

ANALYZING COCKPIT COMMUNICATION: THE LINKS BETWEEN LANGUAGE, PERFORMANCE, ERROR, AND WORKLOAD

J. Bryan Sexton & Robert L. Helmreich¹
The University of Texas Team Research Project
Department of Psychology
The University of Texas at Austin
Austin, Texas USA

ABSTRACT

The importance of communication on the flightdeck is discussed and an application of a new computer-based linguistic method of text analysis is introduced. Preliminary results from a NASA B727 simulator study indicate that specific language variables are moderately to highly correlated with individual performance, individual error rates, and individual communication ratings. Also, language use was found to vary as a function of crew position and level of workload during the flight. Use of the first person plural (we, our, us) increases over the life of a flightcrew, and CA's speak more in the first person plural than FO's or FE's. Language use in initial flights was associated with performance and error in subsequent flights. These are preliminary data, in that this method of linguistic analysis is currently being developed and integrated with a content-coding method of communication analysis and models of threat and error.

INTRODUCTION

Communication in Aviation

NASA researchers analyzed the causes of jet transport accidents and incidents between 1968 and 1976 (Cooper, White & Lauber, 1980; Murphy, 1980) and concluded that pilot error was more likely to reflect failures in team communication and coordination than deficiencies in technical proficiency. Human factors issues related to interpersonal communication have been implicated in approximately 70% to 80% of all accidents over the past 20 years. Correspondingly, over 70% of the first 28,000 reports made to NASA's Aviation Safety Reporting System (which allows pilots to confidentially report aviation incidents) were found to be related to communication problems (Connell,

1995). In order for cockpit crewmembers to share a "mental model," or common understanding of the nature of events relevant to the safety and efficiency of the flight, communication is critical. This is not to say that effective communication can overcome inadequate technical flying proficiency, but rather the contrary, that good "stick & rudder" skills can not overcome the adverse effects of poor communication. Ruffell Smith's (1979) landmark full-mission simulator study showed that crew performance was more closely associated with the quality of crew communication than with the technical proficiency of individual pilots or increased physiological arousal as a result of higher environmental workload. No differences were found between the severity of the errors made by effective and ineffective crews, rather, it was the ability of the effective crews to communicate that kept their errors from snowballing into undesirable outcomes.

Language

The interdependent nature of the cockpit crew requires language in order to issue commands, acknowledge commands, conduct briefings, perform standard callouts, state intentions, ask questions, and convey information (Kanki, 1995). The role of language use in communication processes has been neglected, and researchers have recognized the need for a deeper understanding of language, its basic characteristics, and how it works (Orasanu & Fischer, 1991; Cushing, 1997). There is a wealth of Cockpit Voice Recorder (CVR) data that provide recordings of crewmember interactions related to aviation accidents. Current methods of content-coding utterances from accident and/or simulation transcripts have provided valuable (albeit labor intensive) insights into communication processes (e.g., Orasanu & Fischer, 1991; Predmore, 1991).

¹ ¹ Research reported here was conducted with the support of FAA Grants 92-G-017, 99-G-004, and a grant from the Daimler-Benz Stiftung, Robert Helmreich, Principal Investigator. Acknowledgements go to James Klinect, John Wilhelm, Dave Musson, and Janel Seagal for reading earlier drafts. Thanks are due to research assistants Kathrine Kirby, Anne-Marie Deville, Renee Mauldin, C.J. Winkler, and Dolly Dimke for their many "less than exciting" hours of transcription.

An automated system of linguistic analysis that could quickly and efficiently identify patterns in communication as well as provide a new level of insight into language use in aviation, would be an ideal tool to complement current methods of investigation². To this end, the present research examines the use of a computer-based linguistic analysis of aviation communication.

Research conducted in the language lab of Professor James W. Pennebaker at The University of Texas at Austin has shown that linguistic styles can be considered individual difference markers, i.e., individuals appear to have a distinct “language use fingerprint” which is relatively stable across time and situations (Pennebaker & King, 1999). His lab has identified language dimensions to be internally consistent, and modestly correlated to objective and self-reported health and performance measures at rates comparable to or greater than traditional trait markers of personality such as the “big five” (McCrae & Costa, 1987).

Linguistic Inquiry and Word Count (LIWC)

In order to provide an efficient and practical method for studying the various emotional, cognitive, structural, and process components present in individuals' language use, Pennebaker and Francis developed and validated a computer-based text analysis program called Linguistic Inquiry and Word Count (LIWC: Francis & Pennebaker, 1993; Pennebaker & Francis, 1999). In essence, LIWC analyzes written or transcribed verbal text files by looking for dictionary matches to words in the text file. LIWC does this on a word by word basis by calculating the percentage of words in the text that match each of up to 85 language dimensions. Standard linguistic dimensions include categories such as word count, sentence punctuation, percentage of words longer than 6 letters, 1st person plural (we, our, us), negations (no, never, not), and assents (yes, OK, mmhmm). Dimensions of psychological processes include categories such as positive emotions (happy, pretty, good), anger (hate, resent, pissed), and cognitive processes (cause, know, effect, maybe, would, should). Other categories include swear words, nonfluencies (uh, er, um), and fillers (you know, I mean).

Identifiable Language Patterns & LIWC Categories

Foushee (1984) observed that most of the research studies of team performance have ignored

communication as a process variable moderating the relationship between input (e.g. team size, team structure, composition) and output factors (e.g., quality, latency, errors). Furthermore, Foushee and Helmreich (1989) have noted that closely examining group member *communications* has often proven fruitful in studies of the links between group processes and outcomes. Unlike previous research however, the current study did not attempt to content-code or classify communication in terms of its final product, or meaning. Linguistic factors were utilized to bypass the final product and instead investigate the tools used to build it (i.e., not what the person is talking about, but how the person is talking). To expand upon this concept, LIWC language dimensions will be used to isolate patterns of language use in the current study.

In a closer examination of the Ruffell Smith (1979) data, Foushee and Manos (1981) found that better performing crews communicated more overall. This positive correlation between performance and communication was also documented in a Bell Aeronautics Company study in 1962 (Siskel and Flexman) and replicated in Foushee, Lauber, Beatge, & Acomb, (1986). The quantity of verbal language used in the cockpit is a pattern that has been identified and related to performance measures, and this pattern is readily captured by the LIWC category **WORDCOUNT** (overall number of words spoken). Furthermore, in an analysis of simulator communication processes, Veinott and Irwin (1995) reported that, in general, captains talked more than first officers and there was more communication during the abnormal phase than the normal phase. Accordingly, the current study will investigate differences in language use as a function of position and workload. Cannon-Bowers, Tannenbaum, Salas, & Volpe (1995) identified a specific cockpit communication skill: conveying information concisely. This finding prompted the author to consider a related linguistic dimension, the size of the words used by pilots (big vs. small words) when examining language use. The LIWC category **SIXLTR** is a tabulation of the percent of words greater than six letters long. In the past, the use of the first person plural (we) has been interpreted as a collective or team focus by the speaker (McGreevy, 1996; White and Lippitt, 1960). The first person plural is a linguistic dimension captured by the LIWC category **WE** (e.g., we, our, us). Another linguistic dimension which is theoretically related to a healthy teamwork orientation is the LIWC category **ACHIEVE** (words indicative of working toward a successful outcome: e.g., try, effort, goal).

Prior research has demonstrated that shared mental models and predictable patterns of behavior

² Current work at our lab is integrating a micro-coding technique (Predmore, 1991) with the error model (Helmreich, Klinect & Wilhelm, this volume).

are imperative to laying the groundwork for effective teamwork and communication in the cockpit. Ginnett (1987) has shown that initial crewmember interactions set the tone and can predict subsequent team performance. Similarly, Hines (1998) demonstrated that the best predictors of line performance are captain leadership and pre-flight briefing content. In this vein, one of the goals of the current study was to investigate the extent to which initial language use is related to subsequent performance and error ratings.

Research Questions

This investigation explores the use of language as an individual difference variable in the context of studying communication processes in the highly technical and task-oriented commercial aviation cockpit. LIWC was used to conduct linguistic analyses conceptually similar to methods already in use by crew factors investigators. Is language use related to flight outcome measures, and how does language use vary across position and or level of workload?

METHOD

The pilot language data for the present study were a study from NASA-Ames Research Center that investigated the effects of captain personality on crew performance (see Chidester, Kanki, Foushee, Dickinson & Bowles, 1990 for a detailed description). These data involved a three person crew (captain-CA, first officer-FO, and flight engineer-FE) flying a simulated B-727 during a 5 segment flight over two days. Transcripts were made available for 12 crews³ across the last four flight segments, A, B, C, and D. Segments A and C were routine (low workload), and segments B and D were abnormal (high workload). As part of the data collection process in the original simulator study, an expert pilot observer was present in the simulator cab. Flight outcome data regarding individual performance, individual errors, and individual communication skill were collected by this observer for each pilot by segment. These data were collected in the original study, and were in no way coded, or summarized for use in the current study. Note that no crew level variables will be reported in the current study, i.e., whenever performance, error, or communication skill is mentioned, it refers to individual pilots, not the cockpit crew. Previous research by our lab (unpublished data) has found that the relationships between crew-level flight outcome measures and LIWC categories are weaker and far

³ Our lab is currently transcribing the remainder of the tapes from this study.

fewer in number. Communication in the cockpit is dynamic and interdependent, hence collapsing data across all positions when investigating language dilutes the effects of individuals - ergo the use of individual language and flight outcome variables.

RESULTS

Language use data were analyzed using a doubly multivariate within-subjects repeated measures ANOVA model 4 X 3 (level of workload⁴ X cockpit position). The LIWC categories (WORDCOUNT, SIXLTR, WE, and ACHIEVE) were used as repeated measures⁵.

The model⁶ yielded an omnibus main effect for position $F(8, 4) = 38.9, p < .001$, indicating that across all crews, language is used differently as a function of cockpit position. A within-subjects multivariate test main effect for level of flight segment workload $F(12, 96) = 8.76, p < .001$, indicated that language is used differently as a function of workload. Another within-subjects multivariate test main effect for cockpit position $F(8, 40) = 12.94, p < .001$, indicated that there are differences in language use between members of the same cockpit crew. A within-subjects interaction between workload and position $F(24, 264) = 4.30, p < .001$, showed that the overall differences in language use by position were interdependent with differences in language use by workload.

Pairwise comparisons using the Bonferroni correction for multiple comparisons at the .05 level revealed further insight into the nature of differences identified by the multivariate F tests. Captains used a higher number of words per segment than FO's or FE's, and the number of words used during abnormal segments was higher than the number of words used in routine segments - a replication of Veinott and Irwin's (1995) findings. Individuals are likely to communicate more during periods of high workload due to the inherent multi-tasking involved in flight deck management.

Captains also tended to use the first person plural (WE) more often than FO's and FE's, particularly during the stressful segments. There was a linear increase in the use of the first person plural across time, i.e., the average rates of WE (collapsed across

⁴ Recall that there were four segments, but only two levels of workload (routine vs. abnormal).

⁵ This design accounts for the autocorrelations amongst members of the same crew, insuring that the MANOVA independence of observations assumption was not violated.

⁶ In a separate between-within MANOVA, no differences in language use were found between groups of crews led by captains with a particular personality profile. The source of these data (an investigation of the influence of captain personality) does not appear to have systematically compromised its' use in a linguistic analysis.

position) increase with each flight. Perhaps this could be considered an indicator of increasing familiarity over time. Flight engineers use higher rates of big words than captains and first officers across all four segments (most likely due to technical communication and checklists).

Links between Language Use, Performance, Error, and Communication Skill

Understanding variations in the way language is used (e.g., MANOVA results) is important to the extent that language use is related to flight safety. Examination of the correlation matrix for language and flight safety outcome measures (performance, error rates, and individual communication skills) revealed consistent and interpretable patterns⁷. Individual language use was moderately to highly correlated with individual performance, individual error rates, and ratings of individual communication skills. In sum, WE, ACHIEVE, and WORDCOUNT were positively related to performance and communication, but negatively related to rates of error. For example, flight engineer WORDCOUNT in segment A (routine) was positively correlated with flight engineer performance in the high workload flights B $r(12) = .81$, $p < .001$, and D $r(12) = .65$, $p = .022$, but negatively correlated with flight engineer error in B $r(12) = -.63$, $p = .03$. In other words, flight engineers who used a high number of words to communicate during segment A had higher subsequent performance ratings⁸ and made fewer subsequent errors.

In contrast to the relationships between WE / WORDCOUNT / ACHIEVE, and flight outcome measures, the use of large words (SIXLTR) was *negatively* related to performance variables and communication skill, but positively related to rates of error⁹. The percentage of large words flight engineers used during segment A was also consistently negatively related to performance in segments B $r(12) = -.71$, $p = .009$, C $r(12) = -.84$, $p = .001$, and D $r(12) = -.69$, $p = .013$. Flight engineers who used relatively low rates of large words tended to receive higher subsequent performance ratings.

There were also intra-cockpit relationships between language use and flight outcome data, and these differences spanned across flights. The effects

were strongest for the influence of CA language use on the performance and error variables of the FO and FE. Language use in earlier segments (particularly flight A) appeared to be more highly related to outcome measures in subsequent flights, than the within segment language and performance / error correlations. Along these lines, establishing predictable patterns of behavior during initial interactions is related to subsequent outcomes. For example, captain WORDCOUNT in segment A was positively correlated with flight engineer performance in B $r(12) = .84$, $p = .001$, and C $r(12) = .84$, $p = .001$. Furthermore, captain's use of achievement words in segment A was positively correlated with flight engineer performance in B $r(12) = .74$, $p = .006$, and negatively related to flight engineer errors in segment B $r(12) = -.73$, $p = .007$. Such relationships between language use by one position and performance / errors in another position were not uncommon. Also, the relationships between language use and flight outcome measures were strongest between segment A language use, and subsequent (B, C, and D) performance and error measures. This is conceptually similar to the relationship between initial interpersonal interactions and subsequent performance identified by Ginnett (1987) and Hines (1998). In the current study, use of the first person plural in segment A was highly correlated with performance, and could be a marker of familiarity or a more collective orientation towards the crew. WE could be an important marker in that crew familiarity has been implicated as a moderating variable of aviation accident (NTSB, 1994).

Interpretation

The performance and error outcomes of cockpit crews are multiply determined, and language use is just one of the many group process factors which can impact flight outcomes. The results of this exploratory study indicate that the language use of pilots varies as a function of who is talking (captain, first officer, or flight engineer) and as a function of workload. The correlations between the LIWC categories and flight outcome measures of safety such as performance, error, and communication skill serve as a form of convergent validity for the utility of linguistic analyses.

Perhaps the fact that flight engineers used bigger words than captains and first officers is due to the role of the flight engineer. Mosier (1991) has noted that the flight engineer, can have tremendous control over information disseminated in the cockpit. They are responsible for checklists and procedures as a "keeper of the information." Part of the reason why flight engineers use larger words could be due to the highly technical nature of their communication, and

⁷ Although the Bonferroni correction was not met for many of the correlations, the magnitude of the correlations suggests a strong linear relationship between language use and the outcome measures.

⁸ Interestingly, it was possible to estimate the performance of the individual by simply looking at the kilobytes of disk space required for each file on my computer.

⁹ WORDCOUNT and SIXLTR were moderately negatively correlated. Also, CA, FO, and FE WORDCOUNT within the same segment were not related.

the large number of checklists they must perform. This of course does not account for why the use of bigger words was negatively correlated with performance. It could be the case that the ability to communicate concisely includes the ability to apply a short and succinct vocabulary. Conceivably, those individuals who expend the cognitive resources necessary to speak more elaborately (using bigger words) do so at the expense of decreased situational awareness.

Evidence was found for the positive correlation between the number of words used and performance (as in Foushee & Manos, 1981), and a negative relationship was found between WORDCOUNT and error. This may be due to the increasing predictability of behavior patterns as the amount of verbal communication increases.

The use of the first person plural was more common for captains relative to first officers and flight engineers. This could be due to the status and role of the captain. This role requires more active team building, and the status of the captain affords the right to use the first person plural ("we need to..," "our problem...," "let's get out the before landing checklist," etc.) when briefing, planning, or simply addressing the crew in conversation.

DISCUSSION

Why should language use be considered as a variable of interest in crew factors? Language is a coping mechanism in that it helps individuals lessen and manage both the causes and the effects of stress. Just as there are links between pilot personality and performance, there appear to be links between pilot language use and flight outcome measures.

Language use is dynamic, as it is sensitive to both workload and position, and varies systematically with flight outcome measures. This is in contrast to a more static measure such as "paper and pencil" measurements of personality, which are stable across time and situation. Furthermore, research into language use has the potential to yield a clearer understanding of thought processes involved *during* communication. Linking "online communication" to personality test results collected at a different time is difficult, if at all possible. Such personality tests are also difficult to link to pilot communication captured by accident CVR's. In contrast, LIWC variables are easily extracted from CVR transcripts, and could potentially add a new level of insight into what took place in the minds of the crewmembers.

Without making a causal statement, it is important to consider the potential implications that language use could have for training. Perhaps the use of more effective language styles is trainable,

whereas personality interventions (aside from selection) are generally futile. The findings of Siegel & Federman (1973), indicate that it is possible to isolate aspects of effective communication and to train those aspects successfully (in this case using anti-submarine helicopter crews). Chidester, Helmreich, Gregorich, & Geis, (1990) have assessed the trainability of specific behaviors and have found that human factors training produces measurable improvements in communication skills.

A possible limitation of this study is that these results should not be generalized to two-person crews, where the social structure is conceptually different from the more hierarchical, more formal three-person crews. Small sample size is a limiting factor in this study, but current work at our lab is adding an additional six crews to the dataset, as well as a growing list of commercial aviation accident CVR's. In addition, we are integrating an error taxonomy with the micro-coding and LIWC techniques to systematically investigate the relationship between what is said (micro-coding), how it is said (LIWC), and how it relates to error rates, error responses, and undesired state responses (Helmreich, Klinect and Wilhelm, this volume).

LIWC is not intended to replace current methods of communication analyses such as content coding, but rather to complement any study already using transcripts of pilot communication. LIWC required 6 seconds to generate the language data used for the current study. Furthermore, this study used only 4 of the possible 85 LIWC categories. LIWC is a fast, easy, and practical addition to any investigation of communication, in that it can quickly identify aspects of communication to be explored in-depth using other methodologies.

REFERENCES

- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E. & Volpe C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations*. (pp. 333-380). San Francisco: Jossey Bass.
- Chidester, T.R., Helmreich, R.L., Gregorich, S.E., & Geis, C. (1990). Pilot personality and crew coordination: Implications for training and selection. *International Journal of Aviation Psychology*, 1, 23-42.
- Chidester, T. R., Kanki, B. G., Foushee, H. C., Dickinson, C. L., & Bowles, S. V. (1990). Personality factors in flight operations: Volume 1. Leader characteristics and crew performance in a full-mission air transport simulation. (NASA Technical Memorandum 102259). Moffett Field, CA: NASA-Ames Research Center.

- Connell, L. (1995). *Pilot and controller communications issues*. In B. G. Kanki & O. V. Prinzo (Eds.), *Proceedings of the Methods & Metrics of Voice Communication Workshop*.
- Cooper, G. E., White, M. D., Lauber, J. K. (Eds.). (1980). *Resource management on the flightdeck: Proceedings of a NASA/Industry workshop* (NASA CP-2120). Moffett Field, CA: NASA-Ames Research Center.
- Cushing, S. (1997). *Language differences in aviation communication: Problems and solutions*. In P. Quigley & P. McElwain (Eds.), *Proceedings of the Aviation Communication: A Multi-Cultural Forum*.
- Foushee, H. C. (1984). Dyads and Triads at 35,000 feet: Factors affecting group processes and aircrew performance. *American Psychologist*, 39 (8), 885-893.
- Foushee, H. C. & Helmreich, R. L. (1989). Group interaction and Flight Crew performance. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Modern Aviation*.
- Foushee, H. C., Lauber, J. K. Beatge, M. M., & Acomb, D. B. (1986). *Crew Performance as a function of exposure to high density, short-haul duty cycles*. (NASA Technical Memorandum 88322) Moffett Field, CA: NASA-Ames Research Center.
- Foushee, H. C. & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings & E. S. Cheaney (Eds.), *Information transfer problems in the aviation system* (NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center.
- Francis, M. E. & Pennebaker, J. W. (1993). *LIWC: Linguistic Inquiry and Word Count*. Technical Report. Dallas, TX: Southern Methodist University.
- Ginnett, R. C. (1987). *First Encounters of the Close Kind: The Formation Process of Airline Flight Crews*. Yale University: Doctoral dissertation.
- Helmreich, R. L., Klinec, J. R. & Wilhelm J. A. (this volume). Models of threat, error, and CRM in flight operations. In *Proceedings of the Tenth International Symposium on Aviation Psychology*, The Ohio State University.
- Hines, W. E. (1998). *Teams and technology: Flight crew performance in standard and automated aircraft*. The University of Texas at Austin: Unpublished doctoral dissertation.
- Kanki, B. G. (1995). *A Training Perspective: Enhancing Team Performance Through Effective Communication*. In B. G. Kanki & O. V. Prinzo (Eds.), *Proceedings of the Methods & Metrics of Voice Communication Workshop*.
- McCrae, R. R. & Costa P. T. Jr. (1987). Validation of the five factor model across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- NTSB (1994). *A review of flight-crew-involved major accidents of U.S. air carriers, 1978-1990*. National Transportation Safety Board Safety Study, Washington D.C.
- Murphy, M. (1980). *Review of aircraft incidents*. Cited in Cooper et al.
- McGreevy, M. W. (1996). *Reporter Concerns in 300 Mode-Related Incident Reports from NASA's Aviation Safety Reporting System*. (NASA Technical Memorandum 110413). Moffett Field, CA: NASA-Ames Research Center.
- Orasanu, J. & Fischer, U. (1991). Information transfer and shared mental models for decision making. In R. S. Jenson (Ed.), *Proceedings of the Sixth International Symposium of Aviation Psychology*. Columbus, Ohio: The Ohio State University.
- Pennebaker, J. W. & Francis, M. E. (1999). *Linguistic Inquiry and Word Count (LIWC)*. Mahmah, N.J.: Erlbaum Publishing
- Pennebaker J. W., & King, L. A. (1999). *Linguistic styles: Language use as an individual difference*. Manuscript submitted for publication.
- Predmore, S. C. (1991) *Microcoding of communications in accident investigation: Crew coordination in United 811 and United 232*. In R. S. Jenson (Ed.), *Proceedings of the Sixth International Symposium of Aviation Psychology* (pp. 350-355). Columbus, Ohio: The Ohio State University.
- Ruffell Smith, H. P. (1979). *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions*. (NASA Technical Memorandum 78482). Moffett Field, CA: NASA-Ames Research Center.
- Siegel, I., & Federman, P. J. (1973). *Communication Content Training as an Ingredient in Effective Team Performance*. *Ergonomics*, 16 (4), 403-416.
- Siskel, M. & Flexman, R. (1962). Study of effectiveness of a flight simulator for training complex aircrew skills. *Bell Aeronautics Company*.
- Veinott, E. S. & Irwin, C. M. (1995). *Communication methodology issues in aviation simulation research*. In B. G. Kanki & O. V. Prinzo (Eds.), *Proceedings of the Methods & Metrics of Voice Communication Workshop*.
- White, R. & Lippitt, R. (1960). *Leader Behavior and Member Reaction in Three "Social Climates." Autocracy and Democracy*. New York: Harper.