Exploring the Characteristics of Multi-Party Dialogues

Masato Ishizaki
Japan Advanced Institute of Science and Technology
Tatsunokuchi, Noumi, Ishikawa, 923-1292, Japan
masato@jaist.ac.jp

Tsuneaki Kato
NTT Communication Science Labs.
2-4, Hikaridai, Seika, Souraku, Kyoto, 619-0237, Japan
kato@cslab.kecl.ntt.co.jp

Abstract

This paper describes novel results on the characteristics of three-party dialogues by quantitatively comparing them with those of two-party. In previous dialogue research, two-party dialogues are mainly focussed because data collection of multi-party dialogues is difficult and there are very few theories handling them, although research on multi-party dialogues is expected to be of much use in building computer supported collaborative work environments and computer assisted instruction systems. In this paper, firstly we describe our data collection method of multi-party dialogues using a meeting scheduling task, which enables us to compare three-party dialogues with those of two party. Then we quantitively compare these two kinds of dialogues such as the number of characters and turns and patterns of information exchanges. Lastly we show that patterns of information exchanges in speaker alternation and initiative-taking can be used to characterise three-party dialogues.

1 Introduction

Previous research on dialogue has been mostly focussing on two-party human-human dialogue for developing practical human-computer dialogue systems. However, our everyday communicative activities involves not only two-party communicative situations but also those of more than two-party (we call this multi-party). For example, it is not unusual for us to chitchat with more than one friend, or business meetings are usually held among more than two participants.

Recently advances of computer and networking technologies enable us to examine the possibility of using computers to assist effective communication in business meetings. As well as this line of computer assisted communication research, autonomous programs called 'agents', which enable users to effectively use computers for solving problems, have been extensively studied. In this research trend, 'agent' is supposed to be distributed among computers, and how they cooperate for problem solving is one of the most important research topics. Previous studies on two party dialogue can be of some use to the above important computer related communication research, but research on multi-party interaction can contribute more directly to the advances of the above research. Furthermore, research on multi-party dialogue is expected to make us understand the nature of human communication in combination with the previous and ongoing research on two-party dialogue.

The purpose of this paper is to quantitively show the characteristics of multi-party dialogues in comparison with those of two-party using actual dialogue data. In exploring the characteristics of multi-party dialogues to those of two-party, we will concentrate on the following problems.

- **What patterns of information exchanges do conversational participants form?** When abstracting the types of speech acts, in two-party dialogues, the pattern of information exchanges is that the first and second speakers alternately contribute (A-B-A-B ...). But in multi-party dialogues, for example, in three-party dialogues, dialogue does not seem to proceed as A-B-C-A-B-C ..., since this pattern seems to be too inefficient if B tells C what B are told by A, which C will be told the same content twice, and too efficient and strict if A, B and C always initiate new topics in this order, in which they have no

occasions for checking one's understanding.

- **How do conversational participants take initiative?** In business meetings, most of which are of multi-party, chairpersons usually control the flow of information for effective and efficient discussions. Are there any differences between in multi- and two-party dialogues? For example, are there any possibilities if in multi-party dialogues the role of chairpersons emerges from the nature of the dialogues?

These are not only problems in exploring multi-party dialogues. For example, we do not know how conversational participants take turns (when do they start to talk)? Or how and when do conversational participants form small subgroups? However, the two problems we will tackle here are very important issues to building computer systems in that they directly relates to topic management in dialogue processing, which is necessary to correctly process anaphora/ellipsis and effective dialogue control.

In the following, firstly, previous research on multi-party dialogues is surveyed. Secondly, our task domain, data collection method, and basic statistics of the collected data are explained. Thirdly, our dialogue coding scheme, coding results and the resultant patterns of information exchanges for two- and multi-party dialogues are shown. Lastly, the patterns of initiative taking behaviour are discussed.

2 Related Studies

Sugito and Sawaki (1979) analysed three naturally occurring dialogues to characterise language behaviour of Japanese in shopping situations between a shop assistant and two customers. They relate various characteristics of their dialogue data such as the number of utterances, the types of information exchanges and patterns of initiative taking to the stages or phases of shopping like opening, discussions between customers, clarification by a customer with a shop assistant and closing.

Novick and Ward (1993) proposed a computational model to track belief changes of a pilot and an air traffic controller in air traffic control (ATC) communication. ATC might be called multi-party dialogue in terms of the number of conversational participants. An air traffic con-

troller exchanges messages with multiple pilots. But this is a rather special case for multi-party dialogues in that all of ATC communication consists of two-party dialogues between a pilot and an air traffic controller.

Novick et al. (1996) extended 'contribution graph' and how mutual belief is constructed for multi-party dialogues, which was proposed by Clark (1992). They used their extension to analyse an excerpt of a conversation between Nixon and his brain trust involving the Watergate scandal. Clark's contribution graph can be thought of as a reformulation of adjacency pairs and insertion sequences in conversation analysis from the viewpoint that how mutual belief is constructed, and are devoted to the analysis of two-party dialogues. They proposed to include reactions of non-intended listeners as evidence for constructing mutual belief and modify the notation of the contribution graph.

Schegloff (1996) pointed out three research topics of multi-party dialogue from the viewpoint of conversation analysis. The first topic involves recipient design. A speaker builds referential expressions for the intended listener to be easily understood, which is related to next speaker selection. The second concerns reasoning by non-intended listeners. When a speaker praises some conversational participant, the remaining participants can make inferences that the speaker criticises what they do not do or behave like the praised participant. The third is schism, which can be often seen in some parties or teaching classes. For example, when a speaker continue to talk an uninteresting story for hours, party attendees split to start to talk neighbours locally.

Eggins and Slade (1997) analysed naturally-occurring dialogues using systemic grammar framework to characterise various aspects of communication such as how attitude is encoded in dialogues, how people negotiate with, and support for and confront against others, and how people establish group membership.

By and large, on multi-party dialogues, there are very few studies in computational linguistics and there are several or more researches on multi-party dialogue, which analyse only their example dialogues in discourse analysis. But as far as we know, there is no research on quantitatively comparing the characteristics of multi-

party dialogues with those of two-party. Research topics enumerated for conversation analysis are also of interest to computational linguistic research, but obviously we cannot handle all the problems of multi-party dialogues here. This paper will concentrate on the patterns of information exchanges and initiative taking, which are among issues directly related to computer modelling of multi-party dialogues.

3 Data Collection and Basic Statistics

For the purpose of developing distributed autonomous agents working for assisting users with problem solving, we planned and collected two- and three-party dialogues using the task of scheduling meetings. We tried to set up the same problem solving situations for both types of dialogues such as participants' goals, knowledge, gender, age and background education. Our goal is to develop computational applications where agents with equal status solve users' problems by exchanging messages, which is the reason why he did not collect dialogue data between between different status like expert-novice and teacher-pupils.

The experiments were conducted in such a way that for one task, the subjects are given a list of goals (meetings to be scheduled) and some pieces of information about meeting rooms and equipments like overhead projectors, and are instructed to make a meeting schedule for satisfying as many participants' constraints as possible. The data were collected by assigning 3 different problems or task settings to 12 parties, which consist of either two or three subjects, which amounts to 72 dialogues in total. The following conditions were carefully set up to make dialogue subjects as equal as possible.

- Both two- and three-party subjects were constrained to be of the same gender. The same number of dialogues (36 dialogues) were collected for female and male groups.

- The average ages of female and male subjects are 21.0 (S.D. 1.6) and 20.8 (S.D. 2.1) years old. All participants are either a university student or a graduate.

- Subjects were given the same number of goals and information (needless to say,

	# of chars.	# of turns
2-p	92637	3572
3-p	93938	3520

Table 1: Total no. of characters and turns in two- and three-party dialogues

	ANOVA of chars.	ANOVA of turns
2-p	3.57, 0.59, 0.02	0.00, 0.00, 0.00
3-p	2.53, 1.47, 0.43	3.91, 1.72, 1.00

Table 2: ANOVA of characters and turns for three problem settings in two- and three-party dialogues

kinds of goals and information are different for each participant in a group).

In these experiments, dialogues among the subjects were recorded on DAT recorders in non-face-to-face condition, which excludes the effects of non-linguistic behaviour. The average length of all collected dialogues is 473.5 seconds (approximately 7.9 minutes) and the total amounts to 34094 seconds (approximately 9.5 hours).

There are dialogues in which participants mistakenly finished before they did not satisfy all possible constraints. It is very rare that one party did this sort of mistakes for all three task settings assigned to them, however in order to eliminate unknown effects, we exclude all three dialogues if they made mistakes in at least one task setting. For this reason, we limit the target of our analysis to 18 dialogues each for two- and three-party dialogues which do not have such kind of problem (the average length of the target dialogues is 494.2 seconds (approximately 8.2 minutes).

Table 1 shows the number of hiragana characters [1] and turns for each speakers, and its total for two- and three-party dialogues. It illustrates that the total number of characters and turns of three-party dialogues are almost the same as those of two-party, which indicates

---

[1]This paper uses the number of hiragana characters to assess how much speakers talk. One hiragana character approximately corresponds to one mora, which has been used as a phonetic unit in Japanese.

the experimental setup worked as intended between two- and three-party dialogues. Table 2 shows ANOVA of the number of hiragana characters and turns calculated separately for different task settings to examine whether there are differences of the number of characters and turns between speakers. The results indicates that there are statistically no differences at .05 level to the number of characters and turns for different speakers both in two- and three-party dialogues except for one task setting as to the number of turns in three-party dialogues. But this are statistically no differences at .01 level. For the experimental setup, we can understand that our setup generally worked as intended.

4 Patterns of Information Exchanges

4.1 Dialogue Coding

To examine patterns of information exchanges and initiative taking, we classify utterances from the viewpoint of initiation-response and speech act types. This classification is a modification of the DAMSL coding scheme, which comes out of the standardisation workshop on discourse coding scheme (Carletta et al., 1997b), and a coding scheme proposed by Japanese standardisation working group on discourse coding scheme(Ichikawa et al., 1998) adapted to the characteristics of this meeting scheduling task and Japanese. We used two coders to classify utterances in the above 36 dialogues and obtained 70% rough agreement and 55% kappa agreement value. Even in the above discourse coding standardisation groups, they are not at the stage where which agreement value range coding results need to be reliable. In content analysis, they require a kappa value over 0.67 for deriving a tentative conclusion, but in a guideline of medical science, a kappa value $0.41 < \kappa < 0.60$ are judged to be moderate (Carletta et al., 1997a; Landis and Koch, 1977; Krippendorff, 1980). To make the analysis of our dialogue data robust, we analysed both coded dialogues, and obtained similar results. As space is limited, instead of discussing both results, we discuss one result in the following. From the aspect of initiation-response, utterances are examined if they fall into the category of response, which is judged by checking if they can discern cohesive relations between the current and corresponding utterances if ex-

Types of speech act for initiating
Want-propose(WP), Inform(IF), Request(RQ)
Types of speech act for responding
Positive_answer-accept(PA), Negative_answer-reject(NA), Content-answer(CA), Hold(HL)
Types of speech act for both
Question-check(QC), Counter_propose(CP), Meta(MT)

Table 3: Types of speech act for coding two- and three-party dialogues

ist. The corresponding utterances must be ones which are either just before the current or some utterances before the current in the case of embedding, or insertion sequences. If the current utterance is not judged as response, then it falls into the category of initiation.

From speech act types, as in Table 3, utterances are classified into five types each for initiating and responding, two of which are used for both initiating and responding. Bar ('-') inserted categories show adaptation to our task domain and Japanese. For example, in this task domain, expressions of 'want' for using some meeting room are hard to be distinguished from those of 'proposal' in Japanese, and thus these two categories are combined into one category 'want-proposal'.

4.2 Patterns of act sequences by speakers

Table 5 shows the frequency ratio as to the length of act sequences represented by different speakers in two- and three-party dialogues. The act sequences are defined to start from a newly initiating utterance to the one before next newly initiating utterance. Let us examine an excerpt in Table 4 from our dialogue data, where the first column shows a tentative number of utterances, the second is a speaker, the third is an utterance type, and the fourth is English translation of an utterance. In this example, there are two types of act sequences from the first to the fifth utterance (E-S-E-S-E) and from the sixth to the seventh (S-H). Our purpose here is to examine how many of the act sequences consists of two participants' interaction in three-party dialogues. Hence we abstract a speakers' name with the position in a sequence. The speaker in

	2 acts	3 acts	4 acts	5 acts	6 acts
2-p	54.2	21.6	11.8	5.3	2.1
3-p	45.1	26.0	12.2	5.4	2.4

Table 5: Frequency ratio (%) for the number of act sequences in two- and three-party dialogues

3 act sequences	
ABA	ABC
62.7	37.3

4 act sequences			
ABAB	ABCA	ABAC	ABCB
53.2	17.1	16.2	13.5

5 act sequences			
ABABA	ABCAB	ABABC ABACA	Others
36.7	16.3	10.2(each)	26.6

6 act sequences			
ABABAB	ABCACB	ABABAC ABCACA	Others
31.8	18.2	9.1(each)	31.8

Table 6: Frequency ratio (%) of 3 to 6 act sequences in three-party dialogues

the first turn is named A, and the one in the second and third turn are named B and C, respectively.

In both two- and three-party dialogues, the most frequent length of act sequences is that of two speakers. The frequencies decrease as the length of act sequences increases. In two-party dialogues, speaker sequences concern only their length, since there are two speakers to be alternate while in three-party dialogues, more than two length of sequences take various patterns, for example, A-B-A and A-B-C in three act sequences. Table 6 illustrates patterns of speaker sequences and their frequency ratios. In three act sequences, the frequency ratios of A-B-A and A-B-C are 62.7% and 37.3%, respectively, which signifies the dominance of two-party interactions. Likewise, in four, five and six act sequences, two-party interactions are dominant, 53.2%, 36.7% and 31.8%, both of which are far more frequent than theoretical expected frequencies (25%, 12.5 and 6.3%). In three-party dialogues, two-party interactions amounts to 70.6% (45.1%+26.0%×62.7%+12.2%×53.2%+ 5.4% × 36.7% + 2.4% × 31.8% = 70.6%) against total percentage 91.1% from two to six act sequences (if extrapolating this number to total 100% is allowable, 77.5% of the total interactions are expected to be of two-party). **The conclusion here is that two-party interactions are dominant in three-party dialogues.** This conclusion holds for our meeting scheduling dialogue data, but intuitively its applicability to other domains seems to be promising, which should obviously need further work.

4.3 Patterns of initiative taking

The concept 'initiative' is defined by Whittaker and Stenton (Whittaker and Stenton, 1988) using a classification of utterance types assertions, commands, questions and prompts. The initiative was used to analyse behaviour of anaphoric expressions in (Walker and Whittaker, 1990).

The algorithm to track the initiative was proposed by Chu-Carroll and Brown (1997). The relationship between the initiative and efficiency of task-oriented dialogues was empirically and analytically examined in (Ishizaki, 1997). By their definition, a conversational participant has the initiative when she makes some utterance except for responses to partner's utterance. The reason for this exception is that an utterance following partner's utterance should be thought of as the one elicited by the previous speaker rather than directing a conversation in their own right. A participant does not have the initiative (or partner has the initiative) when she uses a prompt to partner, since she clearly abdicates her opportunity for expressing some propositional content.

Table 7 and 8 show the frequency ratios of who takes the initiative and χ^2 value calculated from the frequencies for two- and three-party dialogues. In two-party dialogues, based on its χ^2 values, the initiative is not equally distributed between speakers in 5 out of 18 dialogues at .05 rejection level. In three-party dialogues, this occurs in 10 out of 18 dialogues, which signifies the emergence of an initiative-taker or a chairperson. To examine the roles of the participants in detail, the differences of the participants' behaviour between two- and three party informa-

#	Sp	Type	Utterance
1	E	WP	Well, I want to plan my group's three-hour meeting after a two-hour meeting with Ms. S's group.
2	S	QC	After the meeting?
3	E	PA	Yes.
4	S	PA	Right.
5	E	PA	Right.
6	S	QC	What meetings do you want to plan, Ms. H?
7	H	CA	I want to schedule our group's meeting for two hours.

Table 4: An excerpt from the meeting scheduling dialogues

E	62.5	55.3	56.0	74.3	70.8	57.1	56.2	50.0	51.7	75.0	61.8	60.9	48.1	53.8	46.9	76.6	49.0	64.0
S	37.5	44.7	44.0	25.7	29.2	42.9	43.8	50.0	48.3	25.0	38.2	39.1	51.9	46.2	53.1	23.4	51.0	36.0
χ^2	3.00	.53	.72	8.26	11.2	.86	.25	.00	.03	18.0	3.07	3.26	.07	.45	.18	13.3	.02	3.92

Table 7: Frequency ratio (%)of initiative-taking and χ^2 values of the frequencies between different speakers in two-party dialogues

tion exchanges in Table 9. The table shows the comparison between two and three speaker interactions in three-party dialogues as to as who takes the initiative in 3 to 6 act sequences. From this table, we can observe the tendency that E takes the initiative more frequently than S and H for all three problem settings in two-party interaction, and two of three settings in three-party interaction. S has a tendency to take more initiatives in two-party interaction than that in three-party. H's initiative taking behaviour is the other way around to S's. Comparing with S's and H's initiative taking patterns, E can be said to take the initiative constantly irrespective of the number of party in interaction.

The conclusion here is that initiative-taking behaviour is more clearly observed in three-party dialogues than those in two-party dialogues. Detailed analysis of the participants' behaviour indicates that there might be differences when the participants take the initiative, which are characterised by the number of participants in interaction.

5 Conclusion and Further Work

This paper empirically describes the important characteristics of three-party dialogues by analysing the dialogue data collected in the task of meeting scheduling domain. The characteristics we found here are (1) two-party interactions are dominant in three-party dialogues, and (2) the behaviour of the initiative-taking

	S	E	H
2-p	39.1, 33.0, 31.1	39.1, 45.4, 43.2	21.8, 21.6, 25.7
3-p	30.9, 21.9, 27.0	40.5, 35.9, 32.4	28.6, 42.2, 40.6

Table 9: Frequency ratio (%) of initiative-taking for 3 to 6 act sequences between two- and three-party interaction in three-party dialogues (Three numbers in a box are for three problem settings, respectively.)

is emerged more in three-party dialogues than in those of two-party. We will take our findings into account in designing a protocol which enables distributed agents to communicate and prove its utility by building computer system applications in the near future.

References

J. Carletta, A. Isard., S. Isard, J. Kowtko, A. Newlands, G.Doherty-Sneddon, and A. H. Anderson. 1997a. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–32.

J. Carletta, N. Dahlbäck, N. Reithinger, and M. A. Walker. 1997b. Standards for dialogue coding in natural language processing. Technical report. Dagstuhl-Seminar-Report: 167.

J. Chu-Carroll and M. K. Brown. 1997. Tracking initiative in collaborative dialogue interactions. In *Proceedings of the Thirty-fifth Annual Meeting of the Association for Computational Linguistics and the Eighth Confer-*

E	26.2	28.1	35.8	13.8	18.5	9.2	14.3	45.8	30.8	51.2	30.4	34.0	39.3	14.5	7.4	56.8	10.0	54.5
S	57.1	45.3	45.3	34.5	38.9	38.5	25.7	25.0	21.2	34.1	46.4	40.4	46.4	54.5	70.4	34.1	42.5	36.4
H	16.7	26.6	18.9	51.7	42.6	52.3	60.0	29.2	48.1	14.6	23.2	25.5	14.3	30.9	22.2	9.1	47.5	9.1
χ^2	11.3	4.2	5.70	6.28	5.44	18.9	11.9	3.50	5.81	8.24	4.75	1.57	4.79	13.3	17.6	15.0	8.19	13.8

Table 8: Frequency ratio (%) of initiative-taking and χ^2 values of the frequencies among different speakers in three-party dialogues

ence of of the European Chapter of the Association for Computational Linguistics, pages 262–270.

H. H. Clark. 1992. Arenas of Language Use. The University of Chicago Press and Center for the Study of Language and Information.

S. Eggins and D. Slade. 1997. Analyzing Casual Conversation. Cassell.

A. Ichikawa, M. Araki, Y. Horiuchi, M. Ishizaki, S. Itabashi, T. Ito, H. Kashioka, K. Kato, H. Kikuchi, H. Koiso, T. Kumagai, A. Kurematsu, K. Mackawa, K. Murakami, S. Nakazato, M. Tamoto, S. Tutiya, Y. Yamashita, and T. Yoshimura. 1998. Standardising annotation schemes for japanese discourse. In Proceedings of the First International Conference on Language Resources and Evaluation.

M. Ishizaki. 1997. Mixed-Initiative Natural Language Dialogue with Variable Communicative Modes. Ph.D. thesis, The Centre for Cognitive Science and The Department of Artificial Intelligence, The University of Edinburgh.

K. Krippendorff. 1980. Content Analysis: An Introduction to its Methodology. Sage Publications.

J. R. Landis and G. G. Koch. 1977. The measurement of observer agreement for categorial data. Biometrics, 33:159–174.

D. G. Novick and K. Ward. 1993. Mutual beliefs of multiple conversants: A computational model of collaboration in air traffic control. In Proceedings of the Eleventh National Conference on Artificial Intelligence, pages 196–201.

D. G. Novick, L. Walton, and K. Ward. 1996. Contribution graphs in multiparty discourse. In Proceedings of International Symposium on Spoken Dialogue, pages 53–56.

E. A. Schegloff. 1996. Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In Eduard H. Hovy and Donia R. Scott, edi-

tors, Computational and Conversational Discourse, pages 3–35. Springer-Verlag.

S. Sugito and M. Sawaki. 1979. Gengo koudo no kijutsu (description of language behaviour in shopping situations). In Fujio Minami, editor, Gengo to Koudo (Language and Behaviour), pages 271–319. Taishukan Shoten. (in Japanese).

M. A. Walker and S. Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segment. In Proceedings of the Twenty-eighth Annual Meeting of the Association for Computational Linguistics, pages 70–78.

S. Whittaker and P. Stenton. 1988. Cues and control in expert-client dialogues. In Proceedings of the Twenty-sixth Annual Meeting of the Association for Computational Linguistics, pages 123–130.