Measuring Conformity to Discourse Routines in Decision-Making Interactions

Sherri L. Condon
Department of English
condo@usl.edu

Claude G. Čech
Department of Psychology
cech@usl.edu

William R. Edwards
Center for Advanced Computer Studies
wre@cacs.usl.edu

University of Southwestern Louisiana/Université des Acadiens
Lafayette, LA 70504

Abstract

In an effort to develop measures of discourse level management strategies, this study examines a measure of the degree to which decision-making interactions consist of sequences of utterance functions that are linked in a decision-making routine. The measure is applied to 100 dyadic interactions elicited in both face-to-face and computer-mediated environments with systematic variation of task complexity and message-window size. Every utterance in the interactions is coded according to a system that identifies decision-making functions and other routine functions of utterances. Markov analyses of the coded utterances make it possible to measure the relative frequencies with which sequences of 2 and 3 utterances trace a path in a Markov model of the decision routine. These proportions suggest that interactions in all conditions adhere to the model, although we find greater conformity in the computer-mediated environments, which is probably due to increased processing and attentional demands for greater efficiency. The results suggest that measures based on Markov analyses of coded interactions can provide useful measures for comparing discourse level properties, for correlating discourse features with other textual features, and for analyses of discourse management strategies.

Introduction

Increasingly, research in computational linguistics has contributed to knowledge about the organization and processing of human interaction through quantitative analyses of annotated texts and dialogues (e.g. Carletta et al., 1997; Cohen et al., 1990; Maier et al., 1997; Nakatani et al., 1995; Passonneau, 1996; Walker, 1996). This program of research presents opportunities to examine the relation between linguistic form and pragmatic functions using large corpora to test hypotheses and to detect covariance among discourse features. For example, Di Eugenio et al. (1997) demonstrate that utterances coded as *acceptances* were more likely to corefer to an item in a previous turn. Grosz and Hirschberg (1992) investigate intonational correlates of discourse structure. These researchers recognize that discourse-level structures and strategies influence syntactic and phonological encoding. The regularities observed can be exploited to resolve language processing problems such as ambiguity and coreference, to integrate high level planning with encoding and interpretation strategies, or to refine statistics-based systems.

In order to identify and utilize discourse-based structures and strategies, researchers need methods of linking observable forms with discourse functions, and our focus on discourse management strategies has motivated similar goals. Condon & Čech (1996a,b) use annotated decision-making interactions to investigate properties of discourse routines and to examine the effects of communication features such as screen size on computer-mediated interactions (Čech & Condon, 1997). In this paper we present a method for measuring the degree to which an

interaction conforms to a discourse routine, which not only allows more refined analyses of routine behavior, but also permits fine-grained comparison of discourses obtained under different conditions.

In our research, discourse routines have emerged as a fundamental strategy for managing verbal interaction, resulting in the kind of behavior that researchers label *adjacency pairs* such as question/answer or request/compliance as well as more complex sequences of functions. Discourse routines occur when a particular act or function is routinely continued by another, and as "predictable defaults," routine continuations maximize efficiency by requiring minimal encoding while receiving highest priority among possible interpretations. Moreover, discourse routines can be exploited by failing to conform to routine expectations (Schegloff, 1986). Consequently, interactions will not necessarily conform to routines at every opportunity, which raises the problem of measuring the extent to which they do conform.

Condon et al. (1997) develop a measure based on Markov analyses of coded interactions, and the measure is employed here with a larger corpus in which students engage in a more complex decision-making task. These measures provide evidence for the claim that participants in computer-mediated decision-making interactions rely on a simple decision routine more than participants in face-to-face decision-making interactions. The measures suggest that conformity to the routine is not strongly affected by any of the other variables examined in the study (task complexity, screen size), even though some participants in the computer-mediated conditions of the more complex task adopted turn management strategies that would be untenable in face-to-face interaction.

Data Collection

The initial corpus of 32 interactions involving simple decision-making tasks was obtained under conditions which were similar, but not identical, to the conditions under which the 68 interactions involving a more complex task were obtained. One obvious difference is that participants in the first study completed 2 simple tasks planning a social event (a getaway weekend, a barbecue), while participants in the second study completed a single, more complex task: planning a televised ceremony to present the MTV music video awards. Furthermore, all interactions in the first study were mixed sex pairs, whereas interactions in the MTV study include mixed and same sex pairs. All participants were native English speakers at the University of Southwestern Louisiana who received credit in Introductory Psychology classes for their participation.

In both studies, the dyads who interacted face-to-face sat together at a table with a tape recorder, while the pairs who interacted electronically were seated at microcomputers in separate rooms. The latter communicated by typing messages which appeared on the sender's monitor as they were typed, but did not appear on the receiver's monitor until the sender pressed a SEND key. The software incorporated this feature to provide well-defined turns and to make it possible to capture and change messages in future studies. In addition, to minimize message permanence and more closely approximate face-to-face interaction, text on the screen is always produced by only one participant at a time.

In the original study, the message area was approximately 4 lines long, and it was not clear how much this factor influenced our results. Consequently, in the MTV study, the message area of the screen was either 4, 10, or 18 lines. Other differences in the computer-mediated conditions of the two studies include differences in the arrangement of information on the screen such as a brief description of the MTV problem which remained at the bottom of the screen. We also used an answer form in the first study, but not the second. More details about the communication systems in the two studies are provided Condon & Čech (1996a) and Čech & Condon (1998).

Data Analysis

Face-to-face interactions were transcribed from audio recordings into computer files using a set of conventions established in a training manual (Condon & Čech, 1992). All interactions were divided into utterance units defined as single clauses with all complements and adjuncts, including sentential complements and subordinate clauses. Interjections like *yeah, now, well,* and *ok* were considered to be separate utterances due to the salience of their interactional, as opposed to propositional, content.

The coding system includes categories for request routines and a decision routine involving 3 acts or functions (Condon, 1986; Condon & Čech, 1996a,b). We believe that the decision routine observed in the interactions instantiates a more general schema for decision-making that may be routinized in various ways. In the abstract schema, each decision has a goal; proposals to satisfy the goal must be provided; these proposals must be evaluated, and there must be conventions for determining, from the evaluations, whether the proposals are adopted as decisions. Routines make it possible to map from the general schema to sequences of routine utterance functions. Default principles associated with routines can determine the encoding of these routine functions in sequences of utterances.

According to the model we are developing, a sequence of routine continuations is mapped into a sequence of adjacent utterances in one-to-one fashion by default. If the routine specifies that a routine continuation must be provided by a different speaker, as in adjacency pairs, then the default is for the different speaker to produce the routine continuation immediately after the first pair-part. Since these are defaults, we can expect that they may be weakened or overridden in specific circumstances. At the same time, if our reasoning is correct, we should be able to find evidence of routines operating in the manner we have described.

(1) provides an excerpt from a computer-mediated interaction in which utterances are labeled to illustrate the routine sequence. P1 and P2 designate first and second speaker (an utterance that is a continuation by the same speaker is not annotated for speaker).

(1) a. P1: [orientation] who should win best Alternative video.
    b. P2: [suggestion] Pres. of the united states
    c. P1: [agreement] ok
    d. P2: [orientation] who else should we nominate.
    e. [suggestion] bush. goo-goodolls and oasis
    f. P1: [agreement] sounds good, [...]

(2) provides an annotated excerpt from a face-to-face interaction.

(2) a. P1: [orientation] who's going to win?
    b. [suggestion] Mariah?
    c. P2: [agreement] yeah probably
    d. P1: [orientation] alright Mariah wins what song?
    e. P2: [suggestion] uh Fantasy or whatever?
    f. P1: [agreement] that's it that's the same song I was thinking of
    g. [orientation] alright alternative?
    h. [suggestion] Alanis?

Coded as "Orients Suggestion," orientations, like (1a,2a) establish goals for each decision, while suggestions like (1b,e) and (2b,e,h) formulate proposals within these constraints. Agreements like (1c,f) and (2c,f), which are coded "Agrees with Suggestion," and disagreements ("Disagrees with Suggestion") evaluate a proposal and establish consensus. The routine does not specify that a suggestion which routinely continues an orientation must be produced by a different speaker: the suggestion may be elicited from a different speaker, as in (1a,b) and (2d,e) or it may be provided by the same speaker, as in (1d,e) and (2a,b). However, an agreement that routinely continues a suggestion is produced by a different speaker, as (1b,c), (1e,f), (2b,c) and (2e,f) attest.

Other routine functions are also classified in the coding system. Utterances coded as "Requests Action" propose behaviors in the speech event such as (3).

(3) a. well list your two down there (oral)
   b. ok, now we need to decide another band to
      perform (computer-mediated)
   c. Give some suggestions
                       (computer-mediated)

Utterances coded as "Requests Information" seek information not already provided in the discourse, as in (1a,2a). Utterances that seek confirmation or verification of provided information, however, are coded as "Requests Validation." The category "Elaborates-Repeats" serves as a catch-all for utterances with comprehensible content that do not function as requests or suggestions or as responses to these.

Two categories are included to assess affective functions: "Requests/Offers Personal Information" for personal comments not required to complete the task and "Jokes Exaggerates" for utterances that inject humor. The category "Discourse Marker" is used for a limited set of forms: *Ok, well, anyway, so, now, let's see,* and *alright.* Another category, Metalanguage, was used to code utterances about the talk such as (3b,c).

In the initial corpus, the categories described above are organized into 3 classes: MOVE, RESPONSE, and OTHER, and each utterance was assigned a function in each of these three groups of categories. In cases involving no clear function in a class, the utterance was assigned a No Clear code. A complete list of categories is presented at the bottom of Figure 1 and more complete descriptions can be found in Condon and Čech (1992). In the modified system used to code the MTV corpus, the criteria for classifying all of these categories remain the same.

The data were coded by students who received course credit as research assistants. Coders were trained by coding and discussing excerpts from the data. Reliability tests were administered frequently during the coding process. Reliability scores were high (80-100% agreement with a standard) for frequently occurring move and response functions, discourse markers, and the two categories designed to identify affective functions. Scores for infrequent move and response functions,
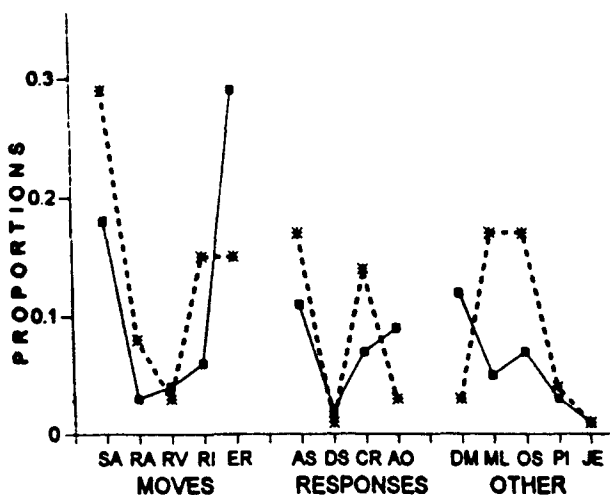
metalanguage, and orientations were somewhat less reliable.

Results

In the initial study, the 16 face-to-face interactions produced a corpus of 4141 utterances (ave. 259 per discourse), while the 16 computer-mediated interactions consisted of 918 utterances (ave. 57). In the MTV study, the 8 face-to-face interactions produced 3593 utterances (ave. 449), the 20 interactions in the 4-line condition included 2556 utterances (ave. 128), the 20 interactions in the 10-line condition produced 3041 utterances (ave. 152) and the 20 interactions in the 18-line condition included 2498 utterances (ave. 125). Clearly, completing the more complex MTV task required more talk.

Figure 1 presents proportions of utterance functions averaged per interaction for each modality in the initial study. Analyses of variance that treated discourse (dyad) as the random variable were performed on the data within each of the three broad categories, excluding the No Clear MOVE/RESPONSE/OTHER functions where inclusion would force levels of the between-discourse factor to the same value. We found no significant effect of problem type or order (for details see Condon & Čech, 1996). However, the interaction of function type with discourse modality was significant at the .001-level for all three (MOVE, RESPONSE, OTHER) function classes. Tests of simple effects of modality type for each function indicated that only four proportions were identical in the two modalities: Requests Validation in the MOVE class, Disagrees in the RESPONSE class, and, in the OTHER class, Personal Information and Jokes-Exaggerates.

Figure 2 presents the proportions of utterance functions for the MTV corpus using the same categories of functions as in Figure 1. The similarity of the results in the two figures is remarkable, especially considering differences in methods of data collection described above. First, it can be observed that

241

Figure 2: Proportions of code categories in face-to-face (triangles), 4-line (squares) and 18-line (circles) conditions

MOVE FUNCTIONS
SA Suggests Action
RA Requests Action
RV Requests Validation
RI Requests Information
ER Elaborates, Repeats

OTHER FUNCTIONS
DM Discourse Marker
ML Metalanguage
OS Orients Suggestion
PI Personal Information
JE Jokes, Exaggerates

RESPONSE FUNCTIONS
AS Agrees with Suggestion
DS Disagrees with Suggestion
CR Complies with Request
AO Acknowledges Only

Figure 1: Proportions of code categories in face-to-face (squares) and computer-mediated interactions (asterisks) in the original study

the screen size in the MTV condition did not influence the proportions of functions in the 4-line and 18-line conditions. The results in both those conditions are nearly identical. Second, similar differences are obtained between face-to-face and computer-mediated conditions in both corpora. For example, all of the computer-mediated interactions produced suggestions at a proportion of approximately .3, while the face-to-face interactions produced suggestions at closer to half that frequency. Similar patterns of difference between face-to-face and computer-
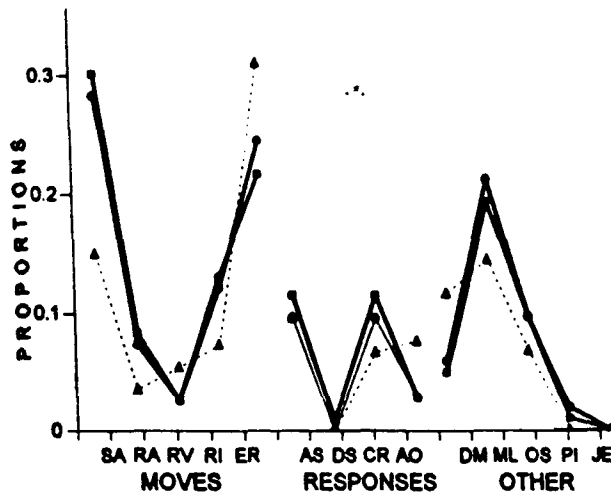
mediated conditions occur in both corpora for the 3 types of requests in the coding system, too.

We anticipated an increase in discourse management functions due to the complexity of the task, and the increase in metalanguage from .05 to .15 in the face-to-face conditions suggests that the more complex task pressured participants to engage in more explicit management strategies. In the computer-mediated interactions, the proportion of functions coded as metalanguage also increases with the complexity of the task, though not as much. The greater proportion of discourse markers in the computer-mediated interactions also reflects an increase in discourse management activity for the more complex task.

The failure to observe an increase in the proportion of utterances coded as "Orients Suggestion" in the MTV interactions is probably a result of the emergence of a turn strategy not observed in the interactions with simpler decision-making tasks. Specifically, while all of the computer-mediated interactions in the initial study and many of the computer-mediated interactions in the MTV study

242

consisted of relatively short turns, some of the latter display a strategy of employing long turns in which participants encode routine functions for several decisions in the same turn, as in (4).

(4) Best Female Video Either we could have Celine Dione's song I'ts all coming back to me or the other one that was in that movie up close and personal. Aany of the clips with her in them would be good. Toni Braxton with that song...gosh I can't think of any of the names of anybody's songs. And show the same clip as before. What about jewel. Who will save your soul. Personally I think she should win we could use the clip of her playing the guitar in the bathroom. We need one more female singer. Did we pick who should present the award? I think Bush should play after the award.

These more parallel management strategies can reduce the number of orientations if a single orientation can hold for several suggestions and a single agreement can accept them all. Of course, this is exactly what happens when participants provide a list of suggestions in a short turn, too. Therefore, the parallel strategy is a minor modification of the decision routine, but it may influence the proportions of routine functions by reducing the number of orientations and agreements.

In fact, the proportions of utterances coded as "Agrees with Suggestion" and "Complies with Request" are lower in the computer-mediated MTV interactions than in the computer-mediated interactions of the initial corpus. Though these proportions are still slightly higher than those in the face-to-face MTV condition, preserving the pattern observed in the initial corpus, the differences are smaller. These differences are reflected even more dramatically if we compare the ratios of suggestions to agreements in the MTV corpus. At approximately 1.5, the ratio of suggestions to agreements in the face-to-face condition of the MTV study resembles the ratio in the face-to-face condition of the earlier study (1.64). Similarly, the ratio of suggestions to agreements in the computer-mediated interactions of the original study is 1.71. In contrast, the ratios of suggestions to agreements in the 4- and 18-line conditions of the MTV corpus are much larger, both at approximately 2.5. We believe that

much of the difference observed is the result of longer turns employing parallel decision management in the MTV corpus.

These results raise the question of the extent to which the interactions conform to a model of the decision routine we have described. The measure developed in Condon et al. (1997) begins by combining the 3 code annotations as a triple and treating those triples as the output of a probabilistic source. Then 0-, 1st- and 2nd-order Markov analyses are performed on the resulting sequences of triples. While the 0-order analyses simply give the proportions of each triple in the interactions, the 1st-order analyses make it possible to examine adjacent pairs of triples to determine the probability that a particular combination of functions will be followed by another particular combination of functions. Similarly, the 2nd-order analyses examine sequences of 3 utterances.

Figure 3: A More Complex Decision Routine Based on Frequency Analyses

Examination of the 2nd-order analyses in the original study revealed that all of the 7 most frequent sequences of 3 utterances trace a path in the model in Figure 3. Using the model in Figure 3, we then calculated the proportions of 0-, 1st- and 2nd-order sequences that trace a path through the model. Of course, the 0-order frequencies simply provide the proportions of utterances that are coded as

Discourse Modality

Markov Order	Oral	Electronic
0 (Single Function)	.34 (.09)	.53 (.13)
1 (Sequence of Two)	.16 (.06)	.32 (.13)
2 (Sequence of Three)	.07 (.04)	.21 (.11)

Table 1: Proportions of Utterance Events Averaged Per Discourse (Standard Deviations in Parentheses) that Conform to the Model in Figure 3 from the Original Corpus

either orientations, suggestions or agreements, but the 1st- and 2nd-order analyses make it possible to examine the extent to which pairs and sequences of 3 utterances conform to the model in Figure 3. Table 1 presents the results of obtaining the measure just described from the initial corpus of face-to-face and computer-mediated interactions. The proportions therefore reflect the average (and standard deviation) per discourse of events that conform to a sequence of routine continuations in Figure 3.

Since conforming to the model is less and less likely as more functions are linked in sequence, it is not surprising that the proportions decrease as the order of the Markov analysis increases. Still, it is encouraging that the proportions of routine continuations in the 1st-

order analyses are approximately equal to the proportions of suggestions in the two types of interactions, since the latter provide an estimate of the number of opportunities to engage in the routine.

Table 2 presents the results of computing the same analyses on the face-to-face, 4-line, 10-line, and 18-line computer-mediated interactions in the MTV corpus. The 0-order results are much the same for both corpora with about 1/3 of the utterances in face-to-face interactions functioning in the decision routine compared to ½ in the computer-mediated interactions. Similarly, proportions of utterance pairs that conform to the routine remain fairly close to the proportions of suggestions in each condition. Screen size appears to have no effect on the results obtained with this measure.

Conclusions

The results are promising both as evidence for our theory of routines and as an initial attempt to devise a measure of conformity to routines. In particular, the fact that an additional corpus with a more complex task has provided measures which are very similar to those obtained in the initial corpus increases our confidence that these methods are tapping into some stable phenomena. Moreover, the similarities of the conformity measures in Tables 1 and 2 occur in spite of the emergence

Discourse Modality

Markov Order	Oral	4-line	10-line	18-line
0 (Single Function)	.29 (.07)	.50 (.12)	.48 (.11)	.45 (.11)
1 (Sequence of Two)	.11 (.05)	.27 (.10)	.25 (.10)	.21 (.11)
2 (Sequence of Three)	.04 (.03)	.17 (.10)	.14 (.08)	.12 (.10)

Table 2: Proportions of Utterance Events Averaged Per Discourse (Standard Deviations in Parentheses) that Conform to the Model in Figure 3 from the MTV Corpus

of new computer-mediated discourse management strategies in which long turns encode decision sequences in parallel. Though these strategies seem to have a strong effect on the ratio of suggestions to agreements in the computer-mediated interactions of the MTV corpus, the conformity measures are still quite similar to the measures obtained in the computer-mediated interactions of the initial study.

The MTV data also confirm the result obtained in the original study that computer-mediated interactions rely more heavily on routines than face-to-face interactions. The much higher conformity measures for all three Markov orders provide clear evidence for this claim with respect to the decision routine. Moreover, a comparison of Figures 1 and 2 shows that the computer-mediated interactions have higher proportions of requests, especially requests for information. If these proportions are indicative of the extent to which request routines are relied on in the interactions, then these data also support the claim that computer-mediated interactions rely on discourse routines more than face-to-face interactions. Given our claims about the effectiveness of discourse routines, it makes sense that participants in an unfamiliar communication environment will employ their most efficient strategies.

The conformity measure that has been devised does not make use of all the information available in the Markov analyses, and we continue to experiment with different measures. It seems clear that Markov analyses can provide sensitive measures that will be useful for identifying differences between interactions and for measuring the effects of experimental factors on interactions.

References

Carletta, J.; Dahlback, N.; Reithinger, N.; and Walker, M. 1997. Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.

Cohen, P.R.; Morgan, J.; and Pollack, M., eds. 1990. *Intentions in Communication*. Cambridge, MA: MIT Pr.

Čech, C. and Condon, S. 1998. Message Size Constraints on Discourse Planning in Synchronous Computer-Mediated Communication. *Behavior Research Methods, Instruments, & Computers*, 30, 255-263.

Condon, S. 1986. The Discourse Functions of OK. *Semiotica*, 60: 73-101.

Condon, S., and Čech, C. 1992. Manual for Coding Decision-Making Interactions. Rev. 1995. Unpublished manuscript available at Discourse Resource Initiative website at http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html

Condon, S., and Čech, C. 1996a. Functional Comparison of Face-to-Face and Computer-Mediated Decision-Making Interactions. In Herring, S. (ed.), *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*. Philadelphia: John Benjamin.

Condon, S., and Čech, C. 1996b. Discourse Management in Face-to-Face and Computer-Mediated Decision-Making Interactions. *Electronic Journal of Communication/La Revue Electronique de Communication*, 6, 3.

Condon, S., Čech, C., and Edwards, W. (1997) Discourse routines in decision-making interactions. Paper presented to AAAI Fall Symposium on Communicative Action in Humans and Machines.

Di Eugenio, B.; Jordan, P.; Thomason, R.; and Moore, J. 1997. Reconstructed intentions in collaborative problem solving dialogues. Paper presented to AAAI Fall Symposium on Communicative Action in Humans and Machines.

Grosz, B. and Hirschberg, J. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada (429-432).

Maier, E.; Mast, M.; and Luperfoy, S., eds., Dialogue Processing in Spoken Language Systems, Lecture Notes in Artificial Intelligence. Springer Verlag.

Nakatani, C., Hirschberg, J. and Grosz, B. 1995. Discourse structure in spoken language: Studies on speech corpora. Paper presented to AAAI 1995 Spring Symposium Series: Empirical Methods in Discourse Interpretation and Generation.

Passonneau, R. 1996. Using centering to relax Gricean informational constraints on discourse anaphoric noun phrases. *Language and Speech*, 39(2-3), 229-264.

Schegloff, E. 1986. The Routine as Achievement. *Human Studies*, 9: 111-151.

Walker, M. (1996). Inferring acceptance and rejection in dialog by default rules of inference. *Language and Speech*, 39(2-3), 265-304.