Coding Structures and Actions with the COSTA Scheme in Medical Conversations

Nan Wang[†‡♯], **Yan Song**[♠], **Fei Xia**[‡]
[†]University of California-Los Angeles, CA, USA
[‡]University of Washington, WA, USA
[♯]Hunan University, Hunan, China
[♠]Tencent AI Lab
nwang3@ucla.edu

Abstract

This paper describes the COSTA scheme for coding structures and actions in conversation. Informed by Conversation Analysis, the scheme introduces an innovative method for marking multi-layer structural organization of conversation and a structure-informed taxonomy of actions. In addition, we create a corpus of naturally occurring medical conversations, containing 318 video-recorded and manually transcribed pediatric consultations. Based on the annotated corpus, we investigate 1) treatment decision-making process in medical conversations, and 2) effects of physician-caregiver communication behaviors on antibiotic over-prescribing. Although the COSTA annotation scheme is developed based on data from the task-specific domain of pediatric consultations, it can be easily extended to apply to more general domains and other languages.

1 Introduction

Conversational understanding has been investigated for long by various fields of study such as philosophy of language (Austin, 1962; Searle, 1969, 1985; Wittgensein, 1953), sociology (Schütz, 1967; Sacks, 1992; Garfinkel, 1967; Goffman, 1983), and artificial intelligence (Grosz and Sidner, 1986; Core and Allen, 1997; Perrault. and Allen, 1980; Pollack, 1986) .

Conversational structures are at the heart of the inquiry. Drawing from the philosophical and sociological views of conversational understanding (Schütz, 1967; Wittgensein, 1953; Weber, 1991), Conversation Analysis (CA) was developed to study the systematic organization of conversation and answer the question: '*How is conversation made possible?*'(Heritage, 1984; Schegloff, 2007; Sacks et al., 1974). In artificial intelligence, researchers also explored various theories and practices in analyzing conversation structures, based on which intelligent dialog systems can be developed to assist human with various types of tasks (Core and Allen, 1997; Carletta et al., 1997; Grosz and Sidner, 1986; Jurafsky et al., 1997; Stolcke et al., 2000; Mayfield et al., 2014). In medicine, research shows that a thorough understanding of physician-patient communication structure is important for delivering quality health care and achieving optimal health outcomes (Heritage and Maynard, 2006; Zolnierek and Dimatteo, 2009; Stivers, 2007).

Despite the enormous contribution that existing research has made to advance our knowledge in conversational structures and understanding, limitations exist and opportunities stand for future research. For CA, although the theory and practices of analyzing conversational structures and actions exist, there has not been any synthesized scheme to analyze the hierarchical structure of complete conversations; nor is there any corpus in which such information is annotated. In artificial intelligence, although existing studies recognized the role of structures and actions in conversation understanding and developed annotation schemes to code such information, most of them has only implemented structural annotations at a shallow layer. Moreover, due to a lack of appropriate language resources and tools, research on medical communication in clinical setting remains limited.

Motivated by these challenges, we propose COSTA (COnversational STructures and Actions) – a scheme for coding hierarchical structures and actions in conversations, and a corpus of medical conversation with such annotations.
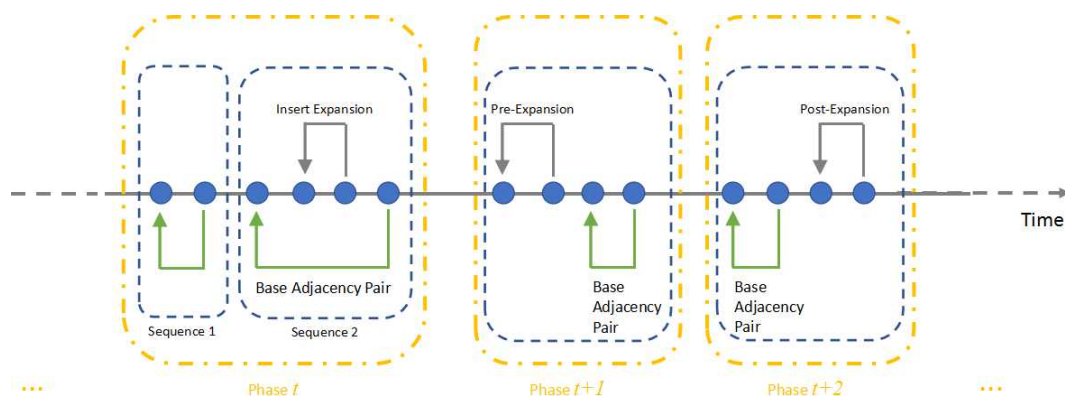
Figure 1: A schematic representation of hierarchical structure of conversation. Blue nodes are turns following a chronological order (the horizontal axis). The arrows link two turns in an adjacency pair. Base adjacency pairs are marked by green arrows; adjacency pairs in sequence expansions are marked by gray arrows. Sequences are marked by blue boxes, and phases are marked by yellow boxes.

2 Conversation Analysis

The COSTA scheme is informed by the sociological theory of conversation analysis (CA). Although CA resembles discourse structure theories such as Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) and that of Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008) in a sense that utterances are considered as structurally organized, what distinguishes CA is that its theory is based on dialogic text rather than monological text (e.g., news articles, academic articles, etc.). Conversation is viewed as organized with 'interaction' orders (Weber, 1991); by contrast, monological text does not take into account recipients' reactions in its immediate context. This means that these two types of discourse are distinctively different and might need to be analyzed with different structural frameworks.

Using naturally occurring conversational data, CA aims to investigate the methods and resources that participants systematically use and rely on to produce intelligible actions and make sense of each other (Heritage, 1984).

Two of the major dimensions of CA involve *sequence organization* and *action formation*. *Sequence organization* addresses questions such as how successive turns are formed up to be 'coherent' with the prior turn, and relatedly, how the overall composition of a conversation gets structured, what those structures are, and how the placement in the overall structure informs the construction and understanding of the talk (Schegloff, 2007). *Action formation* refers to the problem as to how the resources of language, the body, the environment of the interaction, and position in the interaction are fashioned into conformations designed to be and recognizable by the recipient as particular action (Schegloff, 2007).

2.1 Conversational Structures

In CA, structural organization of conversation can be conventionally analyzed at three layers (Schegloff, 2007).

(1) *Turn*: Turns are segmented at each change of speakership. A turn is analyzed in terms of how it is designed to implement some social actions (e.g., a question, a proposal, etc. (Drew, 2014)).

(2) *Sequence organization*: Sequence organization examines how successive turns are formed up to be 'coherent' with the prior turn to accomplish some courses of social actions (e.g., question-answer, proposal-acceptance, greeting-greeting, etc. (Schegloff, 2007)). Relatedly, *adjacency pairs* are the most basic unit of sequence organization (Schegloff, 2007). The idea is that social actions are produced to either initiate a possible sequence of action or to respond to an already initiated action (Stivers, 2014). By initiating a sequence of actions, social actors impose a normative obligation on co-interactants to provide a type-fitted response at the first possible opportunity (Stivers, 2014; Sacks, 1992). Yet, an adjacency pair may, but need not, be expanded, with one or multiple forms of sequence *expansions* (i.e., pre-, insert, and post-expansion) (Schegloff, 2007; Stivers, 2014). Therefore, a cluster of turns in conversation can be analyzed as to whether they form up a coherent sequence with one base adjacency pair and its expansions.

(3) *Overall organization*: A single conversation is viewed as conducted to accomplish some social

77

TID	APP	SL	PS	PR	Speech Text	Action	Outcome
58	1-B	0	P4	D	感冒 了 . (He's) got a cold.		
59	2-B	58		M	嗯 . Ok.		
60	1-B	0	P5	D	吃 点 药 吧 ? Take some oral medicine, ok?	B2	
61	1	60		M	啊 ? Huh?	C0	
62	2	61		D	先 吃 点 药 ，好 吧 ? Take some oral medicine first, ok?		
63	2-B	60		M	哦 ，好 . Oh, ok.		
64	3-B	60		D	嗯 . Ok.		
65	1-B	0	P5	M	开 点 青霉素 给 我们 好 吧? Could you prescribe us some Penicillin?	A1	
66	2-B	65		D	嗯，行 . Ok. Alright.		D1

Table 1: An example of annotated excerpt. TID: Turn ID; PR: Participant Role (M: Mother, D: Doctor); APP: Adjacency Pair Part; SL: Sequence Link; PS: Phase; Action: Conversational Action; Outcome: Prescribing Outcome. Note: Prescribing outcome is annotated at the last turn of a conversation. D1 is included in this table for illustration purpose.

activities, and the social activities can be viewed as involving multiple, normatively ordered sequences of actions (Sacks, 1992; Robinson, 2014; Sacks and Schegloff, 1973). For example, the social activity of telling a trouble to a friend usually involves approaching, arriving at, delivering, working up, and exiting from the trouble (Jefferson, 1988); the activity of dealing with acute medical concerns involves presenting, gathering information about, diagnosing and treating the concern in the American primary care settings (Robinson, 2003, 2014).

Based on the CA theory, we synthesize and formalize the CA analytical practices by developing an annotation scheme of conversational structures which has the following four layers:

(1) Turn: A conversation is segmented at each change of speakership. A turn consists of all its construction units before the speakership changes.
(2) Adjacency Pair: Turns are analyzed as to whether they form up adjacency pairs. An adjacency pair has two parts - a first pair part (FPP) initiates an action, and a second pair part (SPP) responds to a FPP action. For instance, FPP could be a request and SPP is a response to the request.
(3) Sequence of Actions: Turns are also analyzed as to whether they form up a sequence of actions. A sequence is composed of a base adjacency pair and zero or more expansions.
(4) Phase: At a highest level, a conversation may consist of several ordered phases. For instance, a medical conversation may include phases for his-

tory taking, diagnosis, treatment, etc. A phase consists of one or more sequences of actions.

This hierarchical structural organization of conversation is illustrated in Figure 1. In this figure, blue nodes are turns in a conversation in a chronological oder. Yellow boxes represent phases in a conversation; blue boxes represent sequences. Within a sequence, an arrow links two turns in an adjacency pairs - green arrows represent base adjacency pairs, whereas gray arrows represent that the adjacency pairs are expansions of a base adjacency pair.

These concepts can be further illustrated with the examples in Table 1. Table 1 is a short excerpt of a medical conversation in which the physician and the mother are engaged in an activity of dealing with the patient's acute respiratory tract infection symptoms.

Phase: The excerpt contains two phases in a medical conversation: Turns 58-59 belong to a diagnosis phase, in which a diagnosis of the patient's condition is provided and received; Turns 60-66 are part of a treatment phase, in which a treatment recommendation is offered and accepted. Note that a phase can contain multiple sequences. For example, there are two sequences in Turns 60-66, in which two treatment recommendations are offered and received (Turns 60-63 and 64-66, respectively).

Sequence: The example contains three sequence. Two of them, Turns 58-59 and 65-66 each contain only one adjacency pair. The third one, Turns

60-64, contains two adjacency pairs: Turns 60-63-64 is the base adjacency pair;[1] Turns 61-62 is an insert expansion of the bases pair, as the mother and the physicians deal with repairing a hearing problem with the physician's turn (Schegloff et al., 1977).

Adjacency pair and Turn: Each Chinese line in Table 1 is a turn, and they form multiple adjacency pairs. For instance, Turn 65-66 forms an adjacency pair, where the mother initiates a request for a Penicillin prescription in Turn 65 and the physician grants it in Turn 66, thereby fulfilling the expectation set up by the request.

2.2 Conversational Actions

Definition of action has long been of considerable interest to many fields. In CA, the central sense of action is the ascription and assignment of 'a main job' that the turn is doing (*i.e., what the response must deal with in order to count as an adequate next turn; whether the turn fits to the overall contextual environment or not*) (Levinson, 2014).

The structural placement of a turn thus is essential for action recognition and ascription in conversations. First, action ascription is informed by the sequential position of a turn in a local adjacency pair (e.g., question-answer, offer-acceptance). A first pair part (FPP), by projecting a matched second pair part (SPP), maps an action onto the second. Thus, the same utterance might be understood as different actions by virtue of its location. For example, Turn 58 *'He's got a cold.'* in Table 1 is understood as delivering a diagnosis, rather than providing an account, because of its sequential context as being an initiating action in the diagnosis phase, rather than an answer responding to a question (e.g., *'Why is him not here today?'*). In sum, CA views the positioning of an utterance in the ongoing conversation as fundamental to the understanding of its meaning as performing some actions. Social actors rely on their shared knowledge or commensense about the sequential context to make sense of each other. This structure-informed theory about conversational actions thus distinguishes CA from other approaches such as Speech Act Theory, which exclusively focuses on the surface composition of an utterance.

In this study, we use this structure-informed tax-

Item	Number
Total Number of Visits	318
Total Number of Hospitals	5
Total Number of Physicians	9
Total Number of Patients	318
Average length of a visit	4.9 minutes
Total length of the recordings	26 hours

Table 2: Statistics of the raw data. Total number of patients are calculated by those accompanied by caregivers.

onomy of action to identify the conversational actions that are hypothesized to affect the prescribing decision outcome of the medical visits. This will be explained in Section 3.3.

3 Corpus Construction and Annotation Scheme

How do we annotate structures and actions in conversation? In this section, we describe the corpus that we constructed for the study and the annotation procedure of the COSTA scheme.

3.1 Video-recording and Transcription

We created a corpus containing 318 medical conversations between pediatricians and patients/caregivers, collected from five hospitals in China in 2013.

Raw Data: The raw data are video-recordings of the medical conversations. Due to its pediatric setting, the conversations were mostly between physicians and patients' caregivers. We call each conversation (i.e., a recording of a complete medical visit) a *visit*. Table 2 shows raw data statistics.

Transcribing: The video-recordings were transcribed to capture both what was said and how it was said in the conversation. The conversation was segmented into turns at each speakership change in two passes. The first pass transcribed the verbatim words of a turn; the second pass transcribed speech production features (e.g., intonations, overlapping, etc.), as well as non-verbal activities (e.g., nodding, coughing, etc.). Example of the transcript is in the *Speech Text* column in Table 1. Details of the transcribing symbols are described in (Jefferson, 2004).

Five undergraduate students and one graduate student transcribed the data. Each conversation was transcribed by two annotators and verified by a third. The inter-annotator agreement was 91%.[2]

---

[1] While an adjacency pair typically contains two turns, there are exceptions such as the one here, where Turn 64 is a sequence closing third turn (see Section 3.2).

[2] The character error rate was 8.9% when one transcript was treated as the reference and the other as system output.

Ethical Consideration: Research procedures were reviewed and approved by the UCLA IRB (Ref: IRB#13-000748). All identifiable information were removed.

3.2 Structure Annotations

To annotate structures in conversation, we create five attributes: Turn ID (TID), Participant Role (PR), Adjacency Pair Part (APP), Sequential Link (SL), and Phase (PS). The first four are at the turn level, and the last one is at the sequence level.

TID (Turn ID) is a sequential number automatically assigned to a turn, indicating the temporal position of the turn in a conversation.

PR (Participant Role) marks the speakership of a turn, using labels from a pre-defined label set (which is task-specific). For example, in Table 1, Label *D* stands for Doctor, and *M* stands for Mother. The PR label is particularly informative when there are more than two participants.

APP (Adjacency pair part) marks the position of a turn in an adjacency pair and it normally has one of the two values:

- *1* marks an initiating action (FPP).
- *2* marks a responding action (SPP).

This can be illustrated in Table 1, lines 58-59. In addition to 1 and 2, APP can have other values:

- *0* marks a turn occupied by a noticeable silence or some non-verbal activities.
- *3* marks a turn as 'sequence closing third (SCT)'. SCT is in fact a minimal form of post-expansion of an adjacency pair, indicating that no further talk is projected beyond this turn. However, it is ritually used and viewed as part of the base adjacency pair, making it a three-part exception of the adjacency pair (Schegloff, 2007). For example, in Table 1 (lines 60–64), a treatment recommendation is delivered at line 60 and accepted at line 63. This sequence can be considered as completed with the second pair part turn fulfills the expectation of the first pair part. Following this, the physician produces an acknowledgment token *'ok'* at line 64, indicating no further talk projected related to the sequence. This turn is thus marked as 3 in the APP attribute.

Although a sequence is ideally composed of a two-part adjacency pair (the minimal form), it can

be and is usually expanded, and thereby consist of one base adjacency pair and one or more expansions.

To distinguish a base adjacency pair from its expansions, we attach label *B* to the APP value of the turns in the base adjacency pair, such as the pair formed by Turns 58-59.

Given that an adjacency pair can be expanded with other turns (e.g., by an insert expansion) and some adjacency pairs can be incomplete (e.g., a question is not answered), APP labels alone will not be sufficient to indicate which turns form an adjacency pair and which adjacency pairs form a sequence. The *SL* attribute is created to solve this problem.

SL (Sequential link) is a pointer to another turn in the same sequence, indicating the dependency-like relation between two turns. The SL values are set according to the following rules:

- *Rule 1*: In an adjacency pair, the non-FPP (e.g., SPP and SCT) always points to its corresponding FPP. That is, the SL value of an non-FPP turn of an adjacency pair is the TID of its corresponding FPP.
- *Rule 2(a)*: The base adjacency pair in a sequence is like the root of a dependency structure; therefore, the SL of the FPP of the base adjacency pair is set to 0.
- *Rule 2(b)*: If a sequence includes any forms of expansion, the expansion pair 'depends' on the base pair; therefore, the SL value of the FPP of an expansion pair is the TID of the FPP of the base adjacency pair.

To illustrate an example of a sequence with an insert sequence, we can look at Turns 60–64 in Table 1. At Turn 60, the physician initiates a recommendation, which sets up an expectation for the mother's acceptance. However, the mother displays a hearing problem before she finally accepts it at Turn 63. In this sequence, Turns 60 and 63 are FPP and SPP of the base adjacency pair, respectively; Turns 61 and 62 are FPP and SPP of an insert expansion of the base adjacency pair. The SL values of Turns 62-64, 60, and 61 are set according to Rule 1, 2(a), and 2(b), respectively.

Note that although not shown in Table 1, expansion adjacency pair can possibly be further expanded with its own expansions. In such cases, the rules above still apply. As a result, the conversational structure of a sequence is a tree, and

it is very similar to the dependency structure for a sentence: the SL attribute is just like the dependency arc, indicating the dependency of the non-FPP turns on FPP turns and that of the FPP of an expansion pair on the FPP of the base adjacency pair. While we are not using dependency type on the arc, the type can be easily inferred from the APP attribute including label -B.

PS (Phase) indicates the nature of sequence (i.e., what phase a sequence belongs to) in a conversation, and it is marked at the first turn in a sequence.

The labels for PS are task-specific and the ones that we used for this corpus are: *P0*: Consultation opening, *P1*: Problem presentation, *P2*: History taking, *P3*: Physical examination, *P4*: Diagnosis, *P5*: Treatment, *P6*: Addressing additional concerns, *P7*: Consultation closing.

In Table 1, Turn 58 is the start of a sequence of actions for delivering diagnosis, thus its PS label is P4. Similarly, Turn 60 and Turn 65 are the start of two action sequences of physician's treatment recommendations, thus their PS labels are both P5. Note that phases can go back and forth. Therefore, a P7 label can precede a P6 Label.

In sum, PS marks the natures of and boundaries of sequences; SL marks the relations of turns within a sequence (similar to a dependency tree); and APP indicates the role of a turn within an adjacency pair.

Based on the CA theory, this multi-layer structure annotation scheme is not only salient in indicating a turn's position in a conversation, but also important for determining the type of action that a turn undertakes (Stivers, 2014; Schegloff, 2007; Sacks, 1992). The hierarchical structural information thus forms a fine-grained contextual constraint for the way of a turn in conversation can be understood. Therefore, by incorporating our shared knowledge or commonsense about the context of a turn in conversation, the COSTA annotation scheme is capable of dealing with problems such as comprehending indirect speech actions, as it no long relies on the surface composition of a turn to classify its action type.

As this is preliminary work, we used code-recode procedure to test the agreement of the structural and action annotations. The overall agreement achieved 94.43% among the APP, SL, and PS attributes[3].

---

[3]Since PR is assigned during the transcribing process and

3.3 Task-specific Annotations

Besides examining conversational structures, we also examine the decision-making process of antibiotic treatment in the specific clinical context of pediatric consultations. This task is motivated by the fact that antibiotic over-prescribing and bacterial resistance is a big global public health crises today, and the problem is particularly severe in China in the pediatric settings (Li et al., 2012; Laxminarayan et al., 2013).

Several kinds of physician-patient/caregiver conversational actions are annotated, as well as prescribing outcome of the visits. For example, the task-specific annotations are marked on the last two columns of Table 1, and explained below:

Caregivers' advocacy for antibiotics (A) is marked in the turn where a caregiver advocates for antibiotic treatment in the medical visits. This attribute has four possible values, indicating a varying degree of overtness of the advocating actions:
- *A1:* Explicit request for antibiotics (e.g., *Can you prescribe me some antibiotics?*)
- *A2:* Statement of desire for antibiotics (e.g., *Her mother wants to put her on antibiotics.*)
- *A3:* Inquiry about antibiotics (e.g., *Does he need antibiotics?*)
- *A4:* Evaluation of treatment effectiveness (e.g., *Antibiotics always work well for her.*)

Physicians' treatment recommendation (B) is used for a turn where physician makes a treatment recommendation. This attribute has three possible values, indicating a varying degree of physician authoritarian style in delivering the treatment recommendation.
- *B1:* Pronouncement (e.g., *She has to take some antibiotics now.*)
- *B2:* Proposal (e.g., *How about we put her on antibiotics?*)
- *B3:* Offer (e.g., *If you'd like, I can prescribe you some antibiotics.*)

Response to treatment recommendation or **Response to antibiotic advocacy (C)** is used for a turn if it contains a response to either an antibiotic treatment advocacy (A) or a treatment recommendation (B). Such a turn normally appears immediately after a turn with an A or B action. Two possible values are: *C1:* Acceptance and *C0:* Non-acceptance [4].

TID is assigned automatically after the transcribing process, they were excluded from the test.

[4]Partial or full rejection are annotated as non-acceptance.

Item	Total	Avg.
Characters	468,162	1472.2
Words	270,042	849.2
Turns	39,216	123.3
Non-verbal turns	5,815	18.3
Adjacency pairs	20,123	63.3
Sequences	9114	28.7

Table 3: Statistics of the annotated corpus. Total number of visit is 318. Avg. refers to that average number is calculated per visit.

Prescribing Outcome (D) marks whether antibiotics are prescribed in a visit. This label is annotated at the end of the conversation as a derived result. It has two possible values: *D1:* Antibiotic treatment and *D0:* Non-antibiotic treatment .

The overall code-recode agreement of the task-specific annotations achieved 97% among the four types of behavior[5].

4 Results

In this section, we first present basic statistics of our corpus; next, we report our findings on 1) the process of treatment decision-making in medical consultation, and 2) the association between physician-patient/caregiver conversational behaviors and antibiotic prescribing outcome in medical consultations.

4.1 Corpus Statistics

In total, our corpus contains 318 manually transcribed conversations, among which 187 are acute visits and 131 are follow-up visits. Table 3 summarizes the statistics of the corpus. The corpus contains nearly 40K turns with 470K Chinese characters, which is considerably large in terms of manually annotated natural human conversations. The Chinese sentences are then automatically word segmented with an in-house CRF model. On average, each visit has three participants (physician might talk to more than one caregiver), and the turns in a visit form 63 adjacency pairs, which in turn form 29 action sequences.

4.2 Treatment Decision-making Process

To investigate the process of treatment decision-making in medical consultation, we focus on the interactive process in which a physician's treatment recommendation is accepted. We found that a physician's treatment recommendation is

---

[5]See Chilisa and Preece (2005) for details of the code-recode strategy. The overall code-recode agreement was calculated based on the average of the four task-specific labels
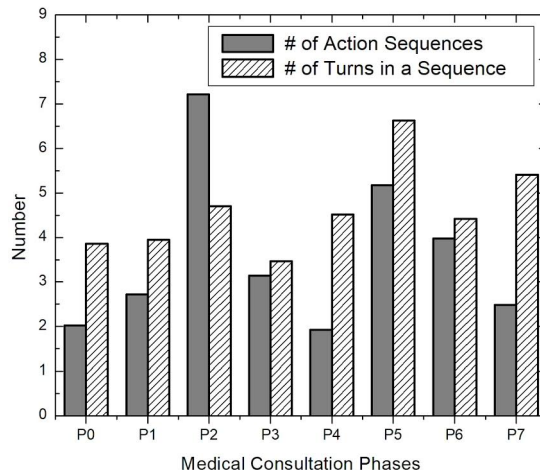

Figure 2: Average number of action sequences and average number of turns in a sequence in medical consultation phases. *P0*: Opening *P1*: Problem presentation *P2*: History taking *P3*: Physical exams *P4*: Diagnosis *P5*: Treatment *P6*: Addressing additional concerns *P7*: Closing.

not always immediately accepted in the next turn; rather, it can be resisted or rejected by a patient or caregiver. In doing so, the patient or caregiver has the opportunity to negotiate for a treatment that is in line with their own wants. As a result, this could lead to rather expanded treatment recommendation action sequence shapes.

After examining our corpus, we found that physicians' treatment recommendations are resisted by caregivers 41% of the time. On average, a treatment recommendation action sequence takes 6.63 turns for its completion. In comparison, other actions in a medical consultation are usually less expanded. A history-taking action sequence takes 4.70 turns, and a problem presentation action sequence takes 3.95 turns to complete on average.

In our corpus, the average number of turns of an action sequence is the greatest in treatment phase (P5) throughout all phases in medical consultation. Figure 2 shows the distribution of the average number of turns for an action sequence and average number of actions in each phase of medical consultations. The long sequence suggests that the treatment phase is where communication problems (understanding or accepting physician's recommendations) are most likely to occur.

4.3 Association between Conversational Behaviors and Antibiotic Prescribing

From the annotated corpus, we can collect various statistics to study the association between physician/caregiver behavior and antibiotics prescribing outcome. Table 4 shows the distribution of advo-

Advocating Action Type	# of Visits	% of Visits
A1 Requests	10	5.3
A2 Statements	14	7.5
A3 Inquiries	50	26.7
A4 Evaluations	26	13.9
Total	100	54.0

Table 4: Frequency and distribution of caregiver advocacy for antibiotics. # of Visit refers to number of visits in which advocacy is observed. % of Visit refers to proportion of visits in which advocacy is observed out of 187 visits.

Prescriptions	V.w.A.	V.w/o.A.	Total	%
Antibiotics	72	39	111	59.4
Non-antibiotics	28	48	76	40.6
Total # of Visits	100	87	187	100.0
%	53.5	46.5	100.0	

Table 5: Frequency and distribution of prescribing outcomes by occurrence of advocacy in number of visits. V.w.A.:visits with advocacy; V.w/o.A.:visits without advocacy. The rightmost column shows the percentage of antibiotic prescriptions out of a total of 187 visits. The bottom row shows the percentage of visits in which caregiver advocacy is observed out of the 187 visits.

cating actions that Chinese caregivers use to advocate for antibiotics. Table 5 shows the distribution of antibiotics prescribing outcomes by occurrence of caregivers' advocacy for antibiotics. The result reveals that caregiver advocacy for antibiotic treatment is significantly associated with antibiotic prescribing outcome. What is more troubling about this finding is that while caregiver advocacy for antibiotic treatment occurred in 54% of the acute visits in our corpus, similar kind of caregiver advocacy for antibiotics were observed only 9% of the time in the similar setting of American pediatric consultations (Stivers, 2002).

In addition, we found that physicians tend to use less authoritarian styles of treatment recommendations (i.e., B2 and B3 combined) than more authoritarian ones (i.e., B1). Table 6 shows distribution of the three types of treatment recommendation actions in the Chinese pediatric context. Moreover, in response to caregivers' advocacy for antibiotic treatment, physicians more frequently resist it than grant it, as shown in Table 7. These findings indicate that physicians play a less dominant role in antibiotic over-prescribing in the medical visits; in contrast, caregivers have a significant influence on the prescribing outcomes.

Multivariate logistic regression results reveal that caregiver advocacy for antibiotic treatment significantly increases the likelihoods of antibiotic prescribing in a visit – caregivers' advocacy was associated with 9.23 times increased likelihoods of antibiotic prescription (Odds Ratio (OR) = 9.23, 95% Confidence Interval(CI): 3.30-33.08); whereas physician's response to caregivers' advocacy has a significant effect on the prescribing outcome – physicians' resistance to caregivers' advocacy reduced the likelihoods of antibiotic prescriptions by 77% (OR=0.23, 95%CI: 0.06-0.68), controlling for the socio-demographic variables in our model.

5 Discussion

Conversational structures have been recognized as critical for conversational understanding in both sociology and artificial intelligence. Although past research has made enormous contributions to this important inquiry, no annotation scheme exists, with which the hierarchical structural organizations of conversation can be captured. Motivated by this gap, we developed the COSTA and created a corpus annotated with this scheme.

5.1 Related Theories and Schemes

Informed by Conversation Analysis (CA), the theoretical framework of the COSTA annotation scheme is largely in line with the existing discourse structure theories and annotation schemes. Although the existing theories have recognized that utterances in conversation have higher-level forms of hierarchical structures (Grosz and Sidner, 1986; Carletta et al., 1997), most have only implemented annotations of conversational structures at turn level and between a pair of turns (e.g., by distinguishing Forward Communicative Function and Backward Communicative Function (Core and Allen, 1997; Jurafsky et al., 1997)).

In addition, the COSTA annotation scheme also presents an innovative method for annotating actions in conversation. Most of the existing annotation schemes of dialog acts for conversations (Core and Allen, 1997; Jurafsky et al., 1997; Stolcke et al., 2000) and particularly, for medical dialogue (Hoxha et al., 2016; Mayfield et al., 2014) were based on Speech Act Theory (SAT); however, the SAT has long been criticized for being difficult in dealing with indirect dialog acts. Different from the SAT, the CA theory considers the sequential position of a turn as critical for action recognition and ascription. The COSTA annotation scheme thus 1) allows multi-layer annotation

Action Type	# of Visits	%
B1 (Pronouncements)	61	41.5
B2 (Proposals)	65	44.2
B3 (Offers)	21	14.3
Total	147	100

Table 6: Frequency and distribution of physicians' treatment recommendations, by three recommendation actions. Percentage of action is out of a total of 147 visits, in which physicians rather than caregivers initiated treatment discussions.

Response to Advocacy	Visit	%
C0 (Non-acceptance)	65	65.0
C1 (Acceptance)	35	35.0
Total	100	100

Table 7: Frequency and distribution of physicians' response to caregiver advocacy for antibiotics. Percentage is out of a total of 100 visits where caregiver advocacy is observed.

at a turn, and 2) depends on the multi-layer structural annotations of a turn for action taxonomy. It thus offers great flexibilities in annotating indirect actions.

5.2 Applications to Different Domains

The COSTA annotation scheme can be used for both general domains and for task-specific domains. While the values for TID, APP, and SL are likely to remain the same for different domains, the values for PR and PS and additional attributes such as A-D labels as described in Section 3.3 are task-specific. In addition,, because the CA theory about conversational structures and actions applies to both ordinary conversation and task-specific conversation, we believe that the same scheme with slight customization (e.g., using a different label set for PS) can accommodate analysis of conversational structures and actions in other task-specific service settings such as airliner hotlines, 911 call centers, etc. Furthermore, since social norms underlying conversations do not tend to vary significantly across cultures, the COSTA scheme can be applied to languages other than Chinese.

5.3 Applications of the Corpus

Although research in medicine has long been concerned with effective communication between physicians and patients, related language resources are still lacking. Our corpus is one of the first to have multi-layer structure annotations of complete natural conversations, in the task-specific setting of physician-patients/caregivers medical consultations.

The findings regarding structural shape of a typ-

ical medical consultation and the process through which a treatment decision is made can be applied to research and practices in medicine and beyond. For example, communication effectiveness can be improved by focusing on phases that are identified as critical in medical consultations (e.g., treatment phase in which sequences are most expanded). In addition, intervention programs can be developed to reduce antibiotic over-prescribing by training physicians to resist caregivers' pressure more effectively. Moreover, the rich information of the corpus can be valuable for building intelligent dialogue system for applications in clinical setting (Campillos et al., 2016).

6 Conclusion

In this paper, we propose a general scheme for annotating multi-layer conversational structures and actions and use that scheme to build a corpus of medical conversations in Chinese pediatric settings. First, our work extends the theory and practice of the sociological field of conversation analysis (CA) by creating an annotation scheme for coding conversational structures and actions. Second, we create a corpus of naturally occurring conversations between physicians and caregivers. The corpus can be used not only for research of general purposes such as conversational understanding, modeling human social behavior of cooperation and coordination, but also for more specific purposes such as identifying risk factors for antibiotic prescribing. Third, we demonstrate that conversational behavior indeed affects medical decisions. We hope our findings can be used to train physicians for effective communication.

For future work, we want to test the usefulness of the scheme in other domains. In addition, we plan to extend COSTA to mark turn construction unit (TCU) [6] . We plan to release the dataset once it is completed.

References

John L. Austin. 1962. *How to Do Things with Words*. William James Lectures. Oxford University Press.

Leonardo Campillos, Dhouha Bouamor, Eric Bilinski, Anne-Laure Ligozat, Pierre Zweigenbaum, and So-

---

[6]In this version, we treated turn as the smallest unit of conversation, but as discussed in (Sacks et al., 1974), a turn may consist of several TCUs (e.g., one TCU answers the question in the previous turn, and another TCU initiates a question).

phie Rosset. 2016. Integrating a dialogue system into a virtual patient consultation.

Jean Carletta, Amy Isard, Stepen Isard, C. Jacquelin Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics* 23(1):13–31.

Bagele Chilisa and Julia Preece. 2005. *Research methods for adult educators in Africa*. UNESCO Institute for Education; Pearson Education, Hamburg, Germany; Cape Town, South Africa.

Mark Core and James Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In *Proceedings of AAAI Fall Symposium on Communicative Action in Humans and Machines*.

Paul Drew. 2014. Turn Design. In J. Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, John Wiley & Sons, West Sussex, UK, pages 131–149.

Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall Inc.

Erving Goffman. 1983. The interaction order: American sociological association, 1982 presidential address. *American Sociological Review* 48(1):1–17.

Barbara Grosz and Candace Sidner. 1986. Attention, Intentions, and the Structure of discourse. *Computational Linguistics* 12(3):175–204.

John Heritage. 1984. A Change-of-State Token and Aspects of its Sequential Placement. In J. Maxwell Atkinson, , and John Heritage, editors, *Structures of social action: Studies in Conversation Analysis*, Cambridge University Press, Cambridge, U.K., chapter 13, pages 299–345.

John Heritage and Douglas W. Maynard. 2006. *Communication in Medical Care: Interaction between Primary Care Physicians and Patients*. Cambridge University Press, Cambridge.

Julia Hoxha, Praveen Chandar, Zhe He, James Cimino, David Hanauer, and Chunhua Weng. 2016. DREAM: Classification Scheme for Dialog Acts in Clinical Research Query Mediation. *Journal of Biomedical Informatics* 59:89–101.

Gail Jefferson. 1988. On the sequential organization of trouble-talk in ordinary conversation. *Social Problems* 35(4):418–441.

Gail Jefferson. 2004. Glossary of transcript symbols with an introduction. In Gene H. Lerner, editor, *Conversation Analysis: Studies from the First Generation*, John Benjamins, Amsterdam / Philadelphia, chapter 2, pages 13–31.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13. Technical report, University of Colorado, Boulder.

R. Laxminarayan, A. Duse, C. Wattal, AK. Zaidi, HF. Wertheim, V. Sumpradit, E. Vlieghe, GL. Hara, IM. Gould, H. Goossens, C. Greko, AD. So, M.Bigdeli, G. Tomson, W. Woodhouse, E. Ombaka, AQ. Peralta, FN. Qamar, F. Mir, S. Kariuki, ZA. Bhutta, A. Coates, R. Bergstrom, GD. Wright, ED. Brown, and O. Cars. 2013. Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet Infectious Diseases* 13(12):1057–1098.

Stephen Levinson. 2014. Action Formation and Action Ascription. In J. Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, John Wiley & Sons, West Sussex, UK, pages 103–130.

Yongbin Li, Jing Xu, Fang Wang, Bin Wang, Liqun Liu, Wanli Hou, Hong Fan, Yeqing Tong, Juan Zhang, and Zuxun Lu. 2012. Overprescribing in china, driven by financial incentives, results in very high use of antibiotics, injections, and corticosteroids. *Health Affairs (Project Hope)* 31(5):1075–1082.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: toward a Functional Theory of Text Organization. *Text: Interdisciplinary Journal for the Study of Discourse* 8(3):243–281.

Elijah Mayfield, M Barton Laws, Ira B. Wilson, and Carolyn Penstein Ros'e. 2014. Automating annotation of information-giving for analysis of clinical conversation. *Journal of American Medical Informatics Association* 21(e1):e122–e128.

Raymond C. Perrault. and James Allen. 1980. A Plan-based Analysis of Indirect Speech Acts. *Computational Linguistics* 6(3):167–182.

Martha Elizabeth Pollack. 1986. *Inferring Domain Plans in Question–answering*. Ph.D. thesis, Philadelphia, PA, USA. UMI order no. GAX86–14850.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Livio Robaldo Aravind Joshi Eleni Miltsakaki, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*.

Jeffery D. Robinson. 2003. An interactional structure of medical activities during acute visits and its implications for patient's participation. *Health Communication* 15(1):27–57.

Jeffery D. Robinson. 2014. Overall Structural Organization. In J. Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, John Wiley & Sons, West Sussex, UK, pages 257–280.

Harvey Sacks. 1992. *Lectures on Conversation*, volume 1 & 2. Basil Blackwell.

Harvey Sacks and Emanuel A. Schegloff. 1973. Opening up closings. *Semiotica* 8(4):289–327.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language* 50(4, Part 1):696–735.

Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Cambridge University Press.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53(2):361–382.

Alfred Schütz. 1967. *Phenomenonology of the Social World*. Northwestern University Press.

John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

John R. Searle. 1985. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press.

Tanya Stivers. 2002. Participating in decisions about treatment: Overt parent pressure for antibiotic medication in pediatric encounters. *Social Science & Medicine* 54(7):1111–1130.

Tanya Stivers. 2007. *Prescribing under pressure: Parent-physician conversations and antibiotics*. Oxford University Press, London.

Tanya Stivers. 2014. Sequence Organization. In J. Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, John Wiley & Sons, West Sussex, UK, pages 191–209.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* 26(3):339–373.

Max Weber. 1991. The Nature of Social Action. In W. G. Runiciman Weber, editor, *Selections in Translation*, Cambridge.

Ludwig Wittgensein. 1953. *Philosophical Investigations*. Oxford University Press.

Kelly B. Haskard Zolnierek and M. Ribin Dimatteo. 2009. Physician communication and patient adherence to treatment: A meta-analysis. *Medical care* 47(8):826–834.