

# Negotiation of Antibiotic Treatment in Medical Consultations: A Corpus based Study

Nan Wang

University of California, Los Angeles

nwang3@ucla.edu

## Abstract

Doctor-patient conversation is considered a contributing factor to antibiotic over-prescription. Some language practices have been identified as parent pressuring doctors for prescribing; other practices are considered as likely to engender parent resistance to non-antibiotic treatment recommendations. In social science studies, approaches such as conversation analysis have been applied to identify those language practices. Current research for dialogue systems offer an alternative approach. Past research proved that corpus-based approaches have been effectively used for research involving modeling dialogue acts and sequential relations. In this proposal, we propose a corpus-based study of doctor-patient conversations of antibiotic treatment negotiation in pediatric consultations. Based on findings from conversation analysis studies, we use a computational linguistic approach to assist annotating and modeling of doctor-patient language practices, and analyzing their influence on antibiotic over-prescribing.

## 1 Introduction

“How to do things with words” has long been a topic interested to researchers from various disciplines, such as pragmatics (Austin, 1962; Levinson, 1983), conversation analysis (CA) (Drew and Heritage, 1992; Heritage and Maynard, 2006), and computational linguistics (Stolcke et al., 2000; Williams et al., 2013; Schlöder and Fernandez, 2015). Although computational methods have been widely used to conduct text mining tasks such as detecting reader bias and predicting mood shift in vast populations (Flaounas et al., 2013;

Lansdall-Welfare et al., 2012; Ritter et al., 2011), studies on computational modeling of human natural conversational acts are rare, especially for investigating associations with social behavioral outcomes.

Doctor-patient conversations have been proved highly consequential on a lot of worldwide public health problems and population health outcomes (Zolnieriek and DiMatteo., 2009; Mangione-Smith et al., 2015). Over-prescription of antibiotics is often related to interaction-generated problems arising from doctor-patient conversations, which has little to do with rational medical judgments (Macfarlane et al., 1997). For example, some parent language practices are frequently understood by physicians as advocating antibiotics, resulting in significantly higher likelihood of inappropriate prescriptions (Mangione-Smith et al., 1999; Stivers, 2002, 2007).

This antibiotic resistance and over-prescription phenomenon also has its presence in China. Prescription rates of antibiotics is high (Li et al., 2012; Wang et al., 2014; Xiao et al., 2012); multiple types of antibiotic resistant pathogens have been discovered nationwide. However, determinants of the over-prescription problem in China have not been well studied, especially the impact of doctor-patient conversation in medical consultations.

In this proposal, we propose a corpus based study to examine doctor-patient conversation of antibiotic treatment negotiation in Chinese pediatric settings, using a mixed methodology of conversation analysis and computational linguistics. Particularly, we aim to discover (1) how parent requests of antibiotic prescriptions are made in doctor-patient conversations and their effects on prescribing decision outcomes; (2) how physicians’ non-antibiotic treatment recommendations are delivered and responded by parents; In conducting this study, our findings about doctor-

patient conversation are expected to be extended beyond medical setting to natural human conversations. These findings include:

- How actions are formulated with various forms of language practice in conversations;
- How meaning of language practices is understood by speakers as performing a certain action;
- How choice of one form of language practices in performing an action is associated with its response of various kinds.

In conducting this study, we attempt to bridge the gap between social scientific methods and computational methods in researching the aforementioned questions.

In the following sections, we will introduce our corpus, preliminary findings from CA, and related computational approaches. This is followed by a discussion of contributions of the proposed study.

## 2 Data

The corpus of this study is constructed from natural human conversations. In order to obtain the conversations, 318 video-recorded doctor-patient conversations were collected in 6 Chinese hospitals between September and December in 2013. Each conversation is around 5 minutes in length, resulting in 30 hours of video-recordings in total. The conversations were mostly between doctors and patients’ caregivers regarding patients’ health conditions and lifestyle-related issues that are commonly discussed in pediatrics.

Video-recordings were then transcribed manually. Six researchers were employed to transcribe the data, including one manager and five annotators. All of them are native speakers of Chinese. The five annotators received basic training in CA and its transcribing conventions before they started transcribing. The manager is a specialist in CA, who controlled the work flow and trouble-shot during the transcribing process.

Following the Jeffersonian transcribing conventions (Jefferson, 2004), the video-recorded conversational data were transcribed with considerable details with respect to speech production, including the speech text verbatim and other paralinguistic features such as intonations, overlaps, visible non-verbal activities and noticeable timed silence (Auer et al., 1992).

TID	SID	RID	UID	Speech text
1	1	0	M	您好. 就是咳嗽. Hello. He’s just coughing.
2	2	0	D	((处理病历文件)) ((Processing medical document))
3	3	0	D	来看过的, 是啊? You have visited us for this, have you?
4	3	3	M	嗯, 那天晚上 来挂了急诊室.= Yeah, we came to the emergency room that night.
5	4	4	D	=嗯, 嗯. Yeah, yeah.

Table 1: An example of annotated conversation.

To answer our research questions, we developed an annotation schema, capturing the following aspects of the conversations, including (1) turn-taking and speakership (TID, UID), (2) multi-turn dependency relations, such as adjacency pair<sup>1</sup> (SID) and rhetorical relations<sup>2</sup> (RID). In addition, the speech text was also word segmented corresponding to Chinese Penn Tree Bank segmentation guideline (Xia et al., 2000). An example of the corpus is shown in Table 1.

The current annotated corpus contains 318 conversations with nearly 40K turns and 470K Chinese characters. It has on average 123 turns and 81 adjacency pairs in each conversation. The average number of participants is 3 in each conversation, with a minimum of 2 speakers and a maximum of 8 speakers.

## 3 Conversation Analysis

Conversation analysis (CA) is used to identify the dialogue acts in the corpus. CA views sequence organization a core feature of conversation that is important for understanding the meaning of an utterance and its significance as an action in conversation (Schegloff, 1968). The idea is that the action which some talk is doing can be grounded in its position, not just its composition. Therefore, some talk (e.g. “It’s raining.”) can be heard as an answer to a question (e.g. “Are we going to the game?”), even they are apparently semantically unrelated. The relationship of adjacency between turns is central to the ways in which talk in conversation is organized and understood (Schegloff,

<sup>1</sup>A basic sequential relationship defined in conventional conversation analysis literature, to be explained in the next section.

<sup>2</sup>RID: Information reflecting topical relevances across turns.

2007). The adjacency relationship most powerfully operates in two ways: (1) backwards - next turns are understood by co-participants to display their speaker's understanding of the prior turn and to embody an action responsive to the prior turn so understood; (2) prospective - a first pair part in an adjacency pair projects some prospective relevance rules for the second pair part. Specifically, it makes relevant a limited set of possible second pair parts, and thereby sets some terms by which a next turn will be understood (Schegloff, 2007).

The methodology of CA relies on audio or video-recordings of naturally occurring conversations, which are then transcribed in details for analyses of turns and sequences in the conversation (Sidnell et al., 2013) and the embodied actions that speakers use to accomplish their goals in social interactions (Drew and Heritage, 1992; Drew et al., 2001). In general, CA looks for patterns in the conversation which form evidence of systematic usage that can be identified as a 'practice' through which people accomplish a social action. To be identified as a practice, a particular communication behavior must be seen to be recurrent and to be routinely treated by recipients in a way such that it can be discriminated from related or similar practices (Heritage, 1984; Stivers, 2005).

Utilizing CA, we identify parent practices of making requests and physician practices making treatment recommendations in our corpus. These findings are then used for developing an annotation schema for computational modeling of these dialogue acts and the associations with their responses or action outcomes.

#### 4 Preliminary Results

Based on conversation analytical study, we find that four parent language practices are recurrently treated by physicians as requesting antibiotic treatment:

- Explicit requests of an antibiotic treatment;
- Desire statements of an antibiotic treatment;
- Inquiries about an antibiotic treatment;
- Evaluations of a past treatment.

Among the four language practices, only the first practice takes a canonical form of request (e.g., "*Can you prescribe me some antibiotics?*"), while the other three practices take less explicit language

formats, putting varying degree of impositions on physicians' responsive acts.

For example, an explicit request of antibiotic treatment is the strongest form of request as it puts the highest degree of impositions on physicians' responsive action, by making physicians' grant or rejection of the request relevant in the next turn. In contrast, a statement of desire for antibiotic treatment does not put physicians under any constraint for granting an antibiotic prescription, but it generates an understanding that prescribing antibiotics is a desirable act under this circumstance. Similarly, an inquiry about antibiotics raises antibiotic treatment as a topic for discussion and implicates a preference for the treatment, yet it does not put physicians under the constraint as an explicit request does. Moreover, a positive evaluation of past experience with antibiotics may be subject to physicians' understanding as desiring for antibiotics for the current condition, yet it does not even require any response physicians as an inquiry about antibiotics does.

The CA study of the requesting practices enables us to identify the utterances that are recurrently understood or subject to speakers' understanding as doing the act of requesting. In addition, we find that explicit requests are least frequently used by parents, while less explicit forms of requests occur more frequently. Table 2 describes the frequency (number of cases) and percentage of the requesting practices out of total number of cases in the corpus.

In order to quantitatively investigate the correlation between the presence of the requesting practices and the prescribing decision outcomes, we conduct a Pearson's  $\chi^2$  test between the two variables  $X$  and  $Y$ , where  $X$  is whether parents use at least one of the four requesting practices, and  $Y$  is whether they receive an antibiotic treatment by the end of the consultation. The  $\chi^2$  test suggests that parents using at least one of the four requesting practices is significantly associated with that they receive an antibiotic treatment ( $\chi^2=5.625$ ,  $df = 1^3$ ,  $p = 0.018^4$ ). It is worth noting that this is an approximation of the correlation between parent use of the requesting practices and the prescribing outcomes. Investigation of correlations between individual parent requesting practices and the prescribing outcomes will be carried out in our ongoing

<sup>3</sup>Degrees of freedom.

<sup>4</sup>At the 0.05 significance level.

Request	Frequency	Percentage
Explicit requests	23	7.23%
Desire statements	38	11.94%
Inquiries	90	28.30%
Evaluations	69	21.70%
Total	220	69.17%

Table 2: Distribution of requesting practices in the corpus. Each cell reports the number of cases (and percentage) containing the practice.

ing work. Moreover, computational methods will also be introduced to examine the correlations.

In examining what kind of treatment recommendations are more likely to be resisted by parents, we investigate the association between physicians’ non-antibiotic treatment recommendations and parents’ responses in the next turn.

One way to distinguish the delivery format of a non-antibiotic treatment recommendation is whether it is negative-format or positive-format (Stivers, 2005). A negative-format recommendation is to recommend against a particular treatment (e.g., “*She doesn’t need any antibiotics.*”); while a positive-format one is to recommend for a particular treatment (e.g., “*I’m gonna give her some cough medicine.*”). Findings from the American pediatric settings show that physicians’ positive-format recommendations are less likely to engender parent resistant response to a non-antibiotic treatment recommendation than negative-format recommendations, and thus suggests that recommendations delivered in an affirmative, specific form are most receptive to parents for non-antibiotic treatment (Stivers, 2005).

Beyond distinguishing the recommendations into positive-format and negative-format, there are many other features which could be taken into consideration regarding to their consequences on parents’ responses (e.g. epistemic stances<sup>5</sup> and deontic stances<sup>6</sup> that are embodied in the recommending practices). For example, physicians’ treatment recommendations can be produced with the following types, including assertions, proposals and offers. The assertions are recommen-

<sup>5</sup>The epistemic stance refers to speakers’ orientation toward the relative primacy/subordination in terms of their knowledge access. See (Heritage and Raymond, 2005) for more details.

<sup>6</sup>The deontic stance refers to speakers’ orientation toward their relative primacy/subordination in terms of their rights to decide future events. See (Stevanovic and Peräkylä, 2014) for more details.

Recommendation	Frequency	Percentage
Assertions	128	51.00%
Proposals	87	34.66%
Offers	31	12.35%
Total	246	98.01%

Table 3: Distribution of recommending practices in the corpus. Each cell reports the number of cases (and percentage) containing the corresponding practice.

dations such as “*You have to take some fever medicine.*”. Proposals are such as “*Why don’t you take some cough syrup?*”. Offers are mostly recommendations that are offered following parent indication of their treatment preference or desires, e.g. “*I’ll give you some fever medicine if you want.*”. The assertions index higher physician epistemic and deontic rights in terms of who knows the best about the treatment and who determines what the patient needs to do respectively. Compared to assertions, physicians claim less epistemic and deontic authority by using the proposal format; and offers embody the least amount epistemic and deontic primacy. Table 3 describes the distribution of physicians’ practices of making treatment recommendations across the corpus.

We also conduct a Pearson’s  $\chi^2$  test between physicians’ choice of recommending practice and parent response. The test shows that we cannot reject the null hypothesis that physicians’ choices of recommending practice type are independent of parent response types ( $\chi^2=0.327$ ,  $df = 2$ ,  $p = 0.849$ ). Thus our ongoing work is to examine other complexities of treatment recommending practices and their effect on parents’ response.

## 5 Computational Approach

Conversation analysis allows us to manually identify language practices that are recurrently understood and subject to speaker understanding of doing a particular act; while computational approach is used to assist tasks such as entity type recognition, dialogue act classification, and analyses of interested correlations in a more scalable way.

Early research (Jurafsky et al., 1998; Stolcke et al., 2000) on computational modeling of conversational language has demonstrated that automatic modeling based on manually transcribed conversational data by including features such as speakership, dependency relations have achieved supe-

rior performance results compared to datasets otherwise. In using the computational approach in our study, several techniques will be used. In general, we can divide our computational tasks into two categories, fundamental and dialogue specific tasks.

## 5.1 Fundamental Tasks

Fundamental tasks mainly involve solving general problems that are across all language processing tasks, e.g. named entity recognition and coreference resolution. This part of work lays foundations for more advanced dialogue specific tasks to be discussed in the next section.

### 5.1.1 Named Entity Recognition

Entities are very important in spoken language understanding, as it conveys key information in determining task objectives, intents, etc. In the medical domain, entity recognition is particularly crucial in identifying information such as treatment or prescriptions. As a fundamental natural language processing (NLP) technique for various tasks, e.g. machine translation (Babych and Hartley, 2003), information retrieval (Guo et al., 2009), named entity recognition (NER) (Nadeau and Sekine, 2007) is also used in our study. Using NER in our study has several challenges. For example, utterances in dialogues are shorter compared to other types of texts. Also, NER is conducted on Chinese. Thus, domain specific word segmentation (Song and Xia, 2012) is a prerequisite if we extend our work to larger datasets in a more scalable way. However, using NER in our study has the advantage that utterances in dialogues are not isolated. The sequential relations between the utterances thus potentially provides us with more information to build a better model. Previous work (Weston et al., 2015) proved that information extraction which takes into account information from previous utterances with recurrent neural networks was more effective. NER in our study can provide more in-depth annotations to the corpus, allowing models trained on the corpus to incorporate more information. To accelerate the annotation process, semi-supervised methods are used for dialogue acts recognition and classification. Specifically, we annotate some seed data, use the trained model to automatically annotate the rest, and finally check the automatically generated annotations manually.

### 5.1.2 Coreference Resolution

In natural language, reference is used widely for communication efficiency. In dialogue environments, person reference and even omissions are very common. Therefore, coreference resolution can help us add useful semantic information into our language models (Recasens et al., 2013; Clark and Manning, 2015). General coreference resolution is usually performed on multiple sentences in a document; however, the relations of these sentences are vague. Based on our multi-turn rhetorical relation annotations, information that are absent or abstract in a turn can be extracted from turns that are rhetorically related. This could effectively enhance the performance of coreference resolution and provide more accurate information about the referent. For example, the pronoun *that* may not be clear about what it refers to in one utterance; however, the co-reference resolution technique links it to previous turns which contain the information of its referent.

## 5.2 Dialogue Specific Tasks

Our research is closely related to the studies on dialogue systems (Henderson, 2015), in which models are built to structure conversations. To achieve our research goals, models are built to track states in a dialogue and to build connections between utterances and action outcomes.

### 5.2.1 Dialogue State Modeling

One important task is to classify types of an utterance and types of the action required. For example, to judge whether an utterance is a question, answer, or other dialogue act, classification can be performed, taking into account turns in previous context. Previous work (Henderson et al., 2013; Ren et al., 2014) demonstrated that using a classifier was effective for modeling user intents and utterance types. In our research, we will use this approach to classify utterances into different types such as dialogue acts, parent responses and treatment decisions. In order to perform such classification, further annotations are conducted based on the findings of conversation analyses, including:

- Dialogue act - parent requests for antibiotic treatment, physician treatment recommendations;
- Treatment type - antibiotic or non-antibiotic treatment;

- Response type - grant or rejection to recommendation.

By using these classifiers, it allows us to investigate the features that are most important for classifying the utterances, and then align them with the qualitative findings from CA studies.

Another way to model dialogue states is treating dialogues as a sequence of observations and then build models (e.g., CRF (Lafferty et al., 2001), LSTM (Hochreiter and Schmidhuber, 1997)) to perform labeling or classification based on that. This is a natural way of modeling dialogues in terms of the problem proximity. Current state-of-the-art studies suggest that LSTM is a good choice for modeling not only sequences of turns, but also sequences of words (or other basic units) within a turn (Zilka and Jurčicek, 2015). Using our corpus, an LSTM model can be trained to achieve the same goal as static classifiers for practice type classification, and to model the sequential relationship between turns in real conversations.

Previous studies (Lee and Eskenazi, 2013; Williams, 2014; Henderson et al., 2014) found that systems combining the classifier approach and the sequence model approach showed competitive results. In doing so, one can train several different models with different sets of parameters and join their results accordingly (Henderson et al., 2014). For the aforementioned classification and sequence modeling tasks, the combined model is expected to outperform individual models.

### 5.2.2 Domain Adaptation

Since our data is of the particular domain of medicine, domain adaptation is another task involved in our research. Almost all of the aforementioned tasks can be affected by domain specific variance. Besides, conversational data in medical domain is also lacking. Therefore, acquiring more data from other or general domain can be useful in completing the tasks in the medical domain, and improving the capability of conversational understanding. Training data selection/acquisition (Axelrod et al., 2011; Song et al., 2012) could be the first step to solve the problem of domain variance, without the need to modify the existing models to fit our domain. Moreover, when this work has to be extended to other domains, e.g., law, education, etc., domain adaptation is required to transfer the knowledge from this domain to another.

## 6 Discussion

In this proposal, we propose a study on doctor-patient conversations based on a corpus of naturally occurring medical conversation that are transcribed and annotated manually. With the combination of the social science research method of conversation analysis and computational methods for language modeling, we aim to discover how language practices in doctor-patient conversation influence antibiotic over-prescribing.

Although previous studies (Macfarlane et al., 1997; Mangione-Smith et al., 1999; Stivers, 2007) proved that doctor-patient conversation were consequential on medical decision-making and population health outcomes, findings from the extant social science research are still limited in answering the question “*in what way the language practices that doctors and patients use in medical consultations influence the decision outcomes*”.

Based on our preliminary findings from the CA studies, we propose to use the computational approach to help answer our research questions. In doing so, language patterns that are interested in CA studies can be automatically modeled and predicted with classifier or sequence models, leading us to more interesting findings. Also, by using the computational approach, we can also build a dialogue system based on our corpus. This system can be useful for analyzing doctor-patient conversation and assisting decision-making process in medical consultations.

In addition, we constructed a manually transcribed and annotated corpus. Our ongoing work involves formalizing and adding additional annotations to the corpus. We will release the corpus to the community in near future. It will be a unique resource for both social scientific and computational linguistic studies of conversations in the medical domain.

## References

- Peter Auer, Frank Muller, and Elizabeth Couper-Kuhlen. 1992. *Language in Time: The Rhythm and Tempo of Spoken Interaction*. Oxford University Press, New York.
- John Langshaw Austin. 1962. *How to do things with words*. William James Lectures. Oxford University Press.
- Amittai Axelrod, Xiaodong He, and Jianfeng Gao. 2011. Domain adaptation via pseudo in-domain data

- selection. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP-2011)*. pages 355–362.
- Bogdan Babych and Anthony Hartley. 2003. Improving machine translation quality with automatic named entity recognition. In *Proceedings of the 7th International EAMT Workshop on MT and Other Language Technology Tools, Improving MT Through Other Language Technology Tools: Resources and Tools for Building MT*. Association for Computational Linguistics, Stroudsburg, PA, USA, EAMT '03, pages 1–8.
- Kevin Clark and Christopher D. Manning. 2015. Entity-centric coreference resolution with model stacking. In *Association for Computational Linguistics (ACL)*.
- Paul Drew, John Chatwin, and Sarah Collins. 2001. Conversation Analysis: A Method for Research into Interactions between Patients and Health-care Professionals. *Health Expectations* 4:58–70.
- Paul Drew and John Heritage. 1992. Analyzing talk at work: an introduction. In Paul Drew and John Heritage, editors, *Talk at work: interaction in institutional settings*, Cambridge University Press, Cambridge, pages 3–65.
- Ilias Flaounas, Omar Ali, Thomas Lansdall-Welfare, Tjil De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. 2013. Research methods in the age of digital journalism. *Digital Journalism* 1(1):102–116.
- Jiafeng Guo, Gu Xu, Xueqi Cheng, and Hang Li. 2009. Named entity recognition in query. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, SIGIR '09, pages 267–274.
- Matthew Henderson. 2015. Machine learning for dialog state tracking: A review. In *Proceedings of The First International Workshop on Machine Learning in Spoken Language Processing*.
- Matthew Henderson, Blaise Thomson, and Steve Young. 2013. Deep neural network approach for the dialog state tracking challenge. In *Proceedings of the SIGDIAL 2013 Conference*. Metz, France, pages 467–471.
- Matthew Henderson, Blaise Thomson, and Steve Young. 2014. Word-based dialog state tracking with recurrent neural networks. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Philadelphia, PA, U.S.A., pages 292–299.
- John Heritage. 1984. *Garfinkel and Ethnomethodology*. Polity Press, Cambridge.
- John Heritage and Douglas W. Maynard. 2006. *Communication in Medical Care: Interaction between Primary Care Physicians and Patients*. Number 20 in Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge.
- John Heritage and Geoffrey Raymond. 2005. The terms of agreement: Indexing epistemic authority and subordination in assessment sequences. *Social Psychology Quarterly* 68:15–38.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9(8):1735–1780.
- 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, Rebecca Bates, Noah Coccaro, Rachel Martin, Marie Meteer, Klaus Ries, Elizabeth Shriberg, Andreas Stolcke, Paul Taylor, and Carol Van Ess-Dykema. 1998. Switchboard discourse language modeling project report. *Center for Speech and Language Processing, Johns Hopkins University, Baltimore, MD*.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ICML '01, pages 282–289.
- Thomas Lansdall-Welfare, Vasileios Lampos, and Nello Cristianini. 2012. Effects of the recession on public mood in the uk. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, New York, NY, USA, pages 1221–1226.
- Sungjin Lee and Maxine Eskenazi. 2013. Recipe for building robust spoken dialog state trackers: Dialog state tracking challenge system description. In *Proceedings of the SIGDIAL 2013 Conference*. Association for Computational Linguistics, Metz, France, pages 414–422.
- Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press.
- Yongbin Li, Jing Xu, Fang Wang, Bin Wang, Liquan 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.
- J. Macfarlane, W. Holmes, R. Macfarlane, and N. Britten. 1997. Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: questionnaire study. *BMJ* 315(7117):1211–1214.

- R. Mangione-Smith, E. A. McGlynn, M. N. Elliott, P. Krogstad, and R. H. Brook. 1999. The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. *Pediatrics* 103(4):711–718.
- R. Mangione-Smith, C. Zhou, J. D. Robinson, J. A. Taylor, M. N. Elliott, and J. Heritage. 2015. Communication practices and antibiotic use for acute respiratory tract infections in children. *Annals of Family Medicine* 13(3):221–227.
- David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Linguisticae Investigationes* 30(1):3–26.
- Marta Recasens, Matthew Can, and Dan Jurafsky. 2013. Same referent, different words: Unsupervised mining of opaque coreferent mentions. In *North American Association for Computational Linguistics (NAACL)*.
- Hang Ren, Weiqun Xu, and Yonghong Yan. 2014. Markovian discriminative modeling for dialog state tracking. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Philadelphia, PA, U.S.A., pages 327–331.
- Alan Ritter, Colin Cherry, and William B. Dolan. 2011. Data-driven Response Generation in Social Media. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA, EMNLP ’11, pages 583–593.
- Emanuel Schegloff. 2007. *Sequence organization in interaction: Volume 1: A primer in conversation analysis*. Cambridge University Press.
- Emanuel A. Schegloff. 1968. Sequencing in Conversational Openings. *American Anthropologist* 70(6):1075–1095.
- Julian J. Schlöder and Raquel Fernandez. 2015. Clarifying Intentions in Dialogue: A Corpus Study. In *Proceedings of the 11th International Conference on Computational Semantics*. London, UK, pages 46–51.
- Jack Sidnell, Tanya Stivers, and Eds. 2013. *The Handbook of Conversation Analysis*. Wiley-Blackwell.
- Yan Song, Prescott Klassen, Fei Xia, and Chunyu Kit. 2012. Entropy-based Training Data Selection for Domain Adaptation. In *Proceedings of COLING-2012*. Mumbai, India, pages 1191–1200.
- Yan Song and Fei Xia. 2012. Using a Goodness Measurement for Domain Adaptation: A Case Study on Chinese Word Segmentation. In *Proceedings of LREC-2012*. Istanbul, Turkey, pages 3853–3860.
- Melisa Stevanovic and Anssi Peräkylä. 2014. Three Orders in the Organization of Human Action: On the Interface between Knowledge, Power, and Emotion in interaction and social relations. *Language in Society* 43(2):185–207.
- Tanya Stivers. 2002. Participating in Decisions about Treatment: Overt Parent Pressure for Antibiotic Medication in Pediatric Encounters. *Social Science and Medicine* 54(7):1111–1130.
- Tanya Stivers. 2005. Non-antibiotic treatment recommendations: Delivery formats and implications for parent resistance. *Social Science & Medicine* 5(60):949–964.
- Tanya Stivers. 2007. *Prescribing under Pressure: Parent-physician Conversations and Antibiotics*. Oxford University Press, London.
- Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26(3):339–373.
- Jin Wang, Pan Wang, Xinghe Wang, Yingdong Zheng, and Yonghong Xiao. 2014. Use and prescription of antibiotics in primary health care settings in china. *JAMA Internal Medicine* 174(12):1914–1920.
- Jason Weston, Antoine Bordes, Sumit Chopra, and Tomas Mikolov. 2015. Towards ai-complete question answering: A set of prerequisite toy tasks. *ArXiv pre-prints* abs/1502.05698.
- Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. The Dialog State Tracking Challenge. In *Proceedings of the SIGDIAL 2013 Conference*. Metz, France, pages 404–413.
- Jason D Williams. 2014. Web-style ranking and slu combination for dialog state tracking. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Philadelphia, PA, U.S.A., pages 282–291.
- Fei Xia, Martha Palmer, Nianwen Xue, Mary Ellen Okurowski, John Kovarik, Fudong Chiou, Shizhe Huang, Tony Kroch, and Mitch Marcus. 2000. Developing Guidelines and Ensuring Consistency for Chinese Text Annotation. In *Proceedings of the Second Language Resources and Evaluation Conference (LREC)*.
- YH Xiao, P Shen, ZQ Wei, YB Chen, and HS Kong. 2012. Mohnarin report of 2011, monitoring of bacterial resistance in China. *Chinese Journal of Nosocomiology* 22:4946–4952.
- Lukás Zilka and Filip Jurčicek. 2015. Incremental lstm-based dialog state tracker. *ArXiv pre-prints* abs/1507.03471.
- Kelly B. Haskard Zolnieriek and M. Robin DiMatteo. 2009. Physician communication and patient adherence to treatment: A meta-analysis. *Medical care* 47(8):826–834.