be identified using synonyms and Boolean expressions to identify natural language phrases that were not envisioned. JAPE enables the provision of rules to cope with all such possible variations.

While there are numerous other methods of doing NLP, a study of these found NLP via GATE to be more effective at identifying specific concepts in the free text of aviation-incident reports.

3.5   The Experiment

The corpus of data selected for the experiment was a subset of ASRS reports about one or more of a selected set of anomalous Outcomes. The total set of about a hundred anomalies as identified by the ASRS Office was described in Volume I of this report (Maille et al. 2006) and is shown in Appendix B.[5] From among these the following ten ASRS anomalies were selected for this experiment:

- Aircraft Equipment Problem
- Altitude Deviation
- Conflict (Airborne)
- Conflict (Ground)
- Ground Incursion

- Ground Incursion (Clearance)
- In-flight Weather Encounter
- Maintenance Problem
- Non-adherence
- Airspace Violation

This is the same set of Outcomes that was used in the experiment described in Volume I to identify correlations between Outcomes and Contexts.

For the present experiment, a subset of 17,155 ASRS reports was identified that pertained to one or more of these ten anomalous Outcomes to explore why these Scenarios occurred. GATE was used in analyses of the free texts of these 17,155 incident reports to try to identify the Shaping Factors (of the fourteen presented in table B) that prevailed (or did not prevail) in each of the ten anomalous Outcomes.

The processing of ASRS narratives was conducted in several phases: pre-processing for language normalization, tokenization and generation of token sequences, exact expressions recognition, and template expressions extraction. Accordingly, a modular architecture was developed comprised of the components and functions depicted in figure 3 (Posse et al. 2004).

---

[5] ASRS Anomaly codes provide classification of abnormal or irregular events, or events at variance with ATC clearances or instructions.
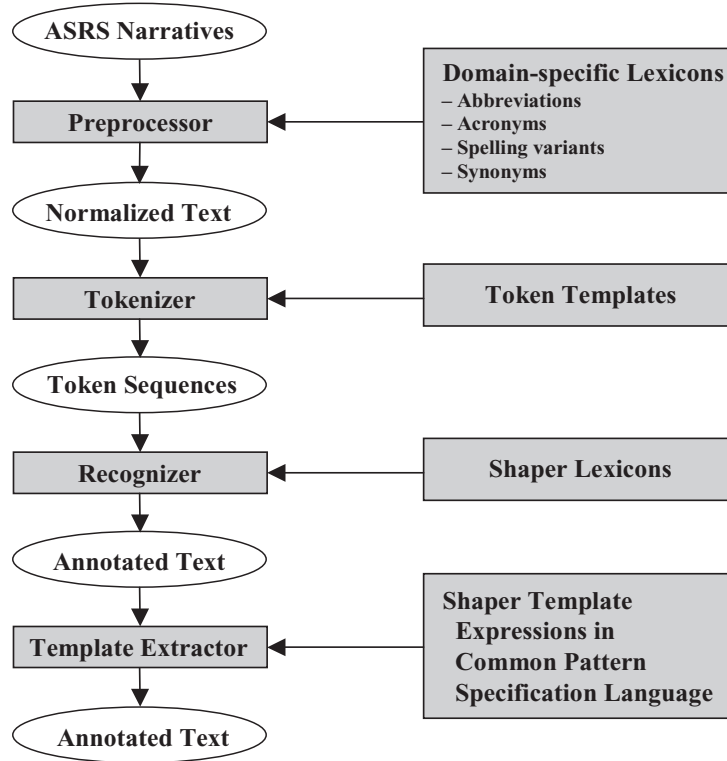
Figure 3. Processing steps for the extraction of shaper references.

Except for the preprocessor phase, all steps in the architecture of figure 3 were implemented within GATE. The preprocessing of each of the 17,155 reports for vocabulary standardization was conducted with PLADS. The lexicons used in this phase were developed by an expert in PLADS working with several experts in the reporting language of aviation-incident reports.

As described previously in this report, aviation-domain and human-factors experts working with NLP experts conceived of and defined the Shaping Factors of table B together with a seed set of natural language expressions for each factor, which were then enriched by using them as inputs for the recognizer and template-extractor phases of figure 3.

The GATE gazetteer was used to mine the incident-report database for each simple expression and to analyze the sentences in which that expression appeared. Whenever this context did not contain any modifier, a simple expression was kept within the gazetteer. Exact matches are much faster than regular expression matches. As was noted earlier, the use of the gazetteer was deemed a good way to start encoding simple expressions to control the false negative rate before proceeding to the development of related JAPE rules for each of the Shaping Factors. Following are examples of synonyms for the notion of pilot fatigue found in the ASRS database using the GATE gazetteer:

- fatigue, tiredness, no sleep, lack of sleep, inadequate sleep, little sleep, short sleep, not sleep, did not sleep, didn't sleep, did not sleep, couldn't sleep, not slept, didn't rest.

- little energy, no energy, no rest, lack of energy, lack of rest, insufficient rest, inadequate rest, stay awake, awake since, on duty since.

- tired, exhausted, dog-tired, fatigued.

- strenuous leg, extended hours, demanding hours, prolonged hours, exhausting day, frantic week.

- tough flight, heavy duty day, it was late, through the night, end of duty time.

Expressions that contain more than one word are more likely to be expressed in different forms; the more terms in an expression, the more variants can be found in the free narratives of the incident reports. For example, in the case of pilot fatigue, "pilot mentions that incident occurred during a certain leg of a several day/leg trip" was used as a seed, and the following variations on that theme were found:

LAST LEG OF A 2 DAY TRIP
LAST LEG OF A TWO DAY TRIP
LAST LEG OF A 3 DAY TRIP
FINAL LEG OF A 3 DAY TRIP
LAST LEG OF A 3-DAY TRIP
FINAL LEG OF A 3-DAY TRIP
LAST LEG OF A LONG 3-DAY TRIP
LAST LEG OF A LONG 5-LEG DAY
LAST LEG OF 3 DAY TRIP
FINAL LEG OF 3 DAY TRIP
LAST LEG OF 3-DAY TRIP
LAST LEG OF A FOUR DAY TRIP
LAST LEG OF A 4 DAY TRIP
FINAL LEG OF A 4 DAY TRIP
LAST LEG OF A 4-DAY TRIP
FINAL LEG OF A 4-DAY TRIP
LAST LEG OF 4 DAY TRIP
LAST LEG OF AN ALL DAY TRIP
LAST LEG OF OUR 3 DAY TRIP
LAST LEG OF A SIX LEG WAY
LAST LEG OF THE DAY
FINAL LEG OF THE DAY
LAST LEG OF THE TRIP
LAST LEG OF A LONG TRIP
FINAL LEG OF THE TRIP
LAST LEG OF A 4 LEG
LAST LEG OF 4 LEGS
LAST LEG OF A LONG DAY
LAST LEG OF A VERY LONG DAY
LAST LEG OF THE NIGHT
LAST FLIGHT OF THE DAY
LAST FLIGHT OF A 2 DAY TRIP
LAST FLIGHT OF THE TRIP
LAST FLIGHT OF THAT DAY
LAST FLIGHT OF THAT DAY
LAST FLIGHT OF A FOUR DAY TRIP

LAST FLIGHT OF A 5 LEG
LAST FLIGHT OF A LONG BYT
UNPROFITABLE DAY
LAST SEGMENT OF A 3-DAY
FINAL SEGMENT OF A LEG WAY
NEXT TO LAST LEG OF TRIP
DAY 2 ON A 4 DAY TRIP
THIRD DAY OF A 3 DAY TRIP
THIRD DAY OF A 4 DAY TRIP
LAST LEG OF DAY #3
3RD LEG OF A 4 LEG
3RD DAY OF A 4-DAY TRIP
THIRD LEG OF A 4 LEG, 1 DAY TRIP
THIRD DAY OF A 3 DAY TRIP
THIRD DAY ON A 4 DAY TRIP
THIRD LEG OF A 2 DAY TRIP
4TH AND LAST LEG OF THE TRIP
4TH LEG OF A LONG DAY
FIFTH LEG OF THE DAY
SIXTH LEG OF A SEVEN LEG DAY
10TH LEG OF AN 11 LEG DAY ON DAY 3
OF A 3 DAY TRIP
11TH LEG OF THE DAY
10 HRS INTO A 12 HR DUTY DAY
NEARING THE END OF A LONG 3-DAY
TRIP
END OF A 2-DAY TRIP
END OF A 3 DAY TRIP
DAY THREE OF A THREE DAY TRIP
DAY 3 OF A THREE DAY TRIP
DAY 4 OF 4 DAYS
DAY 4 OF A 4 DAY TRIP
THIRD DAY OF A THREE DAY TRIP
SECOND DAY OF A THREE DAY TRIP
LAST NIGHT OF A 6 NIGHT TRIP

Trying to capture all of these expressions and the likely variants that can appear in future reports within gazetteers is too laborious. A better approach is to understand the patterns in the set of variant expressions and encode these patterns into templates to be matched. GATE's JAPE was used to describe patterns to match and to create annotations corresponding to the tags with respect to the Shaping Factors assigned to each report. The next step was to write the rules in JAPE computer code to capture the wide set of variations on the natural expressions for each Shaping Factor found in the database and new likely variants to be found in future aviation-incident reports. (See Posse 2004 for more detail on this process.)

Posse (2004) reports on a small experiment to evaluate this approach in which the GATE encoding of the 14 Shaping Factors was run on a random sample of 20 incident reports from commercial passenger flights filed by pilots. Aviation-domain experts considered the results to be plausible and probable. On the basis of the success of that small experiment, the full experiment on the 17,155 incident reports of the ten selected anomalous Outcomes to identify the Shaping Factors associated with each was conducted.

3.5.1   Analyses
PLADS plus GATE was used to analyze the narratives of the 17,155 incident reports to identify the Shaping Factors. The graphic of figure 4 is an example of some of the results of these analyses.
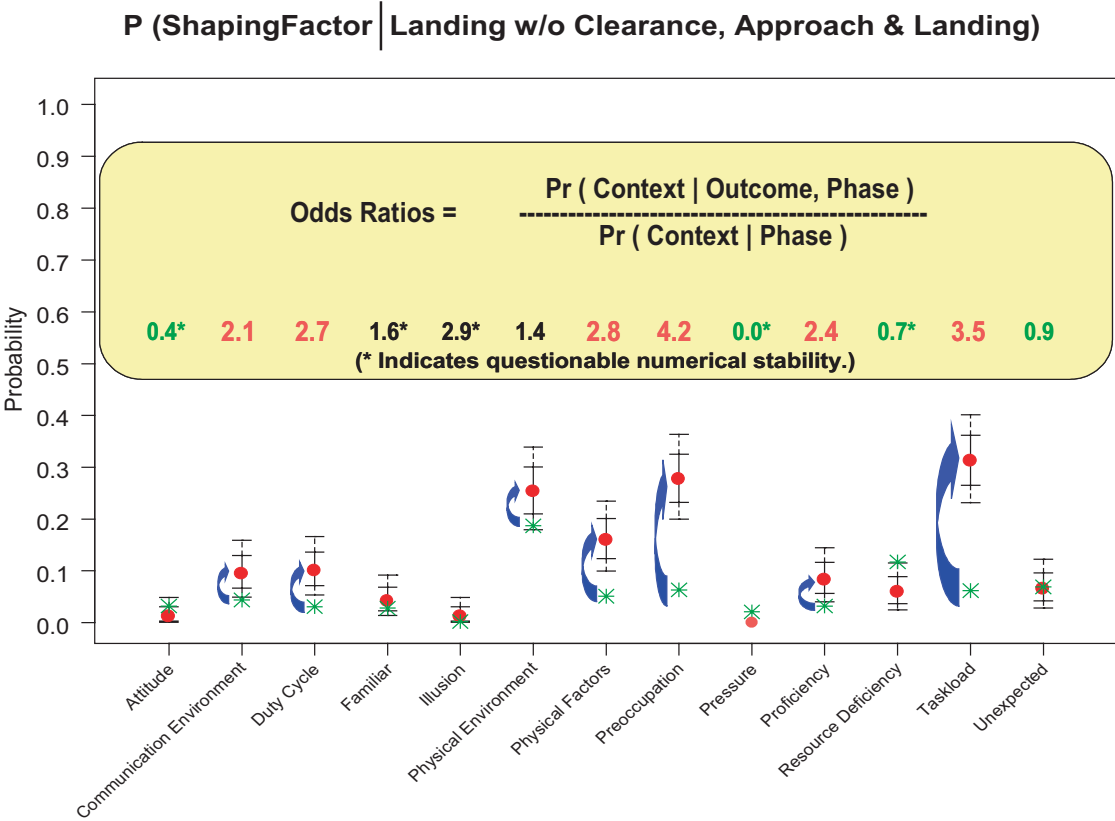

Figure 4. Example of probability of a shaping factor given the outcome and phase of flight.

The example considered in figure 4 is for the Scenario of Landing without a Clearance[6] and is associated with the Approach & Landing phase of flight. The green asterisks ✱ are the probabilities of the occurrences of each of the 13 Shaping Factors (the Other category was omitted in this discussion) during an Approach & Landing phase of flight. The red dots ● are the probabilities of the occurrences of the Shaping Factors during the Scenario of Landing without a Clearance in the Approach & Landing phase of flight. The hash marks on either side of the red dots are the 80% and 99% uncertainty intervals. A Shaping Factor is significantly associated with this Scenario when the ratio of the latter probability (i.e., the red dot) to the former (the green asterisk) is greater than 1.0, as in the cases of those indicated by the blue arrows ↰ for Communication Environment, Duty Cycle, Physical Environment, Physical Factors, Preoccupation, Proficiency, and Task Load. When both probabilities are small, the numerical stability of their ratio is questionable, as in the cases of Attitude, Familiarity, Illusion, Pressure, and Resource Deficiency:

3.5.2   Results

In the following results of this experiment, the significant Shaping Factors are identified and ranked (by odds ratio) for anomalous Outcomes in associated phases of flight for each of the ten ASRS anomalies selected for this study:

Ground Phase

Maintenance Problem is associated with:
  Pressure
     Pr (Pressure | Maintenance Problems, Ground Phase) = 6.4%
     Pr (Pressure | Ground Phase) = 3.0%
     Odds Ratio = 2.1

Ground Incursion is associated with:
  Familiar
     Pr (Familiar | Ground Incursion, Ground Phase) = 8.0%
     Pr (Familiar | Ground Phase) = 3.9%
     Odds Ratio = 2.0
  Communication Environment
     Pr (Communication Environment | Ground Incursion, Ground Phase) = 8.0%
     Pr (Communication Environment | Ground Phase) = 4.2%
     Odds Ratio = 1.9
  Pre-occupation
     Pr (Pre-occupation | Ground Incursion, Ground Phase) = 9.9%
     Pr (Pre-occupation | Ground Phase) = 5.6%
     Odds Ratio = 1.8
  Taskload
     Pr (Taskload | Ground Incursion, Ground Phase) = 7.8%
     Pr (Taskload | Ground Phase) = 4.4%
     Odds Ratio = 1.8

---

[6] Landing without a Clearance is one of the subsets of Scenarios under the anomalous Outcome called Airspace Violation in the set of ASRS anomalies.

Take-off Phase

Conflict (airborne) is associated with:
    Taskload
        Pr (Taskload | Air Conflict, Take-off Phase) = 10.1%
        Pr (Taskload | Take-off Phase) = 5.5%
        Odds Ratio = 1.9
    Communication Environment
        Pr (Communication Environment | Air Conflict, Take-off Phase) = 6.2%
        Pr (Communication Environment | Take-off Phase) = 3.6%
        Odds Ratio = 1.7

Non-adherence is associated with:
    Pre-occupation
        Pr (Pre-occupation | Non-adherence, Take-off Phase) = 8.8%
        Pr (Pre-occupation | Take-off Phase) = 5.5%
        Odds Ratio = 1.6

Ascent Phase

Altitude Deviation is associated with:
    Pre-occupation
        Pr (Pre-occupation | Altitude Deviation, Ascent Phase) = 20.3%
        Pr (Pre-occupation | Ascent Phase) = 7.5%
        Odds Ratio = 2.7
    Proficiency
        Pr (Proficiency | Altitude Deviation, Ascent Phase) = 6.4%
        Pr (Proficiency | Ascent Phase) = 3.0%
        Odds Ratio = 2.1
    Physical Factors
        Pr (Physical Factors | Altitude Deviation, Ascent Phase) = 8.1%
        Pr (Physical Factors | Ascent Phase) = 3.9%
        Odds Ratio = 2.1
    Duty Cycle
        Pr (Duty Cycle | Altitude Deviation, Ascent Phase) = 4.2%
        Pr (Duty Cycle | Ascent Phase) = 2.2%
        Odds Ratio = 1.9

In-Flight Weather Encounter is associated with:
    Taskload
        Pr (Taskload | Altitude Deviation, Ascent Phase) = 13.5%
        Pr (Taskload | Ascent Phase) = 5.9%
        Odds Ratio = 2.3

Cruise Phase

Altitude Deviation is associated with:
Pre-occupation
Pr (Pre-occupation | Altitude Deviation, Cruise Phase) = 15.8%
Pr (Pre-occupation | Cruise Phase) = 4.8%
Odds Ratio = 3.3
Physical Factors
Pr (Physical Factors | Altitude Deviation, Cruise Phase) = 8.4%
Pr (Physical Factors | Cruise Phase) = 4.3%
Odds Ratio = 2.0

In-Flight Weather Encounter is associated with:
Pressure
Pr (Pressure | In Flight Wx Encounter , Cruise Phase) = 3.9%
Pr (Pressure | Cruise Phase) = 2.2%
Odds Ratio = 1.8

Conflict (airborne) is associated with:
Communication Environment
Pr (Communication Environment | Air Conflict, Cruise Phase) = 6.8%
Pr (Communication Environment | Cruise Phase) = 4.2%
Odds Ratio = 1.6

Descent Phase

In-Flight Weather Encounter is associated with:
Pressure
Pr (Pressure | In Flight Wx Encounter, Descent Phase) = 4.8%
Pr (Pressure | Descent Phase) = 1.6%
Odds Ratio = 3.1

Altitude Deviation is associated with:
Pre-occupation
Pr (Pre-occupation | Altitude Deviation, Descent Phase) = 15.5%
Pr (Pre-occupation | Descent Phase) = 8.9%
Odds Ratio = 1.8
Proficiency
Pr (Proficiency | Altitude Deviation, Descent Phase) = 6.9%
Pr (Proficiency | Descent Phase) = 4.2%
Odds Ratio = 1.6
Familiar
Pr (Familiar | Altitude Deviation, Descent Phase) = 3.7%
Pr (Familiar | Descent Phase) = 2.4%
Odds Ratio = 1.5

Aircraft Equipment Problem is associated with:
    Weather
        Pr (Weather | Equipment Problem, Descent Phase) = 18.0%
        Pr (Weather | Descent Phase) = 11.8%
        Odds Ratio = 1.5

Approach and Landing Phase

Airspace Violation (Landing without Clearance) is associated with:
    Pre-occupation
        Pr (Pre-occupation | Altitude Deviation, Approach & Landing Phase) = 27.6%
        Pr (Pre-occupation | Approach & Landing Phase) = 6.6%
        Odds Ratio = 4.2
    Physical Factors
        Pr (Physical Factors | Altitude Deviation, Approach & Landing Phase) = 15.9%
        Pr (Physical Factors | Approach & Landing Phase) = 5.7%
        Odds Ratio = 2.8
    Duty Cycle
        Pr (Duty Cycle | Altitude Deviation, Approach & Landing Phase) = 10.0%
        Pr (Duty Cycle | Approach & Landing Phase) = 3.7%
        Odds Ratio = 2.7
    Proficiency
        Pr (Proficiency | Altitude Deviation, Approach & Landing Phase) = 8.2%
        Pr (Proficiency | Approach & Landing Phase) = 3.5%
        Odds Ratio = 2.4

Confirmation of these results was obtained through expert opinion. Only the important Shaping Factors are shown above. However, those factors that were identified as unimportant were equally interesting and were considered in the validation of the results. When the results of the important and the unimportant Shaping Factors that had been identified with each of these anomalous Outcomes were presented to a group of experts in aviation operations, they all agreed that, in each case, these seemed entirely reasonable and plausible. This experiment demonstrated a capability that was validated by expert opinion to analyze the narratives of experiential reports for identification of Shaping Factors of the human behavior in a group of reports of the same anomalous incident.

Although this experiment used ASRS incident reports, the approach is applicable to any similar safety-reporting system. The methodology is sufficiently generic to be used with any database of textual reports of safety-related incidents. This was confirmed by applying the process to a set of Aviation Safety Action Program (ASAP) reports that were in a format quite different from those in the ASRS database. It was demonstrated that these reports could be reliably categorized by their Context and Outcomes into the anomalies that had been defined by the ASRS Office. This process could be adapted to any other set of anomalies or new ones given their definitions and an exemplary report.

It has been noted throughout the discussions of this approach that it relies heavily on the domain-specific knowledge of experts in aviation operations, and this may be seen as a hindrance to its

continuing development and evolution. However, as Posse et al. (Posse (2004)) point out, new machine-learning techniques could reduce the amount of human intervention in the learning process. Some of these require only a small set of seed expressions to extract automatically template expressions correlated to these seeds. These methods hold promise of reducing the learning phase while, at the same time, providing more complex expressions to match those of the domain experts or even uncovering expressions not anticipated by experts that are subtly related to the concept of interest.

4 SUMMARY AND CONCLUSIONS

While this was just the initial step in developing a first-generation capability, the results of the experiments are sufficiently encouraging to believe that it will be possible to achieve the level of reliability of human analysis of narrative reports. Furthermore, automated analyses of textual reports have the importance of advantage of consistency that is not achievable with human analysts.

The merits of an approach that built upon the concept of the Scenario described in Volume I of this report (Maille et al. 2006) have been demonstrated. Even the simplified model of the Scenario may be sufficient to meet the objectives of automated analysis of aviation-incident reports. It was particularly useful to the objective of this study to discover that reporters typically try to present their perceptions of why they performed as they did, whereas they seldom (if ever) describe the dimensions of their loss of situational awareness. Some success has also been achieved in pushing forward, just a little, the state of the art of practical application of Natural Language Processing using it in conjunction with more classical techniques of statistical analyses. This approach to automated analysis of experiential reports of aviation incidents is expected to have wide application to extracting useful information about why an incident occurred from any of the very large variety of textual databases that exist across the aviation industry.

5 POSTSCRIPT

This important research on the development of efficient and reliable automated analyses of incident reports continues. In fact, studies conducted while this report was in preparation of new statistical techniques have put into question the value added by using Natural Language Processing in certain steps of the process. Preliminary results indicate that categorization of a report by event type (i.e., what happened) using these new statistical methods based solely on the free narrative (i.e., without the fixed fields) may perform as well, or better, than NLP. It is still to be seen how well these new statistical methods will work on extracting the more subjective information related to Shaping Factors.

The work described in the two volumes of this report was carried out over five years, and a great deal was learned about the problem of trying to understand why human operators of the aviation

system sometimes make mistakes by analyzing their reports. Some of these lessons are worth documenting.

1. The effective application of the process requires collaboration among experts in aviation operations, computational linguistics, computer sciences, human factors, and statistical analyses. Access to experts in aviation operations throughout the development is essential to maximize the efficacy of automated data mining.

2. Pre-processing the text to the language of the domain is useful regardless of the subsequent analysis methodology employed. However, PLADS would have to be modified, with the help of the domain expert, for application to textual reports in some other domain.

3. Statistical text mining has been very effective, especially when the data set conforms to a well-defined taxonomic structure, when there is coherence in the concept to be identified, and when there are "truth data" available for training.

4. Natural Language Processing, albeit limited and costly at present, has shown itself to be effective at identifying moderately illusive concepts. However, the level of performance achieved is highly dependent on the effective use of experts, pre-processing, and statistical analyses mentioned above.

6   REFERENCES

Chappell, S.: Using Voluntary Incident Reports for Human Factors Evaluations. Johnston, N.; McDonald, N.; and Fuller, R. (Eds.), Aviation Psychology in Practice, Ashgate, VT, 1997, pp. 149–169.

Chidester, T. R.: An Overview of the Enhanced Aviation Performance Measuring System. Fifth GAIN World Conference Proceedings and Products, Miami, FL, December 5–6, 2001.

Chidester, T. R.: Understanding Normal and Atypical Operations through Analysis of Flight Data. Proceedings of the 12th International Symposium on Aviation Psychology, Dayton, OH, 2003.

Connell, L.: Incident Reporting: The NASA Aviation Safety Reporting System. GSE Today, 1999, pp. 66–68.

Cook, R. I.; and Woods, D. D.: Operating at the Sharp End: The Complexity of Human Error. Bogner, M. S. (Ed.), Human Error in Medicine. Lawrence Erlbaum, Mahwah, NJ, 1994.

Cunningham, H.; Maynard, D.; Bontcheva, K.; and Tablan, V.: GATE: An Architecture for Development of Robust HLT Applications. Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL 2002), July, 2002.

Durso, F. T.; and Gronlund, S. D.: Situation Awareness. Durso, F. T.; Nickerson, R.; Schvaneveldt, R.; Dumais, S.; Lindsay, S.; and Chi, M. (Eds.), The Handbook of Applied Cognition, Wiley and Sons, New York, NY, 1999, pp. 283–314.

Endsley, M. R.: Design and Evaluation for Situation Awareness Enhancement. Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim, CA, Oct. 24–28, 1988, pp. 97–101.

Endsley, M. R.: Situation Awareness in Dynamic Human Decision Making: Theory. Gilson, R. D.; Garland, D. J.; and Koonce, J. M. (Eds.), Situational Awareness in Complex Systems, Embry-Riddle Press, Orlando, FL, 1994, pp. 27–58.

Endsley, M. R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, vol. 37, no. 1, 1995a, pp. 32–64.

Endsley, M. R.: A Taxonomy of Situation Awareness Errors. Fuller, R.; Johnson, N.; and McDonald, N. (Eds.), Human Factors in Aviation Operations: Proceedings of the Western European Association of Aviation Psychology 21st Conference, Dublin, Ireland: Avebury Aviation, 1995b.

Endsley, M. R.: Automation and Situation Awareness. Parasuraman, R.; and Mouloua, M. (Eds.), Automation and Human Performance. Lawrence Erlbaum, Mahwah, NJ, 1996, pp. 163 – 181.

Endsley, M. R.: Theoretical Underpinnings of Situational Awareness: A Critical Review. Endsley, M. R.; and Garland. D. J. (Eds.), Situation Awareness Analysis and Measurement. Lawrence Erlbaum, Mahwah, NJ, 2000a.

Endsley, M. R.: Situation Models: An Avenue to the Modeling of Mental Models. Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA, 2000b.

Ericsson, K. A.; and Simon, H. A.: Protocol Analysis: Verbal Reports as Data. Revised edition. MIT Press, Cambridge, MA, 1993.

Ferryman, T. A.; Willse, Alan; and Cooley, Scott: Cluster Analysis of Digital Flight Data for the Aviation Performance Measurement System. Proceedings of American Statistics Association, Joint Statistical Meetings, Atlanta, GA, August 5–9, 2001.

Ferryman, T. A.; White, A. M.; Posse, C.; and Swickard, A. R.: State Transition Model Investigation of Relationships Between Aviation Contexts and Incidents. PNWD-3342, Battelle Pacific Northwest National Laboratory, Richland, WA, 2003.

Gibson, J.; Orasanu, J.; Villeda, E.; and Nygren, T. E.: Loss of Situation Awareness: Causes and Consequences. Proceedings of the 9th International Symposium on Aviation Psychology, Ohio State University, Columbus, OH, April 27–May 1, 1997, pp. 1417–1421.

Gronlund, S. D.; Ohrt, D. D.; Dougherty, M. R. I.; Perry, J. L.; and Manning, C. A.: Aircraft Importance and its Relevance to Situation Awareness. DOT/FAA/AM-98/16, Office of Aviation Medicine, Washington, DC, 1998.

Hartel, C. E.; Smith, K.; and Prince, C.: Defining Aircrew Coordination: Searching Mishaps for Meaning. Proceedings of the 6th International Symposium on Aviation Psychology, Department of Aviation, Ohio State University, Columbus, OH, 1991.

Heinrich, H.: Industrial Accident Prevention: A Scientific Approach (4th ed.). McGraw-Hill, New York, NY, 1959.

Jones, D. G.; and Endsley, M. R.: Sources of Situation Awareness Errors in Aviation. Aviation, Space, and Environmental Medicine, vol. 67, no. 6, 1996, pp. 507–512.

Maille, N. P.; Ferryman, T. A.; Rosenthal, L. J.; Shafto, M. G.; and Statler, I. C.: What Happened and Why: Toward an Understanding of Human Error Based on Automated Analyses of Incident Reports—Vol. I. NASA/TP–2006-213466, 2006.

Manning, C. D.; and Schültz, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA, 1999.

Posse, C.; Matzke, B.; Anderson, C.; Brothers, A.; Matzke, M.; and Ferryman; T.: Extracting Information from Narratives: An Application to Aviation Safety Reports. IEEE AC paper no. 1182, 2005.

Reynard, W. D.; Billings, C. E.; Cheaney, E. S.; and Hardy, R.: The Development of the NASA Aviation Safety Reporting System, NASA RP 1114, 1986.

Shively, R. J.; Brickner, M.; and Silbiger, J.: A Computational Model of Situational Awareness Instantiated in MIDAS. Proceedings of the 9th International Symposium on Aviation Psychology, Department of Aviation, Ohio State University, Columbus, OH, April 27–May 1, 1997.

Sohn, Y. W.; and Doane, S. M.: Predicting Individual Differences in Situation Awareness: The Role of Working Memory Capacity and Memory Skill. Proceedings of the First Human Performance Situation Awareness and Automation Conference, Savannah, GA, 2000, pp. 293–298.

Sasou, K.; and Reason, J.: Team Errors: Definition and Taxonomy. Reliability Engineering and System Safety, vol. 65, pp. 1–9, Elsevier, 1999.

Statler, I. C.; Morrison, R.; and Rosenthal, L.: Beyond Error Reporting Toward Risk Assessment. Proceedings of the 12th International Symposium on Aviation Psychology, Dayton, OH, 2003.

Swain, A. D.: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Application. Rept. No. SAND-80-0200, Sandia National Labr., Albuquerque, NM, Aug. 1983.

Van Delden, S.; and Gomez, F.: Retrieving NASA Problem Reports: A Case Study in Natural Language Information Retrieval. Data & Knowledge Engineering, vol. 48, no. 2, Feb. 2004, pp. 231–248.

Willse, A. R.; Cooley, S. K.; Chappell, A. R.; Daly, D. S.; and Ferryman, T. A.: Analysis of Rotary-wing Aircraft Incident and Accident Reports. PNWD-3070, Battelle Pacific Northwest National Laboratory, Richland, WA, 2002.

Woods, D. D.; Johannesen, L. J.; Cook, R. I.; and Sarter, N. B.: Behind Human Error: Cognitive Systems, Computers, and Hindsight. CSERIAC SOAR 94-01, Information Analysis Center, Wright-Patterson Air Force Base, OH, 1994.

APPENDIX A: WORKSHOP PARTICIPANTS

Name	Areas of Expertise	Position (at the time of the Workshop)
Mr. Alan Brothers	Automated text analyses, statistical analyses	Battelle PNWD** Research Scientist
Captain Charles Drew	Retired corporate flight captain, ASRS database and ASRS research	Battelle* Principal Research Scientist
Dr. Gaston Cangiano	Modeling situation awareness, human performance modeling, computational linguistics	SJSU Post-doc student
Dr. Kevin Corker	Human performance modeling, human factors, computer sciences	Professor and Assistant Dean of Engineering, SJSU
Dr. Thomas Ferryman	Multivariate statistics, intelligent systems	Battelle, PNWD** Chief Scientist
Ms. Stephanie Frank	ASRS operations and ASRS database	Battelle* Assistant ASRS Program Manager
Captain Robert Lynch	Retired flight captain, flight operations	Battelle* APMS Manager
Mr. Vince Mallone	Retired air traffic controller, ASRS analyst	Battelle* ASRS Manager
Dr. Michael McGreevy	Human factors, automated text analysis	NASA Research Scientist
Ms. Rowena Morrison,	English language grammarian, ASRS research specialist	Battelle* Senior Research Scientist
Dr. Christian Posse	Natural language processing, computer sciences	Battelle PNWD** Scientist
Mr. Loren Rosenthal	ASRS operations and data utilization, designer of the X-Form	Battelle* Manager, Mountain View Operations
Dr. Michael Shafto	Human factors, computer sciences, human-automation interaction	NASA Research Scientist
Dr. Irving Statler	ASMM Project Manager, human factors, data analysis	NASA Research Engineer

*Battelle, contractor to NASA for ASRS and APMS programs

**Battelle Pacific Northwest Division (PNWD), home of DOE's Pacific Northwest Laboratory

APPENDIX B: ASRS ANOMALIES

ASRS Anomaly codes provide classification of abnormal or irregular events, or events at variance with ATC clearances or instructions. An Anomaly is typically a negative descriptor, i.e., an abnormality, irregularity, or a deviation from an expected operation or rule. The "Other" field is not used to enter acts/events that are neither illegal nor deviate from published procedure. For example, a controller technique issue that is not illegal or does not violate published procedure is not entered in "Other."

Category	Test Set
Aircraft equipment problem	
– *Critical*	
– *Less severe*	#1
Airspace violation	
– *Entry*	
– *Exit*	#10
Altitude deviation	
– *Overshoot*	
– *Undershoot*	#2
– *Excursion from assigned altitude*	
– *Crossing restriction not met*	
Other spatial deviation	
– *Altitude heading rule deviation*	
– *Controlled flight towards terrain*	
– *Track or heading deviation*	
– *Uncontrolled traffic pattern deviation*	
– *Descent below MSA*	
Ground excursion	
– *Ramp*	
– *Runway*	
– *Taxiway*	
Ground incursion	
– *Taxiway*	
– *Landing without clearance*	#6
– *Runway*	#5
Ground encounters	
– *Animal*	
– *Birds*	
– *Foreign object damage (FOD)*	
– *Person*	
– *Vehicle*	
– *Gear-up landing*	

- *Jet Blast*
- *Other*

Conflict
- *Airborne, less severe*
- *Airborne, critical*   **#3**
- *Near mid-air collision (NMAC)*
- *Ground, less severe*
- *Ground, critical*   **#4**

In-flight encounter
- *Birds*
- *Turbulence*
- *Skydivers*
- *Wake turbulence*
- *Weather*   **#7**
- *VFR in IMC*
- *Other*

Maintenance problem
- *Improper maintenance*
- *Improper documentation*   **#8**
- *Non-compliance with MEL*

Cabin event
- *Galley fire*
- *Passenger misconduct*
- *Passenger illness*
- *Passenger contraband*
- *Passenger electronic device*
- *Other*

Non-adherence
- *Clearance*
- *FAR*
- *Published procedure*   **#9**
- *Company policies*
- *Required legal separation*
- *Other*

Other anomaly
- *Loss of aircraft control*
- *Unstabilized approach*
- *Hard landing*
- *Tail strike*
- *Speed deviation*
- *Smoke or fire*
- *Hazardous material violation*
- *Fumes*
- *Other*

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
28/12/2006	Technical Publication	02/2005 – 04/2006

4. TITLE AND SUBTITLE

What Happened, and Why: Toward an Understanding of Human Error Based on Automated Analyses of Incident Reports—Vol. II

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thomas A. Ferryman[1]; Christian Posse[1]; Loren J. Rosenthal[2]; Ashok N. Srivastava, Ph.D.[3]; and Irving C. Statler, Ph.D.[3]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

457280.02.07.01.07

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

[1]Battelle/Pacific Northwest National Laboratory, Richland, WA 99352
[2]Battelle Memorial Institute, Mountain View, CA 94043
[3]Ames Research Center, Moffett Field, CA 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

A070001

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D. C. 20546-0001

10. SPONSORING/MONITOR'S ACRONYM(S)

NASA

11. SPONSORING/MONITORING REPORT NUMBER

NASA/TP–2006-213490

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified—Unlimited   Distribution: Standard
Subject Category: 05, 53
Availability: NASA CASI (301) 621-0390

13. SUPPLEMENTARY NOTES

Point of Contact: Irving C. Statler, Ames Research Center, MS 262-4, Moffett Field, CA 94035-1000
(650) 960-6003 or (650) 604-6655

14. ABSTRACT

The objective of the Aviation System Monitoring and Modeling project of NASA's Aviation Safety and Security Program was to develop technologies to enable proactive management of safety risk, which entails identifying the precursor events and conditions that foreshadow most accidents. Information about what happened can be extracted from quantitative data sources, but the experiential account of the incident reporter is the best available source of information about why an incident happened. In Volume I, the concept of the Scenario was introduced as a pragmatic guide for identifying similarities of what happened based on the objective parameters that define the Context and the Outcome of a Scenario. In this Volume II, that study continues into the analyses of the free narratives to gain understanding as to why the incident occurred from the reporter's perspective. While this is just the first experiment, the results of our approach are encouraging and indicate that it will be possible to design an automated analysis process guided by the structure of the Scenario that can achieve the level of consistency and reliability of human analysis of narrative reports.

15. SUBJECT TERMS

Aviation safety, Textual-data mining, Data analysis, Human error, Statistical analysis, Text analysis

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
Unclassified	Unclassified	Unclassified	Unclassified	34

19a. NAME OF RESPONSIBLE PERSON

Irving C. Statler

19b. TELEPHONE *(Include area code)*

(650) 960-6003 or (650) 604-6655