An Empirical Study of Topic Transition in Dialogue

Mayank Soni† **Brendan Spillane**† **Leo Muckley**† **Orla Cooney** ‡ **Emer Gilmartin**†

Christian Saam† **Benjamin R. Cowan** ‡ **Vincent Wade**†

†ADAPT Centre, †Trinity College Dublin
leo_muckley@hotmail.com, {sonim, spillab, saamc, gilmare, vincent.wade}†@tcd.ie,
{benjamin.cowan, orla.cooney}‡@ucd.ie,

Abstract

Although topic transition has been studied in dialogue for decades, only a handful of corpora based quantitative studies have been conducted to investigate the nature of topic transitions. Towards this end, this study annotates 215 conversations from the switchboard corpus, perform quantitative analysis and finds that 1) longer conversations consists of more topic transitions, 2) topic transition are usually lead by one participant and 3) we found no pattern in time series progression of topic transition. We also model topic transition with a precision of 91%.

Turns	Dialogue Text
Turn 1: A:	All right um well [laughter-uh] let's see i'm twenty
Turn 2: B:	How old are you Lisa. Okay that i'm older
Turn 3: A:	Yeah how old are you. Older [laughter]
Turn 4: B:	Older than you [laughter-are]
Turn 5: A:	[laughter-okay]
Turn 6: B:	Okay we are supposed to talk about places we like to go so i'm gonna and where are you from where are you calling from ?
Turn 7: A:	I'm calling from uh Provo Utah but I'm from Plano Texas
Turn 8: B:	Oh you are from Plano my sister lives in Plano yes her husband is the new Director of Admissions at uh University of Texas at Dallas
Turn 9: A:	Oh really. Oh wow my dad used to work at UTD also
Turn 11: B:	Yeah so I [vocalized-noise]. Anyway so where's your favorite place to go ?
Turn 12: A:	Um. Generally we just go on family vacations to Arizona my grandparents live there that's generally our usual summer vacation

Figure 1: Hand-picked example of topic transition in the Switchboard corpus. Colors represent segments of conversation about the same topic

1 Introduction

Human conversation consists of multiple natural topic transitions, from introductions, to topics of interest, and on to leave talking, and thus relies on topic change and shading mechanisms to allow participants to maintain and change topics[1]. An example of topic transition can be seen in Figure 1, participants first begin by talking about each others age, then move on to the places they want to visit and finally move on to talking about the state of Arizona in the USA. Although topic transition has been studied in linguistics for decades (Gardner, 1984; Lambrecht, 1996; Riou, 2015; Van Dijk, 1977a), there are only a few corpora based studies investigating the nature of topic change. This is because of the labour intensive task of manually annotating datasets. Even though the task of annotation is labour intensive and manual, it is necessary to empirically understand how human participants engage in topic transition in a conversation.

Towards this end, this work annotates 215 conversations from the Switchboard (Godfrey and Holliman, 1993) corpus and studies different aspects of topic transition. To the authors best knowledge, this is the the largest quantitative study conducted on the nature of topic transition in social conversations till date. The dataset curated and code utilized can be found at [2].

2 Background Theory

Definitions of topic in the literature fall into two categories; sentence level (Lambrecht, 1996) and discourse level (Van Dijk, 1977a). Gardner (1984) emphasizes the presence and identification of a topic to be a *intuitive* phenomenon answering the question of 'being about'. Multiple Sentence-level topics about the same thing may consist of a discourse-level topic (Van Dijk, 1977b). As this study is concerned with discourse level topic annotation, we adopt the definition of Bonin et al. (2012) which maintains that topic at a discourse level is the *"segments of the discourse sharing coherent information (about the same thing)"*.

Topic transition has been categorized by Gardner (1984), whose model of topic development in

[1]Our annotated dataset and models do not differentiate between the types of topic transition (change, shift, shading, fading etc.) depicted in Gardner's model (Gardner, 1984). For simplicity, this paper uses 'topic transition' to describe all forms. Where necessary, it uses specific terms to differentiate.

[2]github.com/Mayanksoni20/topictransitiondialog

Proceedings of 3rd Workshop on Computational Approaches to Discourse, pages 92–99
October 16 – 17, 2022. ©2022 International Conference on Computational Linguistics

spoken interaction details the multiple means by which humans introduce, maintain, and change topics. Two areas which have received particular attention in the literature are topic change and topic shift. They have been defined as the point between two pieces of discourse which are considered to have different topics. Bublitz (1988) differentiates between topic change and topic shift as having low and high degrees of connectivity respectively to the previous topic. Topic shift includes both topic shading and topic fading (Maynard, 2009; Brown and Yule, 1983; Garcia and Joanette, 1997). Topic change includes reintroduction and full blown change. We annotate all such topic transitions under one common label.

3 Related Work

Related work in the literature is primarily found in the domains of *manual topic annotation* and *automatic topic segmentation*.

3.1 Manual Annotation or Segmentation

Early work to manually annotate topic transition was mainly done for the purpose of conversation analysis. Planalp and Tracy (1980) were among the first to annotate topic transition. They showed that information integration by the interlocutors impacts their topic transition strategies. Crow (1983)'s analysis of topic shift in couples' conversations showed that it occurred fairly frequently; every 48 seconds on average. Later work by Ries (2001) showed that speaker initiative and style can also be indicative of topic transition. Recently, Konigari et al. (2021) annotated a subset of the switchboard corpus (Godfrey and Holliman, 1993) into *major*, *minor* and *other* topics. Sevegnani et al. (2021) introduced a one-turn topic transition corpus by asking annotators to produce bridging sentence connecting two sentences of different topics.

3.2 Automatic Segmentation

There have been many studies to segment text based on topic or detect topic transitions. Unsupervised methods utilize annotated topic transition dataset for testing the algorithms while supervised methods train and test an algorithms on an annotated dataset. Our annotated dataset will be useful in both approaches. A detailed overview of early work is provided by Purver et al. (2011). Among the earliest relevant works is that of Reynar (1994) who proposed a method of identifying topic boundaries based on lexical cohesion and dot plots. Hearst (1997) developed an unsupervised method to separate texts into multiple paragraphs representing subtopics. Passonneau and Litman (1997) developed two algorithms that use utterance features to segment dialogue by topic. Boufaden et al. (2001) used Hidden Markov Models to segment transcriptions of telephone conversations into topics. Galley et al. (2003) tackled the difficult problem of topic segmentation in multiparty speech by focusing on the content of the transcripts and their form, *i.e.* the linguistic cues in the speech. Hsueh et al. (2006) built on the work of Galley et al. (2003) by combining Automatic Speech Recognition (ASR) with existing text based methods of topic segmentation. Arguello and Rosé (2006) also adopted a hybrid approach by combining linguistic features with local context indicators in the text. Sapru and Bourlard (2014) demonstrated that latent topic features are effective predictors of topic transition in transcripts of multiparty speech from office meetings. Joty et al. (2011) developed a supervised method of segmenting topic in email conversations. More recently, Zhang and Zhou (2019) introduced a method based on BERT and Temporal Convolution Network (TCN) . Xing and Carenini (2021) introduced an unsupervised method for topical segmentation of dialog by utterance-pair scoring. There are other relevant techniques and we skip them in the interest of brevity.

4 The Annotation Framework

We annotate 215 conversations from the Switchboard-1 Release 2 corpus (Godfrey and Holliman, 1993). Annotations are based on previous studies demonstrating that naive annotators are capable of annotating topic transition with success. (Mann et al., 1977; Passonneau and Litman, 1997; Planalp and Tracy, 1980).

Switchboard Corpus The Switchboard-1 Release 2 Corpus consists of recordings of about 2400 telephone conversations between 543 distinct speakers who did not know each other (Calhoun et al., 2010). All interlocutors spoke American English. They choose a topic from a list of about 70 topics and were connected to another interlocutor by a switchboard robot. About 50 of the 70 topics were chosen regularly. The conversation is not limited to the initial topic and participants could transition topics at any time. The individual conversation transcripts have been transcribed and annotated to the utterance level and include conversation IDs, time stamps, and label for speakers identity.

Annotation Framework 215 conversations were drawn at random from the switchboard corpus are annotated at sentence-level. The annotation were performed for start *(S)* and end *(E)* of the conversation, greeting and leave taking *(GIL)*, topic, topic transition *(C)*, and failed topic transition *(X)*. Detailed annotation guidelines can be seen in appendix D. This manually annotated corpus consists of $20,566$ turns from 215 conversation. The average number of turns per conversation is 96 with the shortest conversation lasting 33 turns and the longest conversation lasting 242 turns. Mean turns per conversation were found to be 8 and mean turns per topic were observed to be 12. The conversations were annotated by two annotators. The inter-annotator agreement (Cohen 's Kappa) obtained on a sample of five conversations is 0.64, signifying substantial agreement.

5 Empirical Studies of Topic Transitions

Having obtained an annotated corpus of 215 corpus, we conducted quantitative analysis on some aspects nature of topic change. The empirical findings are discussed in the subsections below.

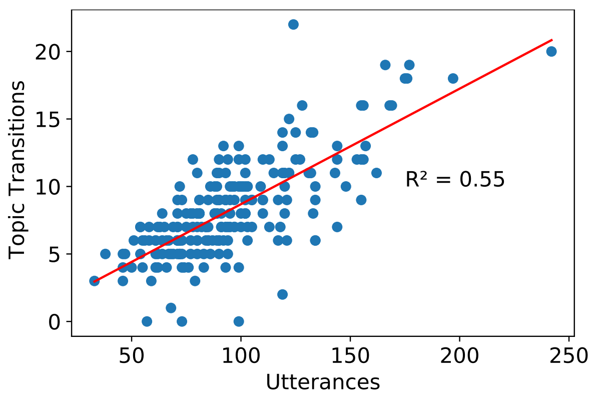

Figure 2: Scatter plot of number of topic transitions and length of conversations

Length of A Conversation and Number of Topic Transitions Longer conversation are a sign of successful and engaging conversation. We wanted to examine if longer conversation consist of more topic transitions than shorter conversation or the number of topic transitions remains similar and some topics are conversed for more turns than others. Towards investigating this relationship, we calculate number of topic transitions per conversation and plot it in Figure 2. The value of Pearson correlation coefficient is found to be 0.74, indicating a positive correlation between length of a conversation and number of topic transitions. We also plot a linear regression line and observe a R^2 value of 0.55 ($p << 0.001$). Figure 2 further highlights that number of topic transitions increase as length of a conversation increases. Most conversations consist of five to thirteen topic transitions. Thus, it is observed that longer conversations have more topic transitions.

Share of Topic Transition by Participants We wanted to explore further if the topic transitions are carried out evenly by both participant or if, one participant carries out more topic transitions. To investigate this, we first calculate the difference in number of topic transitions carried out by each participant for each conversation. We observe that only about 38% of conversations had an equal or only one more topic transition than the other per participant. In about 62% of conversations, one participant initiated at least two more topic transition than the other. It is thus observed that topic transitions are unequally carried out between participants ($\tilde{\chi}^2 = 403.41, p << 0.005$).

Time Series Analysis of Topic transition Next, the study investigate the distribution of utterances per topic as the conversation progresses. Mean and standard deviation of turns/topic is computed for all conversations. It is observed that standard deviation from mean of number of utterances is significant for all topics within a conversation. Hence, we use median to construct a line chart as median is a better measure of central tendency when there are outliers. The correlation between topic time series and number of utterances is observed to be 0.21 signifying only a weak correlation. Figure 3 shows a line plot of the number of turns per topic across the manually annotated dataset. Thus, this study did not find any pattern topic transition time series and number of utterances (($\tilde{\chi}^2 = 11.27, p = 0.98$)).

6 Modelling Topic Transition

In addition to the empirical studies performed, we also modelled topic transition on the manually annotated switchboard corpus, described in section 4. Before describing the modelling in detail, we briefly describe the approaches to model topic transition in literature.

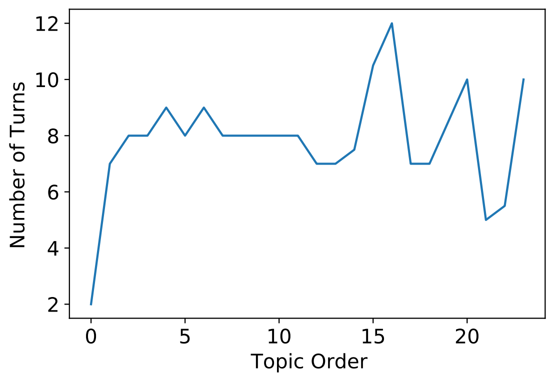

Figure 3: Line plot of turns/topic in a conversation

Approaches to topical segmentation in dialogue include unsupervised and supervised methods. Unsupervised algorithm work on finding similarity or dissimilarity between segments of text, TextTiling (Hearst, 1997) is a seminal work in unsupervised topic segmentation. Supervised approaches work with hand-crafted features or deep learning based methods such as used (Arguello and Rosé, 2006; Xing and Carenini, 2021; Konigari et al., 2021). Following the related research work, we formulate topic transition turn detection as a binary classification problem. We implement TextTiling Hearst (1997) as a baseline and then proceed to implement classical machine learning as well as deep learning based classification algorithms

Since, dialogue is inherently context based *i.e.* the next utterance is influenced by previous utterances and a topic can span across multiple turns, consecutive utterances are grouped by speaker and termed *turn*.

TextTiling (Hearst, 1997) is implemented (employing the code from NLTK (Bird et al., 2009)). Turns are formatted as paragraphs separated by two line breaks ($\backslash n \backslash n$) as required by TextTiling which works with Lexical Cohesion. The last turn of a paragraph, obtained from Texttiling, is labelled as topic transition turn and all other turns are labelled as topic continuation turns. Additionally, as classic machine learning classifiers, Naive Bayes and LightGBM are implemented. Finally, utilizing modern deep-learning based classification algorithms, XLNet (Yang et al., 2019) is implemented using Hugging Face's Transformers (Wolf et al., 2019).

Results and Error Analysis Results in table 1 show that turns where topic transitions occur can be differentiated from turns where topics are continued. Evaluation is performed on a test set which is a subset of annotated switchboard corpus (described in section 4.2). It is observed from this study that TextTiling (Hearst, 1997) is more suitable for expository text since it works with lexical cohesion and requires input text to be in paragraphs, which is a property of expository text and not necessarily of a text conversation. Previous studies Konigari et al. (2021) have also demonstrated that TextTiling (Hearst, 1997) is more suitable for text with clearly defined topics. In terms of precision, LightGBM performs better than other algorithms with a precision of 0.91. In terms of recall and f1 score, XLNet-base performs better than other algorithms. XLNet is state-of-the-art in text classification tasks (Minaee et al., 2020). XLNet-base is fine-tuned with 4 epochs using AdamW (Adam with weight decay) optimizer with Learning Rate of $1e-5$. More than 4 epochs reduce the train error rate but the difference in valid and train error rate increases. The fine-tuning was done on a single GPU. One epoch took about 28 minutes to complete. The performance of algorithms is evaluate against macro averaged precision, recall and f1 score. Precision is a metric indicating how accurately topic transition turn is detected and the values obtained can be seen in Table 2.

Model	Precision	Recall	F1
Naive Bayes	0.55	0.57	0.40
LightGBM	**0.91**	0.50	0.46
TextTiling	0.58	0.59	0.58
XLNet-base	0.68	**0.61**	**0.62**

Table 1: Evaluation scores for various algorithms on test set

7 Limitations and Future Work

Future work will include the application of insights derived from empirical studies to apply them in open-domain dialogue systems such as using the topic transition trained to re-rank responses on topicality. A limitation of this work is the inter-annotator agreement could only be obtained on a small sample of conversation. Future work will include obtaining inter-annotator agreement for all 215 Switchboard.

8 Conclusion

Empirical study on how participants engage in topic transitions in a dyad is presented. It is observed that longer conversations have more topic transitions, topic transition is generally carried out more by one participant and there is no particular pattern observed in time series of topic transition. This study was also able to detect topic transition in dialogue with 91% precision.

Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (D-REAL) under Grant No. 18/CRT/6224 and Science Foundation Ireland (SFI) under Grant Agreement No. 13/RC/2106 at the ADAPT SFI Research Centre at Trinity College Dublin. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant Number 13/RC/2106. We would like to thank anonymous reviewers from CODI 2022 for their valuable comments.

References

Jaime Arguello and Carolyn Rosé. 2006. Topic-segmentation of dialogue. In *Proceedings of the Analyzing Conversations in Text and Speech*, page 42–49.

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*, 1st edition. O'Reilly Media, Inc.

Francesca Bonin, Nick Campbell, and Carl Vogel. 2012. Laughter and topic changes: Temporal distribution and information flow. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 53–58. IEEE.

Narjès Boufaden, Guy Lapalme, and Yoshua Bengio. 2001. Topic segmentation: A first stage to dialog-based information extraction. In *In Natural Language Processing Pacific Rim Symposium, NLPRS'01*. Citeseer.

Gillian Brown and George Yule. 1983. *Discourse Analysis*. Cambridge University Press. Google-Books-ID: ZUnEAgAAQBAJ.

Wolfram Bublitz. 1988. *Supportive Fellow-speakers and Cooperative Conversations: Discourse Topics and Topical Actions, Participant Roles and "Recipient Action" in a Particular Type of Everyday Conversation*. John Benjamins Publishing. Google-Books-ID: d85TIjf7odQC.

Sasha Calhoun, Jean Carletta, Jason M. Brenier, Neil Mayo, Dan Jurafsky, Mark Steedman, and David Beaver. 2010. The nxt-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language resources and evaluation*, 44(4):387–419.

B Crow. 1983. *Topic shifts in couples' conversations.* SAGE Publications, Inc.

Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. 2003. Discourse segmentation of multi-party conversation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, page 562–569. Association for Computational Linguistics. Event-place: Sapporo, Japan.

Linda J. Garcia and Yves Joanette. 1997. Analysis of conversational topic shifts: A multiple case study. *Brain and Language*, 58(1):92–114.

Roderick Gardner. 1984. Discourse analysis: implications for language teaching, with particular reference to casual conversation. *Language Teaching*, 17(2):102–117.

John J. Godfrey and Edward Holliman. 1993. Switchboard-1 release 2 - LDC97s62 - linguistic data consortium.

Marti A Hearst. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, 23(1):33–64.

Pei-Yun Hsueh, Johanna D. Moore, and Steve Renals. 2006. Automatic segmentation of multiparty dialogue. In *11th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Shafiq Joty, Giuseppe Carenini, Gabriel Murray, and Raymond T Ng. 2011. Supervised topic segmentation of email conversations. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Rachna Konigari, Saurabh Ramola, Vijay Vardhan Alluri, and Manish Shrivastava. 2021. Topic shift detection for mixed initiative response. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 161–166.

Knud Lambrecht. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press. Google-Books-ID: bsXLCgAAQBAJ.

William C. Mann, James H. Carlisle, James A. Moore, and James A. Levin. 1977. *An Assessment of Reliability of Dialogue-Annotation Instructions*. ISI/RR-77-54.

Douglas W. Maynard. 2009. Placement of topic changes in conversation. *Semiotica*, 30(3-4):263–290.

Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2020. Deep learning based text classification: A comprehensive review. *arXiv preprint arXiv:2004.03705*.

Rebecca J. Passonneau and Diane J. Litman. 1997. Discourse segmentation by human and automated means. *Comput. Linguist.*, 23(1):103–139.

Sally Planalp and Karen Tracy. 1980. Not to change the topic but...: A cognitive approach to the management of conversation. *Annals of the International Communication Association*, 4(1):237–258.

Matthew Purver, Gokhan Tur, and Rento De Mori. 2011. *Topic segmentation*, page 291–317. John Wiley ‘l&’ Sons.

Jeffrey C Reynar. 1994. An automatic method of finding topic boundaries. *arXiv preprint cmp-lg/9406017*.

Klaus Ries. 2001. Segmenting conversations by topic, initiative, and style. In *Information Retrieval Techniques for Speech Applications*, Lecture Notes in Computer Science, pages 51–66. Springer, Berlin, Heidelberg.

Marine Riou. 2015. A methodology for the identification of topic transitions in interaction. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, (16).

Ashtosh Sapru and Hervé Bourlard. 2014. Detecting speaker roles and topic changes in multiparty conversations using latent topic models. In *INTERSPEECH*, page 2882–2886.

Karin Sevegnani, David M Howcroft, Ioannis Konstas, and Verena Rieser. 2021. Otters: One-turn topic transitions for open-domain dialogue. *arXiv preprint arXiv:2105.13710*.

Teun A. Van Dijk. 1977a. Sentence topic and discourse topic. *Papers in slavic philology*, 1:49–61.

Teun A Van Dijk. 1977b. Sentence topic and discourse topic. *Papers in Slavic philology*, 1(1977):49–61.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *ArXiv*, pages arXiv–1910.

Linzi Xing and Giuseppe Carenini. 2021. Improving unsupervised dialogue topic segmentation with utterance-pair coherence scoring. *arXiv preprint arXiv:2106.06719*.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding.

Leilan Zhang and Qiang Zhou. 2019. Topic segmentation for dialogue stream. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1036–1043. IEEE.

A Utterance Count Per Topic

In addition to plotting median utterances per topic, we also plot mean, minimum and maximum number of utterances as topic order progresses.

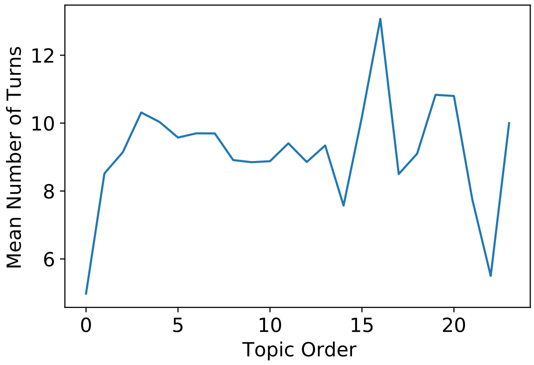

Figure 4: Line plot of mean utterances per topic

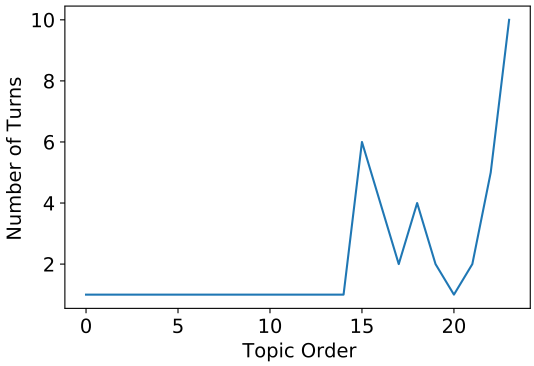

Figure 5: Line plot of minimum utterances per topic

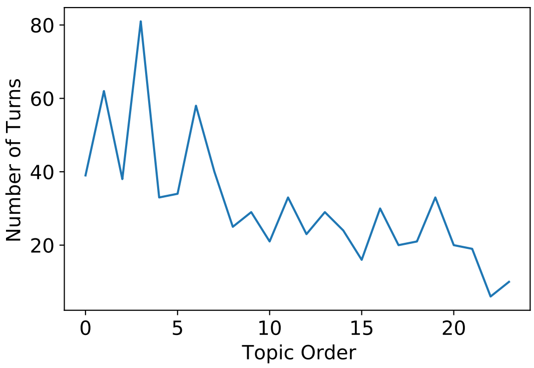

Figure 6: Line plot of maximum utterances per topic

B Share of topics by participants

Below we plot difference of topic transitions per participants across conversations. Figure below shows a bar plot of topic transition difference and percentage of conversations with such difference.

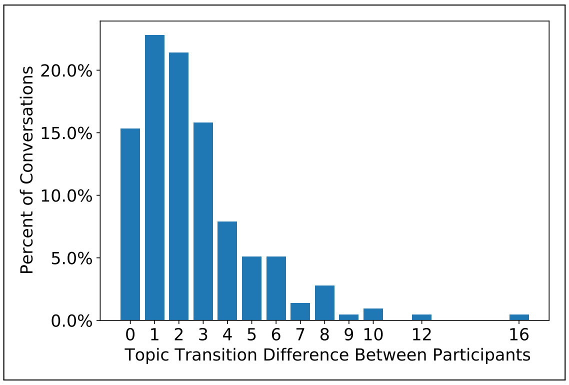

Figure 7: Share of topics per participants across conversations.

C t-SNE Visualizations

To empirically understand the separation of topic transition turns and topic continuation turns, we visualize the two classes using a t-SNE plot.

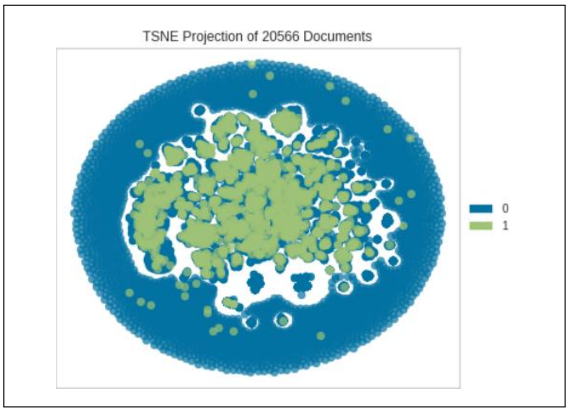

Figure 8: t-SNE visualization of utterances

D Annotation Guidelines

For a conversation, first, a topic is identified and then the topic transition is marked. For some conversation, it could be more difficult to mark the topic transition and may require reading the whole conversation.

start of topic: The first utterance, pertinent to a conversation, is marked as 's'. Here the first utterance is "[noise]" and therefore not pertinent. But the next line, a topic is introduced and therefore pertinent. This will be the starting a point for the conversation so will be marked with an "s". Non-pertinent utterances include greetings/introductions and leave-taking (GIL) as this is not the focus of this part of the project.

topic transition: This is the point when a new topic is introduced. For example, if Speaker B

H	I
"[noise]"	
"{alrighty} uh i guess our topic today is air pollution and we are to just discuss what substances do you think contribute most to air pollution"	s
"as well as what society can do to improve the air quality of the atmosphere around us"	
"right"	

introduces a new topic and then speaker A complies with the change in topic by either contributing or acknowledging the change in topic. This point of topic shift/change is marked with a "c". Here is an annotated example.

H	I	J
"yeah"		recycling
"or it ends up costing a lot more to recycle it than it does just to make it so "		recycling
"yeah [noise] well uh do you have a recycling program where you live"	c	recycling program
"oh yeah uh i go to Indiana University here and its real big and all the buildings are recycling bins for all kinds of materials"		recycling program

This example shows the point of topic transition. This can be seen when the point of the conversation changes from being about "recycling" to being about "recycling programs". This is then marked with a "c".

"well reduce reuse and recycle right [laughter]"		recyclable products
"right [laughter] words to live by"		recyclable products
"[laughter] that's right i try to remember them_1 i guess we've covered it"	e	recyclable products
"[laughter]"		gil
"okay"		gil
"[noise] it was nice talking to you bye-bye now"		gil
"you too bye-bye"		gil
"[noise]"		na

end of a topic: We also denote the end of a topic. This is like beginning the topic where utterances may not be pertinent. When marking the end of the topic, it is marked with an "e" on the last utterance pertinent to the current topic. Here is an example.