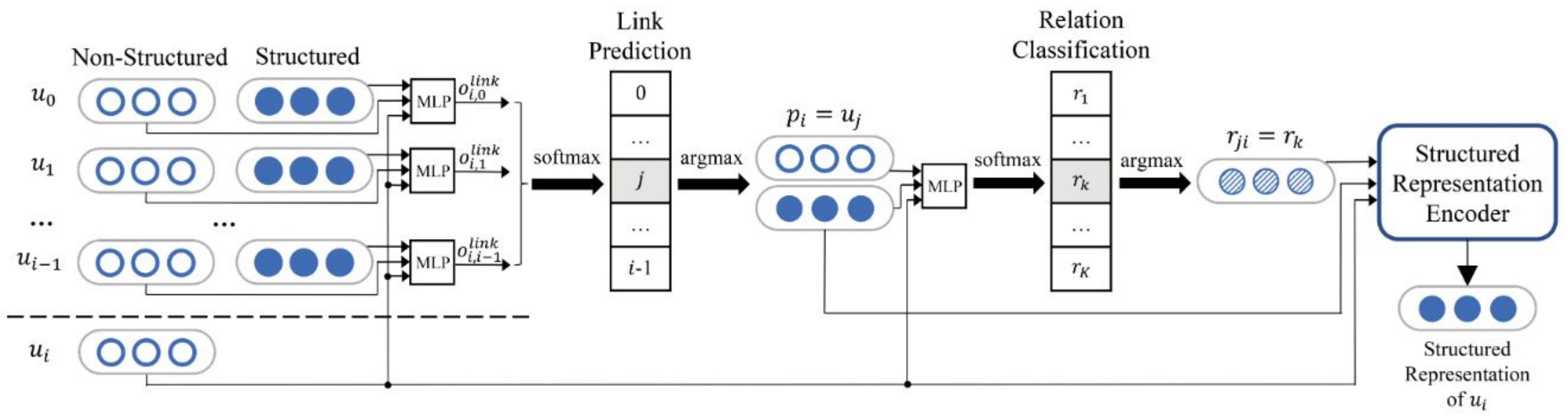
What are you saying? - Dialogue act annotation

Albert Millert, Anar Yeginbergenova



Deep sequential model for discourse parsing on multi-party dialogues

Classification task is based on the notion of ***Elementary Discourse Units*** - utterance being sequence of clause-like units; there are two common classification tasks considered: ***Link prediction*** which is a prediction of the relation between two *EDUs*; and ***relation classification*** is a prediction of the relation's type. Joint prediction of the two - ***link & relation type prediction*** provides an abstract structure of discourse.

The main objective of the conducted research was to investigate the influence of the different criteria on the overall performance of the ***Deep Sequential Model***, specifically developed for the ***STAC*** research of gamers' conversations in the act of exchanging goods and negotiating. As the dataset representative of the primary domain of discourse, we have used the ***DAIC*** dataset. This dataset does not contain punctuation and is an interview between two participants exchanging the speakership in the act of dialogue discourse. We approached the problem of investigating whether the model is capable of representing knowledge in a naive yet universal manner.

Dataset Sizes	Dialogues	Utterances	Relations	Punctuation
STAC (NP)	1026	11432	11109	YES (NO)
Molweni (NP)	9000	79487	70452	YES (NO)
STAC x Molweni (NP)	1026	90919	81561	YES (NO)
DAIC cont full	188	47153	25780	NO

Dataset Sizes	Dialogues	Utterances	Relations	Punctuation
STAC (NP)	111	1156	1126	YES (NO)
Molweni (NP)	500	4430	3911	YES (NO)
STAC x Molweni (NP)	611	5586	5037	YES (NO)
DAIC cont short	10	2563	1467	NO

Types of used corpora and their sizes

In the ***DAIC*** dataset, no interview with a patient's share under 50% exceeded the length of 200 turns in total, indicating that shorter interviews have a higher chance of having been conducted with a bit less talkative patient (turn-wise). Shorter interviews (turn-wise) correspond to a lower share of patient's speakership in the whole interview, while longer - patient's speakership share tends to be higher. On average, the share of patient's speakership in the interview is close to 60% (~140 turns), while the average interview consists of roughly 230 turns.

The average token's length observed in the patients' turns is 3.6 long. Words of lengths 4, 2, 3, 5 have the biggest share among other word lengths. 4-character words make up 23.78%, 2-character - 23.26%, 3-character - 19.84%, 5-character - 4.66%. This group of the most common words' lengths altogether makes up roughly 72% of all the tokens. The average amount of tokens within a single patients' turn is 9.56, with a minimum value of 1 and a maximum of - 125. The shorter the turn is, the more probable it is to occur in patients' utterances. Single token utterances make up to 19.99%, 2-token - 9.19%, 3-token - 7.21%, 4-token - 6.05%. It is important to note that most of the single-token turns seem to be responses to yes/no-questions or - backchannels (encouragements making speaker keep talking).

Discourse Representation Theory considers sequence of sentences; examination of how the representation of new discourse units affects already observed data; construction of a logical representation; two assumptions: 1) Hearer builds the mental representation of sentences; 2) Each consecutive sentence is an addition to the representation.

Rhetorical Structure Theory emphasizes representation learning by transforming surface features into a latent space; allows to jointly learn a projection of the surface features with parsing the discourse.

Train \Test	STAC	STAC NP	Molweni	Molweni NP	S x M	S x M NP	DAIC full	DAIC short
STAC	**47.733**	43.962	24.470	18.736	25.984	21.150	17.831	17.142
STAC NP	12.954	**45.700**	16.298	16.411	18.839	19.035	3.077	2.770
Molweni	19.975	15.545	**55.184**	24.695	42.460	33.858	9.198	10.769
Molweni NP	17.635	17.300	37.494	**45.676**	33.467	35.493	10.990	11.471
STAC x Molweni	31.509	26.828	20.880	21.061	**51.910**	35.386	25.468	27.117
STAC x Molweni NP	31.676	34.099	19.413	19.458	18.733	**44.633**	12.862	13.422

Train \Test	STAC	STAC NP	Molweni	Molweni NP	S x M	S x M NP	DAIC full	DAIC short
STAC	**71.515**	68.199	53.860	51.716	57.283	55.221	45.025	42.731
STAC NP	**71.291**	71.017	63.138	62.619	64.872	64.552	46.669	48.537
Molweni	43.544	44.964	**86.612**	75.643	68.657	69.119	36.691	38.978
Molweni NP	43.711	42.791	75.395	**86.080**	68.657	69.173	32.322	34.452
STAC x Molweni	71.041	69.118	77.652	77.111	**84.254**	75.411	45.991	48.030
STAC x Molweni NP	69.536	70.455	74.199	75.102	73.207	**83.932**	45.675	48.381

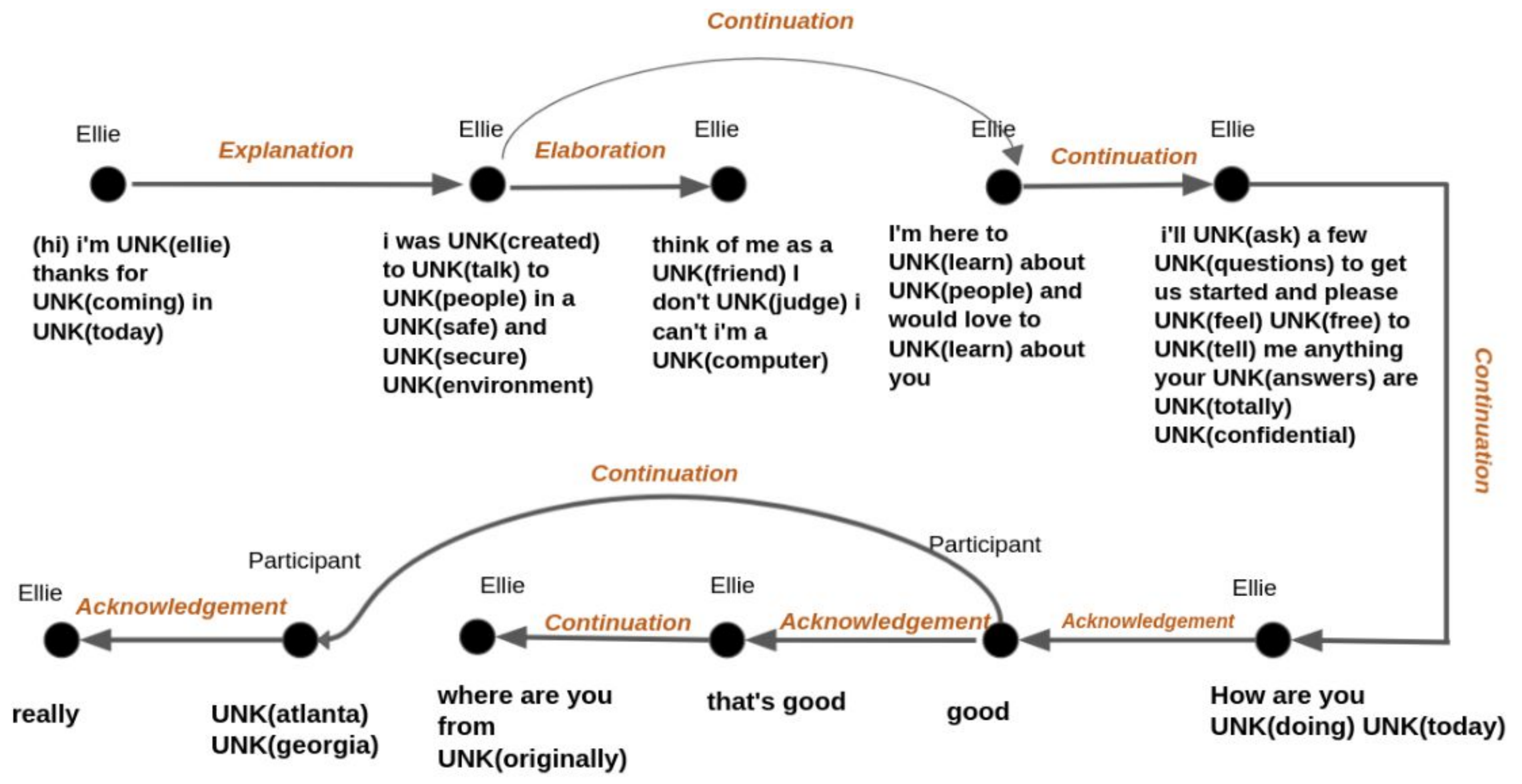
Types of used corpora and their sizes

Token	Tokens share in the category %
um	25.56
yeah	8.16
no	8.1
uh	7.35
yes	6.83
<laughter>	4.45
mhm	3.53
so	2.78
mm	2.55
okay	1.91

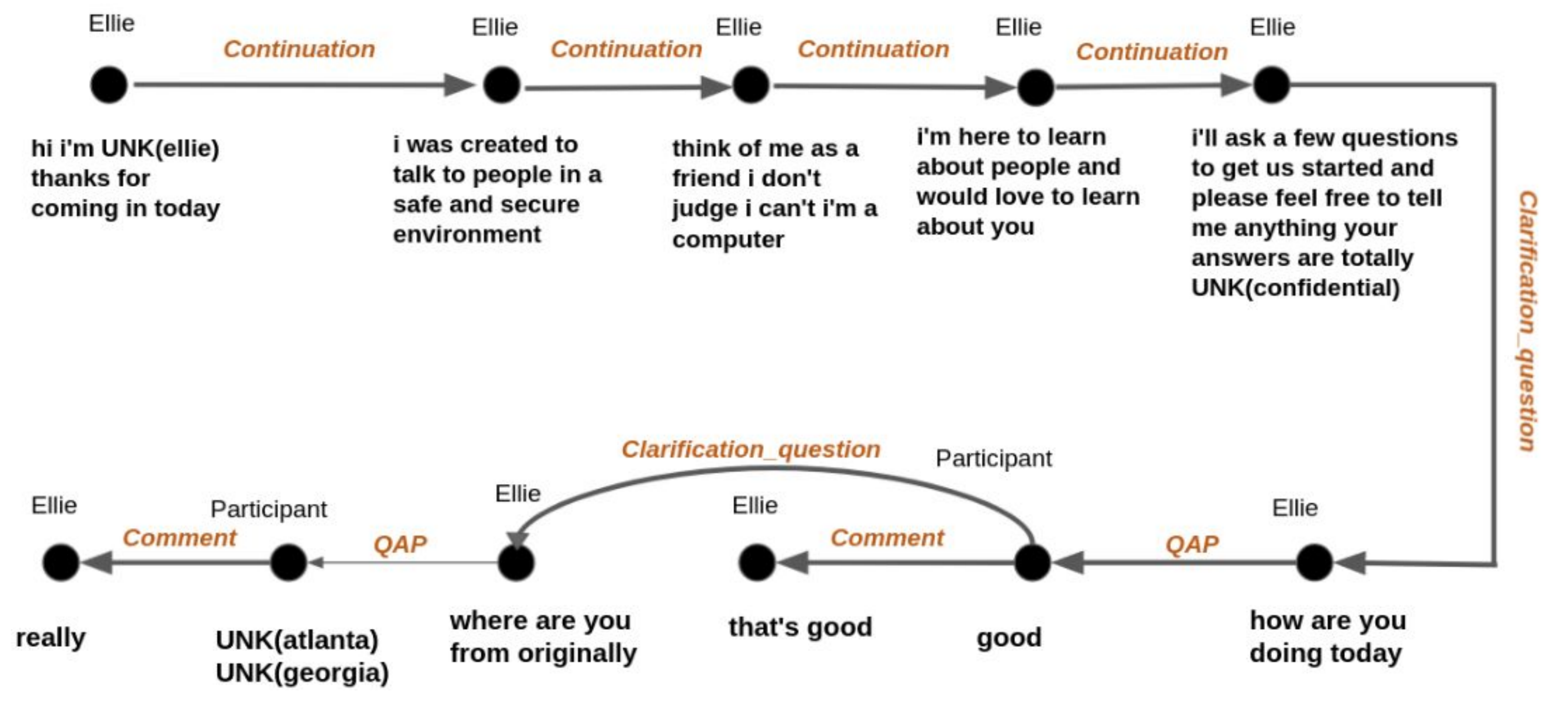
The most frequent single-token utterances in *DAIC* dataset

Segmented Discourse Representation Theory follows the motivation of *DRT* and adds discourse coherence theories; 16 possible relations' types: *Question-answer pair, Comment, Question Elaboration, Acknowledgement, Elaboration, Alternation, Explanation, Result, Continuation, Parallel, Correction, Conditional, Contrast, Clarification question, Narration, Background*; relation types connect the utterances, resulting in a coherent structure.

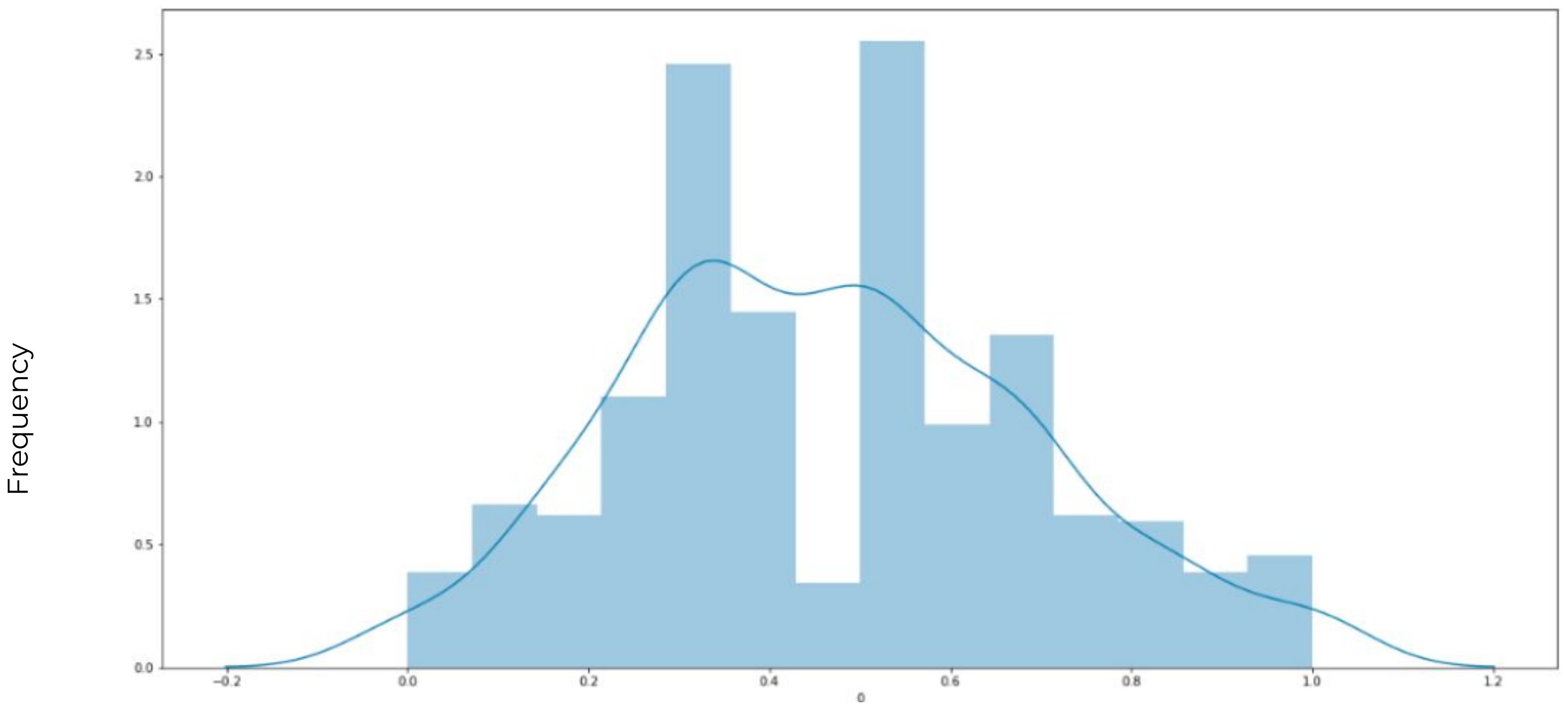
The F1 scores from test data illustrate that the predictions are very diverse and sometimes the model has highly accurate predictions, and sometimes it is lower than 0.5. It depends on the context, length, and structure of the dialogues in the corpus.


Continuation
Ellie
Explanation
Ellie
Elaboration
Ellie
Ellie
Continuation
Ellie
(hi) i'm UNK(ellie) thanks for UNK(coming) in UNK(today)
i was UNK(created) to UNK(talk) to UNK(people) in a UNK(safe) and UNK(secure) UNK(environment)
think of me as a UNK(friend) I don't UNK(judge) i can't i'm a UNK(computer)
I'm here to UNK(learn) about UNK(people) and would love to UNK(learn) about you
i'll UNK(ask) a few UNK(questions) to get us started and please UNK(feel) UNK(free) to UNK(tell) me anything your UNK(answers) are UNK(totally) UNK(confidential)
Continuation
Continuation
Participant
Participant
Ellie
Acknowledgement
Ellie
Continuation
Ellie
Acknowledgement
Acknowledgement
Ellie
really
UNK(atlanta) UNK(georgia)
where are you from UNK(originally)
that's good
good
How are you UNK(doing) UNK(today)

Predictions of the model trained on STAC corpusc


Ellie
Continuation
Ellie
Continuation
Ellie
Continuation
Ellie
Continuation
Ellie
hi i'm UNK(ellie) thanks for coming in today
i was created to talk to people in a safe and secure environment
think of me as a friend i don't judge i can't i'm a computer
i'm here to learn about people and would love to learn about you
i'll ask a few questions to get us started and please feel free to tell me anything your answers are totally UNK(confidential)
Clarification_question
Clarification_question
Participant
Ellie
Participant
Comment
QAP
Ellie
Ellie
Comment
QAP
Ellie
really
UNK(atlanta) UNK(georgia)
where are you from originally
that's good
good
how are you doing today

Predictions of the model trained on Molweni corpus



F1 score of each dialogue in test data

For the ***STAC*** dataset, the length of the utterances was short (on average), compared to the ***Molweni***. The average length of the utterance in ***STAC*** data is 3.3, whereas in Molweni this number equals 10.8. Hence, the ***STAC*** model performed worse when tested on ***Molweni*** because the model never learned to classify long sentences. On the other hand, the Molweni-trained model worked relatively good when tested against the long data and slightly worse on the short ones. Another problem of ***STAC*** is that it has an extremely limited vocabulary compared to the other dataset. It was produced in the gaming environment where the interactions were in shortened form. Whereas on ***Molweni,*** all the sentences are constructed fully in order to let the addressee understand the inquiry.