

Activity-based dialogue analysis as evaluation method

Bilyana Martinovski

Ashish Vaswani

Institute for Creative technology

University of Southern California

martinovski@ict.usc.edu

vaswani@ict.usc.edu

Abstract

This paper uses linguistic activity-based dialogue analysis in order to characterize, evaluate and compare activities. We find that human-human and human-machine interaction via radio are equally efficient but offer different styles of learning and instruction.

1. Introduction

How can we recognize speech activities? How does the language and media determine and reflect the speech activity? Which descriptive features are more basic and what story do they tell us about the activity? In this paper we present activity-based dialogue analysis of radio dialogue during Call for Fire (CFF) training. The quantitative analysis is meant to describe the activity but also to give basis for comparisons with other speech activities. We compare instructive radio talk between humans (HH) and between humans and a machine (HM).

2. Method

Studies of automatic systems of communication typically involve a post hoc evaluation by the users in the form of interviews/questionnaires or scored on performance efficiency. In this study, we suggest to observe dialogue styles in order to characterize, compare and evaluate the relation between activities and between human-human and human-machine interaction.

Most studies on language style have concentrated on written language. Biber [1] analyses speech but studies mostly lexical and grammatical features. Allwood [2] and Martinovski [3], for instance, include features specific to spoken language, such-as feedback, turn-taking, pause, etc. The method we adopt here, namely activity-based dialogue analysis [4] involves qualitative and quantitative description of the activity.

2.1. Data and Activity

The data consist of thirteen (13) audio recordings of CFF training sessions, which are part of the ICT Spoken Language Corpus, USC. Seven of the recordings represent radio interaction between humans and six represent interaction between human and a machine, which simulates a human. Dialogue involved in interactive accomplishment of a task, such as CFF dialogues, can be seen as a collaborative activity with set moves, which correspond to tasks. The dialogue involves two participants. The forward observer (FO) is one (or more) soldier who is close to the target. He or they are supposed to observe

and identify the target and to communicate what is observed and identified to the fire direction center (FDC). The FDC is located further away from the target and closer to the guns, which are far from the target. The FDC assists and directs the actions of the FO and the guns. It also assists and directs the communication between FO and guns, among FOs and among guns. The goal is to fulfill a mission; most often this means to destroy a target.

2.2. Units and Coding Categories

A turn is identical to a radio transmission. Thus one may have turns followed by the same speaker distinguished by the switching off of the radio for more than one second. A qualitative analysis of the data resulted in three (3) coding categories and twenty five (25) subcategories:

Dialogue Moves: identification (ID), target location (TL) and description (TD), message to observer (MTO), engagement, fire, method of fire and control (MFC), shot, splash, rounds complete (RC), check, adjustment, observer coordinates (OBCO), end of mission (EM), intelligence report (INTEL), situation report (SITREP);

Dialogue Regulators (or Keywords): over, out, roger, standby;

Grounding Acts: confirmation, prompt, preparation, correction, say again.

The moves are specific to the activity (see appendix, also [5]). In short, the process involves a preparatory stage of identifications of participants (ID), an initial stage of suggestions/requests for action by the FO (e.g moves TL, TD, OBCO, SITREP, see appendix), which are considered by the FDC (e.g. INTEL, SITREP) who then (third stage) informs the FO of the actions of the guns (which may follow the FO's requests or may not, e.g moves MTO, MFC, SHOT, SPLASH, see appendix); finally the FO reports the end results of the collaborative mission (e.g. EM).

By confirmations we mean mainly feedback giving repetitions of the previous utterance; prompts are initiation eliciting expression such as 'go ahead, give obco'; preparations are acts dedicated to preparing for a move, for ex. 'be prepared for intel'; corrections initiate repair of own or other speech; 'say again' is an other-repair initiation. The dialogue regulators listed above are part of the feedback structure of the military speech genre and indicate end of transmission, end of move, confirmation and readiness for initiation of move, respectively.

3. Quantitative analysis

In the following section we observe what story quantitative language analysis has to tell us about the activity style and functions as a whole, in relation to other activities and in relation between HH and HM interaction within one activity. We use simple measures and ranks of parts of speech, moves, acts, turns, pauses as well as more complex measures of liveliness and caution. The HM data are more than the HH.

3.1. Characteristics of the dialogue style

First we will observe the activity as a whole and in comparison to other activities.

3.1.1. Parts of speech

By checking the parts of speech we can notice that this activity is greatly nominalized and that it is oriented towards precision, because the most frequent parts of speech are nouns and cardinals. We can conclude that many utterances do not consist of grammatical sentences but of listing of numbers, verbs as nouns and nouns as names. Most of the nouns appear without determiners.

Table 1. Parts of speech (POS) ranks in numbers.

POS	Num	POS	Num
noun	2626	pronoun	154
cardinals	2568	adverb	121
preposition	1227	'to'	102
verb	774	coordinat-	54
determiner	570	ing con-	
adjective	529	junction	
hesitation	194	modality	18
sounds		existential	5
Total			8951

Many moves, (e.g. orders, information, and confirmations), are given in an elliptic form and counted as nouns, f.ex. 'fire', 'splash', 'shot' etc. The pronouns are more rare than the hesitation sounds, which indicates the need of precise identification of speakers.

3.1.2. Move and acts

The most frequent moves and acts in the activity are identification, confirmation and dialogue regulators. Identification is used to identify the speaker and the addressee. Confirmations are repetitions of already given information or order which purpose is establishment of common ground. The keywords' function is to regulate the dialogue i.e. they are part of the grounding efforts. Thus the most dominant function of the activity is exercise of security and reliability of information including credibility of participants of which the establishment of common ground is an essential part. The acts contributing to grounding are in bold in Table 2.

Table 2. Moves and acts ranks for all CCF data.

rank	move and num of words	rank	move and num of words
1	id 1984	12	splash 164
2	confirm 1462	13	shot 144
3	keyword 1306	14	rc 93
4	sitrep 617	15	obco 91
5	tl 590	16	engage 88
6	intel 560	17	adjustment 52
7	mto 408	18	say again 52
8	prompt 383	19	mfc 46
9	em 376	20	correction 39
10	td 230	21	prep 33
11	fire 216	22	check 10
		23	standby 6

36.7% of all words are dedicated to grounding. If we include in the grounding the identification 'ritual' of the participants it will dominate the activity. 13.2% of the words are involved in intelligence and situation reports. These are given usually by the FDC and situation reports, which could be given by both participants are also of high frequency higher than moves which realize the desired missions such as shot and splash. This means that there is more effort spent in coordination and preparation of action than on action. We observe though that there are much more prompts than preparations which compensates for the reports. This is explained by the fact that the FDC uses the reports as a medium for explicit instruction of trainees and after such instructions often gives prompts for action.

3.1.3. Cross-activity and cross-linguistics comparison

Can statistical measures indicate something about the functions of the utterances or the predominant type of communicative acts in an activity? Although the correspondence between the function of utterances/activity and the overall impression of the language style has yet to be studied we offer here a short evaluation and characterization of the studied activity by comparing the data across-activities and languages using two complex measures, namely measures of liveliness [6] and caution [7]. The measure of interactive liveliness is created by comparison of simple measures such as stress (StressP-Tok – the percent of stressed words as tokens in relation to the total amount of words as tokens), overlap (ToOVP-tok – percent of overlapped words as tokens in relation to the total amount of words as tokens), mean length utterance (MLU - the mean of the number of tokens in an utterance) and pauses (PausPTok - the percent of pauses as tokens in relation to the total amount of words as tokens). In the caution measure, the more pauses, stressed words, own communication management (OCM e.g. hesitations, self cut-offs) [8] and numerals (cardinals) there is the more cautious the activity is. The applied formulas (see also Appendix for explanation of measures) are:

Interactive liveliness

$$(2. \text{ToOVPTok}) + \text{OvePUtt} + (2. \text{StressPTok}) - (2. \text{PausPTok}) - \text{MLU}$$

Total number of tokens

Interactive caution

$$(2. \text{Pauses}) + \text{Stress} + (2. \text{OCM}) + (2. \text{FB}) + \text{Numerals} - (2. \text{Pronouns}) - (2. \text{Overlap})$$

Total number of tokens

Table 3. Cross-activity comparison of liveliness and caution measures applied to American English and Swedish (<http://www.qualitative-research.net/fqs-texte/3-00/3-00allwoodetal-e.htm>).

Complex measure	Average number of speakers	Liveliness	Caution
All	9	30.1	
Sermon	5	-82.6	
Auction	7	-70.3	
CFF radio training	2	-0.002	1.4
Courts	6	9.4	-0.1
Tasking	3	11.3	
Meeting	11	33.9	
Dinner	9	91.2	-0.7

The US CCF army training is far more cautious than Swedish courtroom interrogation and formal dinner and by implication may be interpreted as a more precision-oriented type of verbal activity. Court examinations are more cautious and precise than dinner talk. Caution is related also to the required commitment to the acts of speech and the activity. Thus caution and precision in references (the large number of numerals, names and nouns) is a reflection of this activity's purpose, namely a secure transfer of precise sensitive information in stressful conditions. This purpose calls for even more caution and precision than the delivery of justice.

With regard to the definition of liveliness the sermon is one of the least lively, it is after all a monologue. The auction, which consists mainly of offers and confirmations, the task-oriented, the court examination and the CCF training activity are much less lively than the dinner conversations. The talk-over-dinner is the most lively activity. Thus we may say that if we imagine the activities as a continuum on a scale of liveliness the auction represents one pole and the dinner the other pole. The CCF training talk style appears to be in the middle of this scale with similarity to the court talk.

We may rather speculatively observe that question-answer fixed activities are less lively but more cautious. As expressed by the complex measures above, the CCF training has much less overlap than court and dinner talk, mainly due to the radio transmission style of interaction and transcription. However, the radio training has a tendency for more stressed tokens and more pauses than the dinner talk, as shown in Table 4.

Table 4. Number of tokens, utterances, overlaps, pauses, and stressed words in the Swedish dinner and court data and American English CFF radio training.

Activity	Word	Type	Turn
Dinner	30738	3971	2365
Court	33409	3667	2045
CCF	8951	630	896
Activity	OvePUtt	StrPTok	PauPTok
Dinner	1998 (84.5 %)	142 (0.5%)	1637 (5.3%)
Court	637 (28.7%)	1127 (3.8%)	2827 (8.5%)
CCF	0	10 (1.1%)	656 (7.3%)

Note:

OvePUtt – percent of overlapped words as tokens in relation to the total number of utterances in a group.

StrPTok – percent of emphatic, stressed words as tokens in relation to the total amount of words as tokens in a group.

PauPTok - percent of pauses in relation to the total amount of words as tokens in a group.

In fact, the pauses in the CCF data have much longer duration than the pauses in all the other activities we reflected on. The average duration of a pause in the CFF radio talk is 1 sec, which is much more than the average conversation pause. Sometimes the pause takes more than 2 minutes. If we involve the pause duration in complex measures we might get different results: the CFF training would appear even more cautious and much less lively. Such a result would reflect the fact the CFF activity is mediated by radio whereas the rest are face-to-face. The emphatics may indicate the instruction pathos but may also reflect the higher level of stress (also, impatience, irritation etc.).

3.2. Comparing human-human and human-machine interaction style

The CFF radio training data consist of human-human talk and human-machine talk, the distinction between which we disregarded in the general characterization of the activity as a whole. We observe how the HH and HM interaction styles within the same activity relate.

3.2.1. Moves and acts

The HH CFF radio talk has 615 unique words which is 8.6% of all HH, whereas the HM has 111 i.e. 6.3% of all HM words. Thus despite the greater amount of reports in HH talk (see Table 5), the vocabulary of the HM is not much different but f. ex. the hesitation sounds are more frequent in and thus characteristic of HH talk.

Table 5. Moves and acts ranks HH/HM talk.

N	Human-human	Human-machine
1	ID 1355 18.8	ID 629 35.7
2	confirm 1197 16.7	confirm 265 15.1
3	keyword 1161 16.2	TL170 9.7
4	SITREP 592 8.2	MTO 162 9.2
5	INTEL 560 7.8	keyword 145 8.2
6	TL 420 5.8	EM 85 4.8
7	prompt 376 5.2	TD 51 2.9
8	EM 291 4.1	fire 41 2.3
9	MTO 246 3.4	RC 39
10	TD 179 2.5	engage 35
11	fire 175 2.4	shot 26 1.5
12	splash 163 2.3	SITREP 25 1.4
13	shot 118 1.6	OBCO 23 1.3
14	OBCO 68 1.0	say again 20 1.1
15	RC 54	MFC 18
16	engage 53	correction 14 0.8
17	adjustment 52 0.7	prompt 7 0.4
18	prep 29 0.4	prep 4 0.2
19	say again 20 0.3	splash 1
20	MFC 32	check 0
21	correction 25 0.4	INTEL 0
22	check 10 0.1	adjustment 0
23	standby 6 0.08	standby 0
sum	7191	1760

The grounding and regulatory moves in HH are 55.2 % of all moves vs. 27.3 % in HM. However, this does not mean that HM is more efficient, both HH and HM have almost the comparable amount of EM reports, 4.1 and 4.8 respectively and similar amount of Shot moves, 1.6 and 1.5 respectively. MTO is more frequent in HM, which means that the radiobot FDC is more active than the human instructor FDC. The human FDC instructs with explanations in INTEL and SITREP while the machine FDC trains by action, with much less explanations or discussions. This is most obvious if we compare the frequency of Intel reports in the HH and the HM data: 7.8 % vs. 0 %, respectively. Also, the HH shows more diversity of moves, e.g. it has moves such as check and adjustment, which were absent or less frequent in HM.

The HH uses meta-grounding acts such as prompts to stimulate the trainee to perform an action whereas the HM simply performs the move.

3.2.2. Parts of Speech

With regards to parts of speech realization, pronouns, hesitation sounds and modals are more frequent in HH talk and thus characterize the HH talk.

Table 6. Parts of speech (POS) ranks in number and in percent of own group for HH and HM talk.

POS ranks	Human-human	Human-machine
noun	2106 (29.3%)	520 (29.6%)
cardinal	1905(26.5%)	663 (37.7%)
preposition	1007(14.0%)	220 (12.5%)
verb	659 (9.2%)	115 (6.5%)
determiner	490 (6.8%)	80 (4.6%)
adjective	411 (5.7%)	118 (6.7%)
hesitation sounds	190 (2.6%)	4 (0.2%)
pronoun	149 (2.1%)	5 (0.3%)

adverb	100 (1.4%)	21 (1.2%)
'to'	90 (1.3%)	12 (0.7%)
coordinating conjunction	52 (0.7%)	2 (0.1%)
modality	18 (0.3%)	0
existential	5 (0.1%)	0
present verb	301 (45.7%)	91 (79.1%)
past verb	127 (19.3%)	13 (11.3%)
Total	7191	1760

The cardinals dominate the HM talk, which indicates that there is greater precision in the HM performance due also to more frequent TL and OBCO moves (see Table 5 above). There is more variation in verb tense forms in the HH talk although and the HM talk is more 'now' oriented, which is partly due to infrequent if any reflection and discussion of events.

In the HH radio talk there are preferences for particular nominalized formulations (in bold in Example 1 below) with indefinite determiners in front of ID names, confirmative regulators, SITREPs, prompts, or target descriptions (TD):

Example 1.

g91: **a roger** over , KEYWORDS
s19: ok . CONFIRM.
g91: **a steel one niner** , ID
this is gator niner one , ID
ah end of mission , EM
break , KEYWORDS
a we'll engage that mission again , SITREP
over . KEYWORDS
s19: roger , KEYWORDS
a go ahead and just send your correction . -
PROMPT over . KEYWORDS
g91: **a roger** , KEYWORDS a distance of a seven
one zero , a direction of five six four zero , TL
a one zsu in open TD over . KEYWORDS
s19: roger , KEYWORDS

These are specific for the HH radio talk and express the ritualistic repetitive fixed format of the sub-activities, which sound like known moves in a game rather than spontaneous speech.

4. Conclusion

In this short paper, we use an activity-based method for analysis and evaluation of dialogue. We start with qualitative analysis in order to produce coding categories, which we then use in a quantitative linguistic study. We find recognizable and specific characteristics of activities, which could be even richer in an extended version of the study. In comparison to other activities across languages, the CFF radio training is livelier only in comparison to sermons and auctions but more cautious compare to courtroom examinations and dinners. Question-answer fixed activities appear less lively but more cautious. The complex measures of liveliness and caution would be improved if information about the duration of pauses is included. Furthermore, the data show that the HH interactive training relies on explicit explanations, coordination and meta-grounding acts whereas the HM relies on repetition of pre-recorded actions with not much explicit explanation, less speech management and coordination, but also less modal expression and hesitation sounds. Thus

both training modes are equally efficient but differ in style of instruction and learning.

Future work on the topic will involve other speech-feature analysis, such as intonation in relation to function of task moves and grounding.

5. Acknowledgements

"The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred."

6. References

- [1] Biber, D. *Variation Across Speech and Writing*. Cambridge University Press, Cambridge, UK, 1988.
- [2] Allwood, J. "Kommentarer till et samtal: kommunikativ verksamhetsanalys" (Comments to a conversation: activity communication analysis), S. Hellberg, U.-B. Kotsinas, and P. Linell, (Eds), *Svenskans Beskrivning*, (Describing Swedish) Goteborg University, Goteborg, Sweden, 1992.
- [3] Martinovski, B. *Speech and Activity Style*. Monographs in Theoretical Linguistics. Goteborg university, Goteborg, Sweden, 1996.
- [4] Allwood, J. "An activity based approach to pragmatics", Bunt & Black (Eds.) *Approaches To Pragmatics*, 1995.
- [5] Martinovski, B., Roque, A., Vaswani, A., Traum, D., and Kennedy, B. "Interactive Action Via Radio. Translation and Coding Manual", Institute for Creative Technologies, USC, California, USA, 2005.
- [6] Allwood, J. & Hagman, J., "Enkla mått på samtal (Simple measures for dialogues)", Frans Gregersen & Jens Allwood (Eds.), *Four special sessions Spoken Language*, Proceedings of the XIV Conference of Scandinavian Linguistics (pp.3-22). Göteborg University, Department of Linguistics, 1994.
- [7] Martinovski, B. *The Role of Repetitions and Reformulations in Court Proceedings – a Comparison of Sweden and Bulgaria*. Gothenburg Monographs in Linguistics. Göteborg University: Department of Linguistics, 2000.
- [8] Allwood, J., E. Ahlsen, J. Nivre, & S. Larsson, *Own Communication Management. Coding Manual V1.0*. Goteborg, Sweden, 1997.

Appendix

Appendix A Example of a CFF dialogue

Note: FO will be G91 (Gator niner one) and FDC will be S19 (Steel one niner)

Phase I:

G91: steel one niner, this is gator niner one, radio check, over

S19: gator niner one, this is steel one niner, radio check good, over

G91: steel one niner, gator niner one, prepare to copy obco, over

S19: go ahead gator niner one

G91: obco follows, grid one two three five, five six six four, over

S19: roger, I read obco is grid one two three five, five six six four, over

G91: that's a good copy steel one niner, over

S19: roger that gator niner one

Phase II:

G91: steel one niner , this is gator niner one, fire for effect , over

S19: fire for effect out

G91: grid, three five four four , three six four six, attitude one sixhundred over

S19: grid , three five four four , three six four six , attitude one six hundred out

G91: five bmps stationary over

S19: gator niner one , steel one niner ,check that grid over

G91: correction , grid , four five four four , three six four six over

S19: correction , grid , four five four four , three six four six out

Phase III:

S19: uh message to observe:r , mike four rounds , h e , over

G91: mike four round h e out

S19: target number , alpha bravo , zero zero zero one , over

G91: target number, alpha bravo , zero zero zero one , break , direction six one hundred over

S19: direction six one hundred out

S19: shot , rounds complete over

G91: shot , rounds complete out

S19: splash over

G91: splash out

Phase IV:

G91: end of mission, target neutralized break, one BMP destroyed, four neutralized

S19: roger one bmp destroyed, four neutralized

Appendix B Description of measures

MLU average number of words per utterance

PauPTok percent of pauses in relation to the total amount of words as tokens in a group

StrPTok percent of stressed words as tokens in relation to the total amount of word- tokens in a group

OvePUtt percent of overlapped words as tokens in relation to the total number of utterances in a group

Measure of interactive caution:

$$(2.Pauses) + Stress + (2.OCM) + (2.FB) + Numerals - (2.Pronouns) - (2.Overlap)$$

total number of tokens

Measure of interactive caution:

$$(2. ToOVptok) + OvePutt + (2.StressPTok) - (2.PausPTok) - MLU$$

total number of tokens