Table 1: Examples of utterances in Alcoholism Treatment. "I" stands for interventionist and "P" for patient. MISC codes are provided in the third column. P and I codes are coded following MISC. Change talk (CT), sustain talk (ST), and follow neutral (FN) codes are also provided.

Role	Conversation	MISC
I	Maybe you could tell me a little bit about what you do on the weekends, what your weekends have been like.	quo
P	Well we go out, but before we go out we just drink in the dorm room.	FN
I	Has this sort of changed your thinking, are things different than they were when you came in?	quc
P	I mean, I feel guilty about drinking,	CT (o+3)
I	Yeah. So it feels like, or it sounds like social social drinking is a big part of how you meet other people.	res
P	It's just, like ...and I don't mean to sound mean, but about the kids who don't drink, and people think that, "Oh, the kids who dont drink are losers".	ST (o-3)

annotation codes into with three categories: "CT": *Change talk* indicates utterances that reflect motivating factors related to change; "ST": *Sustain talk* indicate the patient has no intentions to change; "FN": *Follow neutral* means there is no indication of patient inclination. An example conversation snippet, highlighting all three sources of information is provided in Table 1. The intention labels (o+3, o-3) are only available for patients, whose '+' and '-' refer to change vs sustain talk (CT vs ST) and the number measures the "strength of client language," which represents a subjective assessment by human annotators, and the 'quo' and 'quc' refer to "open question" and "closed questions", which are only for interventionist (see (Borsari et al., 2015) for details regarding the coding strategy). While the MISC codes of client utterances within MISC are more complex and comprise other types of annotations, we focus on human intention modeling (i.e., CT vs. ST vs. FN) only.

How the theme of dialogue shift overtime?

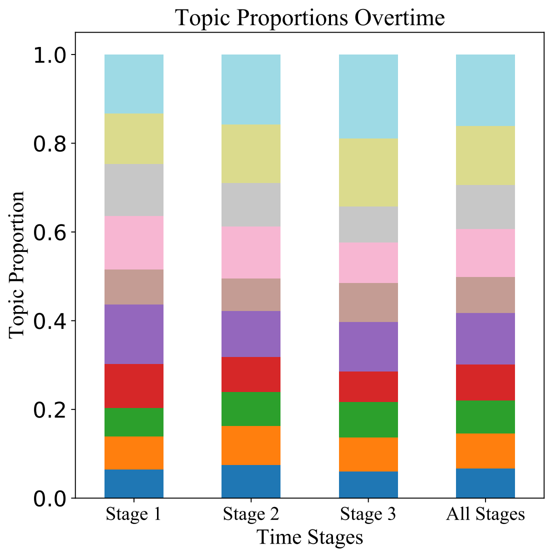

Figure 1: 10 topic proportions of patient's utterance across time stages.

We qualitatively examined how the distribution of content changes across different time stages. To measure the distribution of content, we trained a topic model with 10 topics using Gensim (Řehůřek and Sojka, 2010) with default parameters. The data doesn't have associated timestamps, thus we empirically split each MI transcript by the number of patient utterances equally into three time stages, stage 1, stage 2 and stage 3. We calculated the proportion of each topic within the same time period by take the average of all transcripts. We then normalized the topic distributions and finally visualize the extent to which distributions of the 10 topics varies by time.

We can observe the varied topic distributions across different stages of conversations, where the topic distributions are plotted from the bottom to the top. There are some topics have more variations, such as topic 4, and some topics are very stable such as topic 1 [1]. Recent research shows the performance of classification tasks might be impacted by the temporal character of language (Huang and Paul, 2018). Thus, it might be desirable to model the temporality in the computational classifiers.

[1]The 5 top words of topic 4 and 1: yeah, go, friends, know, people; beer, alcohol, games, meeting, playing.