Annotating the discourse and dialogue structure of SMS message conversations

Nianwen Xue, Qishen Su, Sooyoung Jeong
Brandeis University
Computer Science Department
{xuen,qsu,jeong}@brandeis.edu

Abstract

In this paper we present a framework for annotating the discourse and dialogue structure of SMS message conversations. The annotation specifications integrate elements of coherence-based discourse relations and communicative acts in conversational speech. We present annotation experiments that show reliable annotation can be achieved with this annotation framework.

1 Introduction

With the pervasive use of mobile devices, Short Message Service (SMS) has been used widely in day-to-day communications. In many cases SMS messages have taken the place of traditional telephone conversations, and have become the preferred method for people to communicate with one another. SMS messages are by definition short, and due to its asynchronous nature, a participant does not have to wait to respond before another participant finishes. As a result, it is often the case that the conversation does not alternate in a rigid manner between participants.

The relations between the messages in an SMS conversation are in some ways very similar to those between utterances in conversational speech, where a conversant may agree or disagree with, respond to, or indicate understanding (or non-understanding) of an utterance by another speaker. To the extent that they are similar, the relations between SMS messages can be characterized in terms of the dialogue annotation framework described in (Core and Allen, 1997). The dialogue structure of the SMS "conversations" also tends to be more complex than that of speech conversations as a result of the more complex turn-taking patterns in SMS messages.

SMS message conversations are also different from conversational speech in that they are primarily in text form. Text within a single message also demonstrates the kind of discourse coherence that is typical of written text.

In this paper we describe a framework for annotating the discourse and dialogue structure of SMS message conversations. Based on the linguistic characteristics of SMS messages, we design an annotation framework that integrates elements of dialogue and discourse annotations, and report experiments that show reliable annotation with this framework.

The rest of the paper is organized as follows. In Section 2, we describe our annotation framework in detail. In Section 3 we report results on annotation experiments that show reliable annotation, and we will also discuss sources of disagreement. In Section 4 we discuss related work. We conclude our paper and describe future work in Section 5.

2 Annotation framework

In this section, we describe key elements of our annotation framework. We first describe basic units of our annotation, and then discuss how the basic units relate to each other to form a dialogue structure. Finally we present the set of relations we use in interpreting this structure.

2.1 Units of annotation

The basic units of annotation are individual text messages. The SMS messages are usually short, and most of the messages consist of single sentences, but there are a small and yet significant proportion of messages that consist of multiple sentences. In our current round of annotation, we do not analyze relations between the sentences inside one message, but we leave that possibility open for future rounds of annotation. Compared

Proceedings of LAW X – The 10th Linguistic Annotation Workshop, pages 180–187,
Berlin, Germany, August 11, 2016.

with discourse annotation of newswire text (Carlson et al., 2001; Prasad et al., 2008), determining the text units to perform annotation on is a relatively simple task, due to the fact that there is a natural boundary between text messages.

2.2 Structure of the SMS message conversations

Due to the asynchronous nature of SMS message conversations, individual messages are often "out of order", and determining which message relates to which is a substantial part of the annotation. This aspect of the annotation is different from the annotation of newswire texts or even conversational speech, where the "normal" order is generally maintained, although in conversation speech, there are often interruptions that break the normal pattern of turn-taking (Stolcke et al., 2000). Although there are some exceptions, in general, we assume that one message is only related to one previous message.[1] We call the message we are annotating the "anaphor", and the previous message that it relates to the "antecedent". Because the messages are "scrambled", the antecedent of a message is not always the immediately previous one, although it is in most cases. In addition, the antecedent of a message may not always be from a different participant. A participant may respond to a prior message by another participant, or continue his/her own line of thought without responding to an outstanding message from the other participant. A short snippet of an SMS message conversation is presented in Figure 1. On the left side of the figure is a graph that shows how the messages are connected. Each message is identified by a numerical number followed by a letter indicating the ID of the participant. For example, "7b" indicates message No 7 by participant "b". As should be clear from the graph, some messages (e.g., 7b,12b,14a, 15b,16a, 17b) are not linked to an immediately previous message, and some messages are connected to a previous message by the same participant. The graph shares many properties of a dependency tree in that there is a single root, and each anaphor is connected to one antecedent. It also more constrained than a dependency tree at the syntactic level in that the antecedent is always before the anaphor. The dependency tree is non-projective, since if all the arcs are drawn one one side, there will be crossing edges. These properties are important in fashioning a strategy for parsing this structure automatically, a topic that is out of the scope of this paper. Linking each message to its antecedent message is the first step of our annotation project.

[1]The assumption always holds except for a negligible number of cases where one message responds to multiple previous messages.

2.3 Relations between the messages

The second aspect of our annotation is to label the edges in graph, that is, to determine the relationship between each pair of connected messages. When annotating these relations, we make the distinction between same-participant message pairs and different-participant message pairs. The relations we use to label same-participant message pairs are drawn from the discourse relations defined in the Penn Discourse TreeBank (Prasad et al., 2008), but some PDTB relations are nonexistent in the SMS data. For example, we did not find cases of temporal relations in our SMS conversation data. This makes senses, since there is not much narrative text in SMS messages as there is in newswire such as Wall Street Journal articles in the PDTB and as a result, temporal relations are mostly unnecessary. On the other hand, there are also relations not covered in the PDTB. For example, there are cases where a participant uses another message to complete a previous message, presumably because s/he hit the "send" button in the middle of a message and later had to complete that message. There are also messages used to correct spelling mistakes of a previous message from the same participant. Such cases are not attested in carefully edited newswire text but they need to be accounted for in our annotation. The complete list of same-participant relations are presented in Table 3.

The different-participant relations are drawn from DAMSL (Core and Allen, 1997), a coding scheme for annotating communication acts in conversational speech. DAMSL is a multilayer annotation framework that annotates both forward and backward communicative functions. Since we focus on the relation between the current message and its antecedent, we limit ourselves to mostly annotating backward communicative functions. The set of different-participant relations are provided in Table 2. Two of our labels, *directives:request* and *directives:suggestion* may bear some resemblance to the forward communicative functions in DAMSL, but they are used to label

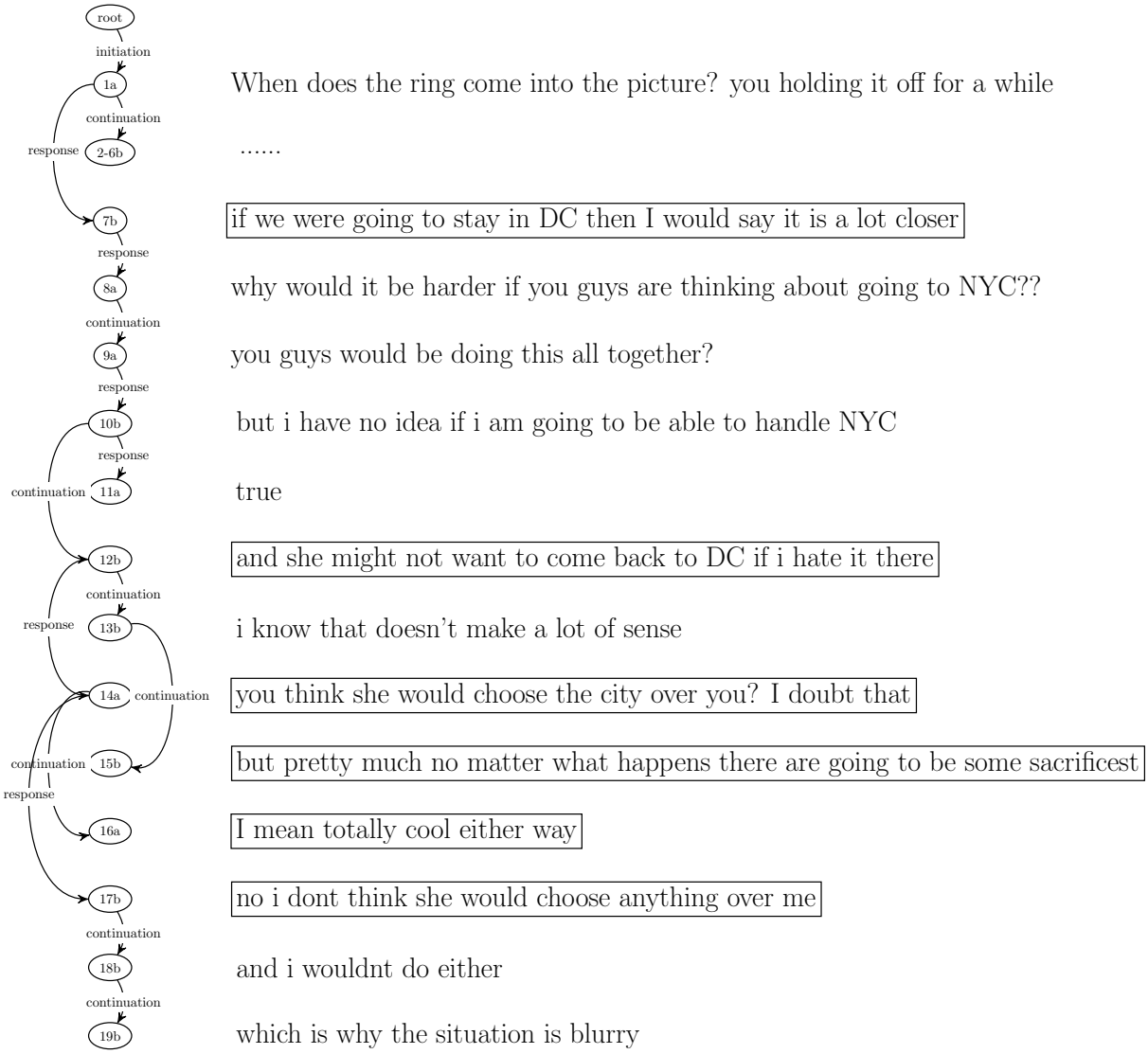

Figure 1: The SMS message as a dependency tree. The suffixes "a" and "b" on the tree nodes are the two participants. Messages in boxes have non-local (not immediately before) antecedents.

requests or suggestions in the context of a previous message. The following example is a case of *directive:suggestion*:

(1) A: I'm hungry.

B: let's go get some food!

It is important to note that unlike DAMSL, the targets of our annotation project are not individual utterances but are relations between pairs of messages. When labeling the backward communicative functions of an utterance in DAMSL, the antecedent of the utterance is assumed to be the immediately previous one, but we cannot make this assumption in our annotation.

There is a third group of labels that don't fit nicely into either group of same-participant or different-participant labels. Those labels are used to label messages that initiate a new topic, get attention, or fulfill a social obligation. These messages are explained in Table 4.

3 Annotation Experiments

The SMS data we performed our annotation experiments on are drawn from an LDC collection of SMS and Chat Messages collected under the DARPA BOLT program. Two annotators performed four rounds of annotation, working on the same documents so that inter-annotator agreement (IAA) statistics can be computed. We started with an initial set of guidelines. After each round of annotation, the annotators met and discussed cases of disagreement. If the differences are due to un-

Label	Description
Agreement:Acceptance	*Acceptance* refers to a positive response to proposals, requests, and suggestions, or agreement to assertions. Common key words of acceptance are "yes", "ok", "alright", etc.
Agreement:Rejection	*Rejection* indicates a negative response to proposals, requests, and suggestions, or disagreement to assertions. Rejection is often signaled by words like "no" or "nah".
Understanding:Acknowledgment	*Acknowledgment* signals a participant's understanding of a previous message. Cue words or phrases for *Acknowledgment* include "ok", "I understand", "yes", "I know", "I see", etc. *Acknowledgment* may also contain words or short phrases that express sentiment such as happiness, excitement, sadness, anger. These words or phrases can be laughing words (such as "haha" and "lol"), words that express surprise or excitement (such as "omg" or "yay") and appreciation (such as "awww"), profanity (such as "what the hell"), or emoticons.
Understanding:Non-Understanding	*Non-understanding* is used when a participant seeks clarification by asking clarification questions.
Directive:Request	This relation is used when a participant asks another participant to perform certain action. The immediate information or context of *Request*, as opposed to *Derivative Request*, comes from the other participant's message.
Directive:Suggestion	This relations is used when a participant offers another participant an idea or plan for consideration. The immediate information or context of *Suggestion*, as opposed to *Derivative Suggestion*, comes from the other participant's message.
Question	This relation is used to mark requests of information and clarification. Unlike the clarification questions mentioned previously, this type of question does not signal non-understanding. Instead it is a general request for additional information. The immediate information or context of *Question*, as opposed to *Derivative Question*, comes from the other participant's message.
Answer:Answer	An answer provides complete or partial information to a question in a previous text message.
Answer:Hold	A participant sometimes signals their acknowledgment of a question, but does not provide an answer to it. Moreover, if a participant responds to another participant's question with a question, such a response is considered as *Hold*
Feedback	This type of relation is used when a participant provides information in response to another participant's message that is neither a question nor a directive.

Figure 2: Dialogue Only Labels

clear instructions in the guidelines or unclear distinctions in the tagset, the guidelines are revised before the next round of annotation starts. We made sure that the document sizes and the number of messages that we annotate in each round stay constant so that we can observe the trend in the agreement statistics after each round of annotation. Before we discuss the IAA, we first present the distribution of the distances between each message and its antecedent in Table 5. The distance is computed by pooling the two sets of annotations by the two annotators. The results show that overall there is a distance of 1 for only 77.97% of the message pairs, meaning that the antecedent mes-

Label	Description
Contingency:Cause	*Cause* indicates that the situations in two text messages influence each other causally, and they are not in a conditional relation.(Group, 2008) This type of relation is used when the argument of the previous message is the result, and that of the following message is the cause.
Contingency:Result	Similar to the *Cause* relation, Result also indicates that the two arguments have a causal relation, and that they are not in a conditional relation. *Result* is used when the argument of a given message is the result caused by the situation of a previous message.
Contingency:Condition	Two text messages are in a conditional relation when the argument of one message is the condition and that of the other message is the consequence.
Expansion:Elaboration	A text message is considered as an elaboration of a previous one, when the current message clarifies or elaborates on the information that the previous message conveys. This relation can apply to two or more messages that are connected by conjunctions "and" and "but".
Expansion:Derivative Question	This type of relation concerns with requests of information and clarification, similar to *Question*. However, the immediate information or context of *Derivative Question*, as opposed to *Question*, derives from the same participant's own messages.
Expansion:Derivative Suggestion	This type of relation is used when a participant provides another participant an idea or plan for consideration of a future action, and its information or context derives immediately from the same participant's own messages.
Expansion:Derivative Request	This elation is used when a participant asks another participant to perform certain action, but its immediate information or context derives from the same participant's messages.
Expansion:Concession	This type of discourse relation is used to highlight prominent differences between two text messages. More specifically, "the highlighted differences are related to expectations raised by one argument which are then denied by the other"(Group, 2008).
Expansion:Alternative	This discourse relation is used when two text messages describe alternative situations. 'or", "instead" and "otherwise" are common cue words for this relation.
Expansion:Completion	Occasionally when a participant uses two or more messages to complete a sentence, and *Completion* is used to describe the relation between these messages.
Reflexive Feedback	This relation is used when a participant answers their own questions or responds to their own statements (such as laughing at their own joke).
Correction	*Correction* is generally concerned with correcting wrong information from a previous text message, such as typos.

Figure 3: Non-Dialogue Only Labels

sage is the immediately previous message in only 77.97% of the cases. For the remaining 22.03% of the cases, the antecedent is not the immediately previous message, indicating there is a significant proportion of messages that do not follow the "normal" order of turn-taking. The amount

Label	Description
Topic Introduction	It is used when a participant initiates a new topic in a new or existing conversation.
Attention Getter	An *Attention Getter* is a word or phrase used to attract the attention of another participant. It can be words like "Hey", "Oh", "Ah", etc., or the name of the other speaker.
Social Obligation	This type of discourse relation is used when a participant complies with certain social norms or obligations, such as apologies, acceptance or rejection of apologies, appreciation, greetings, farewell, etc. When a participant is signaling their desire for ending a conversation, that message is considered farewell, and is thus labeled as *Social Obligation*.
Other	Occasionally, a participant might send an empty message, and in that case, the relation of the empty message to its immediate previous message should be annotated as *Other*. *Other* is also used when a given message is nonsensical in relation to any previous message, or when the relation between two messages are not formalized in any of the categories above.

Figure 4: Dialogue and Non-Dialogue Labels

of "scrambling" is even higher between different-participant message pairs, where one participant is responding to a message of another participant.

The inter-annotator agreement statistics for the four rounds of annotation are presented in Table 1. Column 4 shows the agreement on connections only, which is computed as the percentage of messages that are linked to the same antecedent for both annotators. Column 5 shows the agreement on relations, which is computed as the proportion of message pairs that are annotated with the same relation, out of the total number of connections that both annotators agree on. So this calculation factors out connections that the two annotators have disagreements on. Column 6 shows the Cohen's $Kappa$ on relation agreement. The results show the agreement on connections stays relatively stable between rounds, indicating this aspect of the annotation is rather intuitive, and does not benefit from additional rounds of training. In contrast, there is significant improvement in the agreement on relations as guidelines are refined and the distinction between the relations are clarified. The final column shows the agreement on both connections and labels. The agreements statistics are lower, indicating a cumulative effect, but overall, it shows that reliable annotation can be achieved.

The inter-annotator agreement (IAA) statistics on connections are calculated with equation 3

$$P = \frac{N_a}{N_t}$$

where N_a is the total number of same connections, and N_t is the total number of connections. The inter-annotator agreement for connections with label is calculated similarly: N_a is the total number of same connections with the same label.

The Cohen's Kappa score for labels on the same connections is calculated as follows:

$$K = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the sum of probabilities of choosing the same label, and P_e is the probability of choosing the same label by chance,

$$P_e = \sum P_i^a \times P_i^b$$

where P_i^a and P_i^b are the probabilities of annotator A and annotator B choosing label *i*, respectively. P_e is the sum of the products of P_i^a and P_i^b for all labels.

3.1 Examples of Inter-annotator Disagreement

There are two main types of disagreement between the annotators: disagreement on connections and disagreement on relations. Disagreement on connections happens when, given a message, the annotators disagree on which previous message is its antecedent. Disagreement on relations occurs when the annotators disagree on the relation between a given pair of messages.

Distance	**1**	**2**	**3**	**4 and greater**
Dialogue Links	73.30%	17.08%	5.72%	3.88%
Non-dialogue Links	84.22%	11.18%	3.22%	1.36%
Dialogue and Non-dialogue Combined	77.97%	14.56%	4.65%	2.80%

Figure 5: Distance Distributions

	Number of Files	Number of Messages	Agreement on connections	Agreement on relations	Kappa on relations	Agreement on both
Round 1	10	898	0.886	0.697	0.649	0.618
Round 2	14	873	0.886	0.722	0.680	0.640
Round 3	10	893	0.848	0.838	0.826	0.710
Round 4	10	890	0.867	0.881	0.875	0.764

Table 1: Inter-Annotator agreement statistics

Message ID	Timestamp	Participant ID	Content
m0007	2010-08-24 19:22:45 UTC	153902	Charming is the audience's subjective interpretation
m0008	2010-08-24 19:22:49 UTC	153901	so you can choose to be condescending?
m0009	2010-08-24 19:23:02 UTC	153902	Yes
m0010	2010-08-24 19:23:06 UTC	153901	but you cannot choose to be charming
m0011	2010-08-24 19:23:14 UTC	153902	You can attempt to be charming

Figure 6:

Disagreement on connections Although determining which message is connected to which previous message is intuitive for the most part, disagreement does happen when a message has more than one possible and meaningful connection. For instance, message m0010 in Figure 6 can be a response to message m0009 or an extension of message m0008. This is one of the cases on which the two annotators disagree.

Disagreement on Relations Certain words or phrases are generally ambiguous and prone to causing confusion and disagreement on labeling. For example, the word "yeah" or the phrase "I know" can either signal acknowledgment or express agreement. Disagreement on labeling often occurs when such words or phrases can be interpreted either way in a given context. Message m0053 in Figure 7 can be either acknowledgment or agreement of the assertion in their previous message, and either interpretation makes sense in this context.

4 Related work

There has been relatively little work on annotating the discourse and dialogue structure of SMS conversations. The work that is most similar to ours is that of (Perret, 2015), where they annotated the discourse structure of multi-party dialogues using a corpus collected from an on-line version of the *The Settlers of Catan* game. They argue that multi-party dialogues need to be modeled with a graph structure and adopted an annotation scheme in the SDRT framework (Asher and Lascarides, 2003). In our annotation, since we are dealing with SMS dialogues that involve two participants, we did not find a graph structure to be necessary. We opted for a simpler (non-projective) dependency structure that is easier to model algorithmically. In fact, (Perret, 2015) developed an automatic discourse parser based on the Maximum Spanning Tree, a tree-based dependency parsing algorithm (McDonald, 2006) instead of a graph-based algorithm. We also make a distinction be-

Message ID	Timestamp	Participant ID	Content
m0051	2015-02-27 13:45:13 UTC	152252	It's so stupid Sofie
m0052	2015-02-27 13:45:36 UTC	152252	I just feel like the general public should take an art class
m0053	2015-02-27 13:45:36 UTC	152212	i know

Figure 7: Disagreement on relations

tween same-participant and different-participant relations, and argue SMS message conversations need to be modeled with an annotation framework based on both discourse coherence and dialogue structures.

5 Conclusion and Future Work

In this paper we presented a framework for annotating the discourse and dialogue structure of SMS message conversations. The annotation specifications integrate elements of coherence-based discourse relations and dialogue structure in conversational speech. We conducted annotation experiments that show reliable annotation. Future work includes additional annotation based on this annotation framework and producing sufficient data that can be used to train a statistical parsing model.

Acknowledgment

We gratefully acknowledge a Faculty/Student Collaborative Research Grant from the Office of the Dean of College of Arts and Science Dean's at Brandeis University that has supported the first and third author of this paper. All errors and shortcomings are that of the authors of course.

References

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue - Volume 16*, SIGDIAL '01, pages 1–10, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mark G Core and James Allen. 1997. Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, pages 28–35. Boston, MA.

The PDTB Research Group. 2008. The PDTB 2.0. annotation manual. *Technical Report IRCS-08-01*, page 28.

Ryan McDonald. 2006. *Discriminative Training and Spanning Tree Algorithms for Dependency Parsing*. Ph.D. thesis, University of Pennsylvania.

Stergos Afantenos Eric Kow Nicholas Asher Jérémy Perret. 2015. Discourse parsing for multi-party chat dialogues. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal.

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

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.