Integrating Non-Linguistic Events into Discourse Structure

Integrating Non-Linguistic Events into Discourse Structure*

Julie Hunter
IRIT, Université Toulouse III
juliehunter@gmail.com

Nicholas Asher
IRIT, CNRS
nicholas.asher@irit.fr

Alex Lascarides
University of Edinburgh
alex@inf.ed.ac.uk

Abstract

Interpreting an utterance sometimes depends on the presence and nature of non-linguistic actions. In this paper, we motivate and develop a semantic model of embodied interaction in which the contribution that non-linguistic events make to the content of the interaction is dependent on their rhetorical connections to other actions, both linguistic and non-linguistic. We support our claims with concrete examples from a corpus of online chats, comparing annotations of the linguistic-only content against annotations in which non-linguistic events in the context are taken into account.

1 Introduction

Embodied conversation enables a speaker to refer to non-linguistic entities in the surrounding situation, using little to no descriptive content. Models of non-linguistic context dependence tend to focus on a fairly well-defined set of expressions, with their reference to entities in the non-linguistic context governed by rules provided by the lexical entry of each expression. Indexical and demonstrative expressions (*I*, *now*, *that*, ...) are the most visible examples (Kaplan, 1989). There are also numerous models of linguistic context dependence, or anaphora, tackling the anaphoric properties of lexical items like *he*, with theories of the *rhetorical structure* of discourse (e.g. Asher (1993), Asher and Lascarides (2003), Hobbs et al. (1993), Mann and Thompson (1987)) that analyze context-dependent relations between entire units of discourse. For example, the eventuality described by one unit might serve to explain the eventuality described by another, or it might stand in contrast to another, and so on.

This paper addresses the discourse interactions between linguistic and non-linguistic events and the resulting contribution of non-linguistic events to semantic content. Situated dialogue makes widespread use of the non-linguistic context, in ways that go beyond demonstrative reference. But comparatively little attention has been dedicated to modeling this interaction, even within theories of discourse. This is perhaps because doing so requires assigning semantic contents to non-linguistic eventualities in the context. Linguistically specified contents carry information about how eventualities should be individuated and conceptualized; contents that are not specified linguistically are left to interpreters to sort out in context. Suppose a waiter approaches you with a bottle of champagne and gives you a certain look. You respond "No, thank you. I'm driving." or you merely shake your head and show him your keys. You have understood that he was offering you champagne and you have coherently responded to his implicit question of whether you would like some. To react appropriately to his action, you not only had to isolate the important features in the visual scene—the look in his eyes, but not the fact that he blinked, for example—you also had to understand that these features came together to produce a *meaningful* act with a particular semantic content and you had to understand that content.

We develop a formally precise pragmatic model of linguistic and non-linguistic interactions by drawing on a corpus of chats taken from an on-line version of the game *Settlers of Catan*. The *Settlers* corpus

*This research was supported by ERC Advanced Grant n. 269427. We thank Eric Kow and reviewers for helpful comments.

Proceedings of the 11th International Conference on Computational Semantics, pages 184–194,
London, UK, April 15-17 2015.

is ideal for our task: firstly, non-linguistic events in the game (dice rolls, card plays, the building of settlements and roads, etc.) are crucial for understanding many of the comments made in the chats; and secondly, the controlled, task-oriented environment minimizes interlocutors' differences in conceptualizing non-linguistic events. The meaning of each non-linguistic event in the game is determined by the game rules and set up. For example, when a blue building appears on the game board, it is clear to all players that the player who was designated as blue at the game's start has just built a settlement on that portion of the board. Minimizing conceptualization problems allows us to examine how non-linguistic events affect the structure of discourse—in particular, how they affect the structures posited by Asher and Lascarides' (2003) *Segmented Discourse Representation Theory* (SDRT), which we adopt as our starting point. Not only do we feel that SDRT is a promising semantic model of discourse, but the *Settlers* corpus is annotated with SDRT's coherence relations. We nevertheless recognize that other approaches to the study of situated dialogue are possible, though we reserve a larger discussion for another occasion.

After introducing our corpus in §2, §3 discusses examples that both highlight the importance of non-linguistic events in our corpus and complicate the task of building discourse structures for our chats. §4 extends SDRT to handle these. §5 addresses some questions that arise from treating non-linguistic events as semantic elements, while §6 situates our project with respect to related work.

2 The *Settlers* Corpus

Settlers is a win-lose game in which players use resources (e.g. wood and sheep) to build roads and settlements. Players build on a game board that is divided into multiple regions, each associated with a certain type of resource and a number between 2 and 12. Players acquire resources in various ways: e.g., through trades with other players or the bank and through rolling two dice. A roll of a 4 and a 2, for example, gives any player with a settlement on a 6 region that region's resource. A player who rolls 7, however, moves the *robber* to a region of her choice (there is no region marked with a 7) and she can then steal from a player whose buildings are on that region. Players with settlements on the region occupied by the robber don't receive its associated resources, whatever the dice rolls.

Trades involve moves that we call *offers*, in which one player proposes to exchange particular resources with another. A successful trade involves an explicit acceptance of an offer. In the original online version of *Settlers*, trades take place non-linguistically, via mouse-clicks, etc. But in our corpus, each player was instructed to negotiate the trade via the online chat interface (see Afantenos et al. (2012) for details). In fact, players chatted not only about trades, but about many aspects of the game state, including building actions and dice rolls. The corpus consists of 59 games, and each game contains dozens of individual negotiation dialogues, each consisting of anywhere from 1 to over 30 dialogue turns. The corpus was collected in three distinct phases: a pilot phase, and two seasons of competitions, the second culminating in a "master's league," with the best players of the second season playing an eliminatory competition to choose an overall winner.

Our corpus manifests a wide variety of examples, including (1)-(3) below, in which linguistically specified contents anaphorically depend on contents of non-linguistic events, and vice versa. When we consider the linguistically specified content alone (b-examples), it is highly ambiguous—*Woo* on its own could be a comment on practically anything, for example—but when the b-examples are taken with their non-linguistic antecedents (a-examples), their interpretations are constrained.

(1) a. [i offers 1 wheat for 1 sheep from j]
    b. i: if you have one

(2) a. [i offers 1 wheat for 1 sheep from j]
    b. i: or an ore

(3) a. [i rolled a 2 and a 1.] [j gets 2 sheep, 2 wheat. i gets 1 wheat.]
    b. i: Woo!

The *Settlers* corpus was originally designed as a tool for studying strategic conversation, not the non-linguistic context, so annotators were given only the verbal exchanges from the chats and told to annotate them for discourse structure in the style of SDRT. A little over 1000 dialogues have been annotated so far. Observations of the effects of excluding non-linguistic events from the annotations (see

section 3) have prompted a second round of annotations that includes non-linguistic events. This involves importing descriptions of the non-linguistic events (dice rolls, card plays, etc.) from the game log into the annotation files. The full game log temporally orders all linguistic and non-linguistic game events, yielding an automatic alignment of each utterance with the current game state and an explicit numbering of each turn in the game. Because not all turns from the game log were originally assigned numbers, some server turns are given decimal numbers (e.g. 222.4, *vide infra*) to preserve the original numbering.

Many descriptions of non-linguistic events from the game log are public to the players. These descriptions, whose interpretations are determined by the game rules and state, give the *Settlers* corpus a major advantage for the study of non-linguistic events in discourse: they minimize the effects of the individuation and conceptualization problems, and they also allow us to presuppose joint attention of the players, ensuring that all information can be considered to have entered the common ground.[1]

One might worry, however, that because the server produces these descriptions it should really be considered a conversational participant and the events that we are treating as non-linguistic should really count as linguistically-specified. We do not think this is a concern. First of all, the fact that the non-linguistic events in the game are assigned a semantic content does not make them any less non-linguistic. Reasoning about any non-linguistic events, not just those in our corpus, requires that they be conceptualized. This blurs the line between linguistic and non-linguistic events, but we think this is called for by the nature of situated dialogue. Second of all, the players do not need to rely on the server messages to know what is going on in the game; the messages are helpful only as a record for annotators and for players who might have a lapse in attention. Players can tell by looking at the game board where the robber is located and can see if he moves; they can tell when the dice have been passed to a new player because a pointer on the screen will move to the part of the screen dedicated to that player; and so on. Consider an analogy with a sports game in which player A pushes player B and B yells, "Hey, you can't do that!". The fact that there may be a sports announcer who described the pushing does not make the pushing any less non-linguistic and does not mean that the B's reaction was a reaction to a linguistically-specified event, even if B can hear what the announcer is saying.

3 Analysis of the data

3.1 Crossover

To see how the non-linguistic context affects the discourse structure of our chats, we re-examined 5 games from different parts of the *Settlers* corpus: two from the pilot phase, one from season 1 of the competition, and two from season 2 including one Master's League game. We found many examples in which linguistic moves depend in various ways on non-linguistic events; e.g., examples (1)-(3). Conversely, there are numerous examples in which linguistic moves serve as antecedents to non-linguistic actions. This is common after trade negotiations: a successful linguistic negotiation will result in the non-linguistic action of offering a trade through the game interface; an unsuccessful negotiation will generally result either in an alternative type of trade, such as a trade with the bank, or with the player leading the negotiation ending his turn, thus passing the dice to the next player.

(4)

234	18:55:02:745	gotwood4sheep	anyone got wheat for a sheep?
235	18:55:10:047	inca	sorry, not me
236	18:55:18:787	CheshireCatGrin	nope. you seem to have lots of sheep!
237	18:55:23:428	gotwood4sheep	yup baaa
238	18:55:32:308	dmm	i think i'd rather hang on to my wheat i'm afraid
239	18:55:47:845	gotwood4sheep	kk I'll take my chances then...

gotwood4sheep's (GWS) decision to take his chances (239) is the result of the failed negotiation (234-238); turns 234-239 then together result in his ending his turn. On its own, 239 doesn't make much sense—what chances would he be taking? and why is he taking them now? It is the connection between 234-239 and the non-linguistic move of GWS ending his turn, as well as the risks of end-turn moves in

[1] Thank you to an anonymous reviewer for raising this last point and the worry that follows.

the larger game state, that restrict the interpretation of 239: in passing the dice, GWS risks the possibility that another player might roll a 7, causing him to lose precious resources.

Because non-linguistic events were ignored, the first-round of annotations on the *Settlers* corpus contains numerous incorrect attachments and thus incorrect discourse logical forms for the associated chats. In (5), for example, 564 is incorrectly attached to 561 as a Comment.

	561	17:47:24:638	Euan	Ooh! Clay :D
	562	17:47:25:424	Server	jon rolled a 4 and a 4.
(5)	563	17:47:25:426	Server	Cardlinger gets 3 wood. Joel gets 2 ore. Euan gets 1 ore.
	564	17:47:26:064	Cardlinger	that was an easy turn for me :D
	565	17:47:39:875	Euan	I like this "getting resources" business.

The linguistic clues used to guide attachment for 564 were misleading: 561 suggests that someone, probably Euan (E), received clay and that E was happy about it. In 564, Cardlinger (C) expresses a positive attitude about a turn and his comment suggests that he benefited from the turn without doing anything, which is common with resource distributions. E's comment in 565 seems to confirm a resource distribution. Similarities between 561 and 564 and the fact that 562-3 are missing in the original annotation file, led annotators to conclude that C was commenting on the same turn and resource distribution as E.

Treating 564 as a comment on 561 is problematic in part because even if E and C had been commenting on the same event, 561 and 564 should have been either related by a relation like Parallel or understood as two independent comments and so not related at all. Moreover, 564 is, of course, a comment on a completely different event: the mistaken annotation of (5) entails that the token of *that* in 564 refers to an event other than the one that it actually refers to.

The first round of annotations also suffers from missing rhetorical links, which means that the discourse logical forms that result from these annotations do not provide enough information to disambiguate the interpretations of 'orphan' turns, i.e. turns with no incoming rhetorical links. In the vast majority of cases, looking at the surrounding non-linguistic events constrains the interpretation of these turns considerably. (6) illustrates this problem: 344 was an orphan in the first round of annotations, meaning that the annotators did not see a coherent relation between 344 and any previous linguistic turns. The result is that the discourse logical form for (6) leaves the interpretation of 344 far more unconstrained than it intuitively is, with no indication that 344 is even a coherent move in the game.[2]

	341	19:05:26:615	Server	gotwood4sheep rolled a 6 and a 3.
(6)	342	19:05:26:616	Server	inca gets 2 wheat. dmm gets 1 wheat.
	344	19:05:29:595	gotwood4sheep	9 nooo!

Once we consider the non-linguistic context, it is clear that GWS is unhappy because he rolled a 9. It is also clear that 344 is a coherent move in the game because it is rhetorically related to another game move which is itself coherently related to the rest of the game state. Adding non-linguistic turns does not eliminate all ambiguity; in (6), 344 might be a comment only on the roll (341) or on both the roll plus the resource distribution (341 + 342). It is likely a comment on both—the robber is probably occupying a 9 hex on which GWS has a settlement—but there is also a possibility that GWS has another reason for being upset. Still, considering the non-linguistic events reduces the space of possibilities considerably.

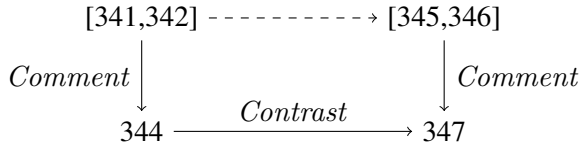
3.2 Attachments

Re-examination of the *Settlers* corpus also revealed anaphoric links that yield surprising discourse structures. For instance, linguistic moves can depend simultaneously on a non-linguistic event and a previous linguistic move. Consider (7), which continues (6):

	345	19:05:34:924	Server	inca rolled a 1 and a 3.
(7)	346	19:05:34:926	Server	gotwood4sheep gets 2 wood.
	347	19:05:39:655	gotwood4sheep	4 better :)

(6) and (7) together yield the graph below, foreign to existing theories of rhetorical structure. The dashed line represents some sort of sequential relation that binds the unit [341,342] to [345,346]; we leave aside its nature here to focus on linguistic/non-linguistic links.

[2]Throughout, we skip turns, e.g. 343, to save space when those turns are irrelevant to our main point.

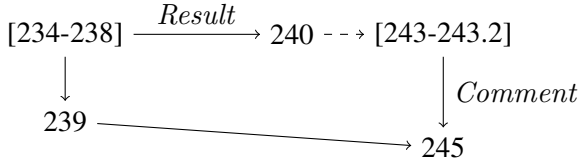


347 is a highly context sensitive fragment that becomes coherent—we understand *what* is better—only once we see its connection to [345,346]. At the same time, *better* signals a contrast with something comparatively worse. In the first round of annotations, annotators used this signal to link 344 (*9 nooo!*) and 347 with Contrast. This is intuitively correct: 344 expresses a negative attitude about the thing that is "comparatively worse", namely events described in [341,342], so the attitudes are in contrast.[3]

Example (8) likewise exhibits simultaneous linguistic and non-linguistic dependence, but it also shows that attachments can span across several game-state changing events, even when those events are related to the linguistic context via coordinating relations like Result, which SDRT predicts should render previously accessible nodes (here [234-239]) inaccessible for future anaphora.

	240	18:55:53:749	Server	inca rolled a 6 and a 2.
	243	18:56:06:835	Server	dmm rolled a 1 and a 6.
(8)	243.1	18:56:06:835	Server	gotwood4sheep needs to discard.
	243.2	18:56:10:630	Server	gotwood4sheep discarded 4 resources.
	245	18:56:13:443	gotwood4sheep	chance fail

(4)+(8) involves a causal chain that passes through 234–239 to a sequence of non-linguistic events, including multiple rolls/turn changes, which update the game state. The last roll (243) results in GWS's losing the gamble (239) that resulted from the failed negotiation (234-238) in (4). 245 was attached to 239 in the first round of annotations and this connection is important for the interpretation of 245. But here again, this link yields a surprising discourse structure because 245 links to 239 despite the presence of Result([234,238],240), which should block accessibility of all events before 240.



The kind of structures that we see once we incorporate non-linguistic events are unlike the dependency trees familiar from *Rhetorical Structure Theory* (RST; Mann and Thompson (1987)) or the directed graphs from SDRT. Given that the shapes of dependency trees and directed graphs are meant to constrain available attachment points, this raises the question: how does the introduction of non-linguistic events affect the set of available attachment points for a given turn? To answer this question, we must investigate the extent to which a set of non-linguistic events has an internal structure and how this structure effects that of the linguistically specified contents in a situated conversation. The examples in our corpus show that treating the non-linguistic events in *Settlers* as forming an unstructured set or a mere sequence of game-changing events, in which either all events are accessible or only the last one is, is insufficient. (6)+(7) and (4)+(8) show that the last turn is not the only available one, and (9) suggests that this is true regardless of whether there is a previous linguistic antecedent facilitating anaphoric dependence on past non-linguistic events.

	154.1	20:21:09:163	Server	gotwood4sheep played a Soldier card.
	154.3	20:21:10:230	Server	gotwood4sheep stole a resource from ljaybrad123
	155	20:21:12:395	Server	gotwood4sheep rolled a 5 and a 1.
	157	20:21:15:027	Server	gotwood4sheep built a settlement.
(9)	158	20:21:19:939	gotwood4sheep	sorry laura
	159	20:21:23:907	gotwood4sheep	needed clay the mean way :D
	159.1	20:21:24:241	Server	ljaybrad123 played a Soldier card.
	159.4	20:21:35:323	Server	ljaybrad123 stole a resource from gotwood4sheep
	163	20:21:40:457	gotwood4sheep	touché

[3]The intuitive connection between 347 and [341,342]—i.e. the comment *4 better :)* expresses the attitude that the roll of the 4 is better than the roll described in 341—should follow at the level of interpretation from the fact that 347 contrasts with 344, which describes an attitude about 341, so there is no reason to draw an extra arc in the discourse graph from [341,342] to 347.

GWS's utterance in 158 ignores his roll in 155 and building in 157, referring back to his steal in 154.3. To make sense of 163, one needs to consider the relation between the steal in 154.3 and that in 159.4, but it is not the linguistic moves in 158 and 159 that give us this structure; even if we ignore these two moves, 163 is coherent given the non-linguistic context.

On the other hand, a fragment like *sorry laura* is highly anaphoric and needs a salient antecedent; therefore, while 158 need not depend on the last move (157), it can't pick up on just any move in the game, either. There are limitations on the accessibility of non-linguistic events. In other words, the non-linguistic events in our game have an internal structure just as the linguistic events do; and what's more, the two structures are integrated. §4 looks at the nature of this integrated structure in more detail.

4 Integrated discourse structures

As in Asher and Lascarides (2003), our annotations assign speech act labels to *elementary discourse units*, which may coincide with or be a proper part of a speaker's turn. We use distinguished variables $\pi^i, \pi^i_1, \pi^i_2, \ldots$ to label EDUs, where π^i labels a speech act performed by i. To build integrated structures, we also treat non-linguistic events in the game as entities or *elementary event units* (EEUs). Each EEU is assigned a first-order formula ϕ that characterizes its content; i.e. ϵ: ϕ is a discourse formula that characterizes the EEU ϵ with ϕ. The interpretation of ϕ in the relevant model, which in this case is determined by the nature of the *Settlers* game, will determine the conceptualization of ϵ. *Complex discourse units* (CDUs), made up of multiple EDUs, EEUs, or a combination of the two, are labelled like EDUs; however, their subscripts reflect a group of speakers when multiple speakers have contributed to their content.

Building appropriate discourse structures for situated discourse requires us to address three problems that already come up for discourse analysis: (a) the *segmentation problem*, i.e. that of individuating the EDUs and EEUs; (b) *the attachment problem*, i.e. how these units link up via their anaphoric dependencies; (c) *the labelling problem*, i.e. which kinds of relations hold between the units. It also requires addressing (d) *the conceptualization problem*, which does not in general arise for the analysis of text. In our *Settlers* corpus study, (a) is moot; the game server segments the relevant events for us, though in general this is a problem (Lascarides and Stone, 2009b). The server messages also help with (d) by assigning a formula ϕ to each ϵ, though these formulas need an interpretation, which we discuss in §5. The discussion in §3.2 shows that (b) needs a solution. We provide this in §4.1-4.2.

The labelling problem, (c), also needs to be addressed because our data require generalizations of certain rhetorical relations that take EEUs and CDUs containing EEUs as arguments. This includes not only Comment (ex. (3)), Alternation (ex. (2)), and Conditional (ex. (1)), but also Result, Explanation, Elaboration and Question-Answer Pair (QAP). For example, we understand gotwood4sheep's action of ending his turn and passing the dice to be the *result* of his failed negotiation attempt. Had he succeeded in trading, he likely would have built something. And in (10), gotwood4sheep's non-linguistic actions provide an *explanation* of his speech act in 538.

(10)

538	21:03:12:661	gotwood4sheep	it may prove a prudent trade, lj...
539	21:03:24:209	ljaybrad123	nope
539.1	21:03:25:530	Server	gotwood4sheep played a Soldier card.
539.4	21:03:28:353	Server	gotwood4sheep stole a resource from ljaybrad123

In the turns leading up to 538, GWS attempts to trade with lj to get the resources that he wants. When she rejects his offer, he plays a Soldier card, which allows him to steal a resource from her. So lj's outcome is worse than had she traded.

57 and 59 in (11) are *restatements* of the non-linguistic offer (56) and accept (58) moves.

(11)

56	16:38:11:641	Server	Joel made an offer to trade 1 wheat for 1 clay.
57	16:38:22:445	Joel	I just sent the trade request
58	16:38:30:436	Server	Joel traded 1 wheat for 1 clay from Euan.
59	16:38:47:583	Euan	I accepted.

Finally, offers to trade function like polar questions, partitioning the subsequent state into two alternatives: the 'addressee' can either accept or reject the offer. Accordingly, we link offers and their responses via QAP. §5 describes the relevant semantics for offer, accept and reject moves.

4.1 Adding EEUs to discourse graphs

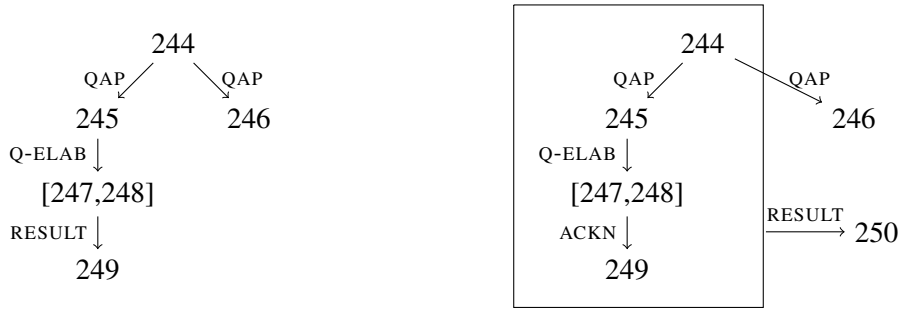
If EEUs enter into rhetorical relations with EDUs—even figuring in mixed CDUs—we cannot simply append a Kaplanian-like context of non-linguistic entities to linguistic discourse structures. Kaplanian contexts are unstructured sets of entities, making them unfit to model the structural relations that we have observed and the accessibility facts discussed in §3.2. Instead, we need to extend our discourse graphs to *situated* discourse graphs with nodes for EEUs and arcs that connect them to other nodes. That is, if a discourse structure for a dialogue d derived only from linguistic utterances is a connected graph $G_d = (V, E_1, E_2)$ where V is the set of EDUs and CDUs, $E_1 \subseteq V \times V$ the set of labelled discourse attachments between elements of V, and $E_2 \subseteq V \times V$ the parenthood relation holding between DUs and their CDU hosts (note that $E_1 \cap E_2 = \emptyset$), then a situated discourse graph G_{sd} for d is the graph: $G_{sd} = (V^{sd}, E_1^{sd}, E_2^{sd})$, where V^{sd} shares with V the same EDUs but also contains EEUs and CDUs that include EEUs, while $E_1^{sd}, E_2^{sd} \subseteq V^{sd} \times V^{sd}$ are the analogues in G_{sd} of E_1 and E_2 in G_d.

Given G_d and G_{sd}, several questions arise. First, is G_d a subgraph of G_{sd} in the following sense: are $V \subseteq V^{sd}$, $E_1 \subseteq E_1^{sd}$, and $E_2 \subseteq E_2^{sd}$? The answer is 'no' for many dialogues: new and different CDUs are created in the presence of non-linguistic events. This happens frequently in our corpus when, for example, a series of linguistic moves *results* in an EEU. For instance, a verbal acceptance of an offer typically leads to a non-linguistic offer to trade, as happens in (12); the non-linguistic offer and its acceptance is how players effect an actual change to the game state so that it matches the agreed trade.

(12)

244	10:55:44:639	mmatrtajova	anyone will trad wheat or sheep?
245	10:55:52:100	Ash	yes for wood
246	10:55:52:379	J	nopes
247	10:56:20:215	mmatrtajova	okay wood for wheat?
248	10:56:32:205	mmatrtajova	and sheep for ore?
249	10:56:41:896	Ash	ok
250	10:56:47:071	Server	mmatrtajova made an offer to trade 1 ore, 1 wood for 1 sheep, 1 wheat.

The semantics of causal discourse relations like Result call for a grouping of all of the linguistic and nonlinguistic events that result in the nonlinguistic offer. Thus, adding the non-linguistic turn 250 from (12) triggers the introduction of the CDU represented by the box in the graph on the right:



Unsuccessful negotiations, like (4), have a similar effect on the discourse graph, as the negotiation as a whole results in an action such as trading from the bank or ending the turn. Our corpus also provides examples of hybrid (EDU+EEU) CDUs and of CDUs from G_d that disappear once EEUs are added. So in general $V \not\subseteq V^{sd}$, and hence $E_2 \not\subseteq E_2^{sd}$. Our data also show that $E_1 \not\subseteq E_1^{sd}$ (see §4.3).

The next question is: do the new relation instances in the graph G_{sd} obey the same structural constraints as the relation instances in G_d? For example, Venant et al. (2013) argues that CDUs have a *no punctures* property such that if π is a CDU then there are no incoming directed edges to proper elements of π whose source is outside π. In the *Settlers* corpus, all CDUs in the linguistic only graphs G_d have this property. The creation of new CDUs, however, makes it more difficult to verify for the graphs G_{sd}; while we have yet to find any punctures, we need a fuller study. Another constraint that holds for discourse-only graphs G_d is the *Right Frontier Constraint* (RFC) on attachment (Asher, 1993; Asher and Lascarides, 2003), which other theories like RST also adopt. We turn now to a discussion of the RFC.

4.2 The Right Frontier

In SDRT, for a discourse graph $G = (V, E_1, E_2)$, E_1 contains two types of edges, *coordinating* and *subordinating*, and we write $e(x, y)$ for e is an edge with initial point x and endpoint y. The *Right Frontier* (RF) for an EDU y, i.e. the set of available attachment points for y, includes: *Last*, the EDU x just before y in a linear ordering of DUs from the discourse; any node that is super-ordinate to x via a series of subordinating relations; and any CDU in which x figures. More precisely,

Definition 1. Let $G = (V, E_1, E_2)$ be a discourse graph. $\forall x, y, z \in V$, $\text{RF}_G(x)$ iff (i) $x = Last$, (ii) $\text{RF}_G(y)$ and $\exists e \in E_1$, $e(x, y)$ and Subordinating(e), (iii) $\text{RF}_G(z)$, $\text{RF}_G(y)$ and $\exists e, e' \in E_1$ such that $e(x, z)$ and $e'(z, y)$ and Subordinating(e) and Subordinating(e'), or (iv) $\text{RF}_G(y)$ and $\exists e \in E_2$, $e(y, x)$

The RFC is much more complicated in multi-party dialogue than in monologue, even within linguistic-only graphs G_d. For multi-party dialogue, we examined a more complex notion of the RF by considering an RF for the subgraph of the connected contributions of each subgroup of conversational participants in G. This choice reflects the fact that there are often several interwoven conversations between subgroups in a dialogue, and even within a subgroup an agent's contribution can attach to several participants' moves, as observed in Ginzburg and Fernández (2005). On this more complex conception, which we call the *supervaluational* RF, we took an attachment of one discourse unit to another to be an RF violation if and only if the attachment violated the RF of the complete graph and the RF of the subgraph for each subgroup of participants. Our extension of the RFC to subgroups shows that crossing conversations are quite common in our corpus. Of the 8829 relation instances in the corpus, 78% of attachments respected a slightly simplified version of the RF defined above. Manually assessing one representative game, we found that 98% of the attachments respected the supervaluational RF.

Understanding how EEUs affect the RF requires understanding their structure and that of CDUs composed from them. The game structure at a high level is a sequence of large events individuated by moves like dice rolls, bargain initiation and building. Each of these high-level events might have as parts other events: for example, the resource distribution or sequence of robber moves that results from a roll, or the various turns in a trade negotiation, and so on. Prior to any linguistic move any EEU ϵ_1 may be commented on *ad libidem* but a linguistic event, like a new offer or a comment on an EEU, will move the RF on, and we predict that an EEU ϵ_1 preceding the linguistic event is no longer accessible for attachment. Any comment on, say, a roll that precedes a bargaining discussion has to involve a definite description allowing for an RF violation and inducing a so-called *discourse subordination* (Asher, 1993). EEUs that follow the bargaining discussion, e.g. an EEU offer, a resulting EEU acceptance, a new roll and shift in bargaining leader and so on, will *all* remain open so long as no commentary is made on these EEUs. In addition, all of the events on the RF of the preceding discussion will also remain accessible. Once linguistic content is attached to one of these later EEUs ϵ_2, however, the RF of the preceding discussion disappears and all EEUs prior to ϵ_2 become inaccessible. We formalize these observations below in an RF constraint for situated dialogue (ignoring subgroups of participants to simplify), in which the RF is parametrized relative to high-level events ϵ (rolls, bargain initiations, buildings, card plays, etc.); $Acc(x)$ means x is on the the RF as defined in Definition 1; $\prec$ is a linear ordering of EEUs and EDUS in V; and π is an EDU or CDU containing at least one EDU.

Definition 2. Let $G_{sd} = (V, E_1, E_2)$ be a situated discourse graph. $\text{RF}(G_{sd}, \epsilon, Last) = \{Last\} \cup Acc(Last) \cup \{\epsilon' : \epsilon' \prec \epsilon \text{ and } \neg\exists\epsilon''\exists\pi \in V \exists e \in E_1\ (\epsilon' \prec \epsilon'' \wedge e(\epsilon'', \pi))\}$

This definition allows for the "rectangular structures" exhibited by (6)+(7) or (4)+(8) but absent from the linguistic-only graphs G_d produced from the first round of annotations on the *Settlers* corpus.

4.3 Divergences

Differences between situated discourse graphs G_{sd} and linguistic-only discourse graphs G_d provide an indirect measure of how much the non-linguistic context affects the comprehension of linguistic discourse moves, at least for the *Settlers* corpus. It is indirect because we are only able to measure the

effects of the non-linguistic context on judgments about attachments and labelling of arcs in the situated discourse graph. Nevertheless, a comparison will be instructive in showing how the non-linguistic context may affect the interpretation of linguistic content.

We categorized the divergences into 5 categories: (i) EEUs missing in G_d that were essential in our judgement for understanding linguistic moves (ME in Table 1); (ii) links missing between EEUs and EDUs (ML); (iii) incorrect links in G_d, which had to be changed in light of the non-linguistic context (IL) (cf. (5)); (iv), missing CDUs in G_d, which the semantics of discourse relations and the presence of EEUs forced us to create (MC) (cf (12)); (v) incorrect dialogue boundaries postulated from the linguistic-only annotations, which changed in light of the non-linguistic context (wrong breaks). Annotators of the linguistic-only part of the dialogue would postulate boundaries when there were two unconnected discourse graphs. These discourse graph boundaries often corresponded to dice rolls but not always. We found two sorts of errors: the first where two distinct discourse graphs G_d and $G_{d'}$ were postulated when in fact there was one connected situated graph, the second where one graph G_d spanned in fact two separate situated graphs G_{sd} and $G_{sd'}$. We counted errors in 3 games: Pilot14, a game with novice players, s1-league1-game3, from season one of the *Settlers* competition, and s2-leagueM-game2, a master's league game. TL and TDU in Table 1 are respectively the total number of links and DUs per game.

game	missing EEU (ME)	missing links (ML)	missing CDU (MC)	incorrect links (IL)	wrong breaks
s1-league1-game3	122	162	26	44	6
s2-leagueM-game2	78	119	15	25	3
pilot14	72	115	6	25	2

game	total # errors	DU error rate (MU/TDU)	link error rate in G_d (IL/TL)	TL in G_d	TDU in G_d
s1–league1-game3	360	17%	6%	722	687
s2-leagueM-game2	340	21%	7%	369	345
pilot14	220	25%	13%	190	214

Table 1: Error rates on Settlers games

The set of dialogues for a given game provided a variable error rate on the existing annotated links in G_d of between 6% and 13%. On the other hand, our analysis showed that the G_d graphs were often seriously incomplete with respect to events deemed essential to understanding the content of the dialogue. We thus concluded: 1) Non-linguistic events are often crucial for understanding the content of dialogue. While this might not be surprising, it motivates an approach to the study of the semantic content of discourse that embraces its potential for radical context sensitivity and attempts to incorporate this sensitivity into a formal semantic/pragmatic model (cf. Ginzburg (2012)). 2) Despite the radical context sensitivity exhibited by our corpus, linguistic clues for inferential relations were remarkably robust: the linguistic-only annotations were quite incomplete but not hopelessly wrong, given the low error rate of dialogue attachments and labelling in the G_d graphs. Whether these observations generalize to different types of a conversation is an open question, but we hope that the techniques elaborated here can facilitate a comparison of different types of conversation.

5 EEU semantics

Rhetorical relations interact with, and are licensed by inferences about, the *content* of EDUs (Asher and Lascarides, 2003). While EEUs are non-linguistic, their *conceptualization* by conversational participants endows them with a content, $\epsilon : \phi$, that enables them to serve as arguments to rhetorical relations. In the *Settlers* corpus, each such formula ϕ is recorded in a server message; the interpretation of ϕ, the conceptualization of ϵ, is determined by the nature of the *Settlers* game. We illustrate our semantics for EEU contents ϕ by looking at trade moves. For instance, the move *offer*, whether linguistic or non-linguistic, has the semantics of a question at an abstract level. *Accept* and *reject* moves, whether linguistic or not, function as answers. (13) is an example.

(13) $\epsilon : \mathit{offer}(i, 1wheat, 1sheep, j) \rightarrow_{\text{qap}} \pi^j$: I DON'T HAVE ANY SHEEP.

More formally, let g be an initial segment of a game; $g.V^*$, the game tree of all legal sequences given g; and E, the Linear Temporal Logic operator *eventually*. We write '$g < g'$' where the sequence g' extends g, and '$g \models \phi$' where the end state of g satisfies the formula ϕ. We note the fact that some actions depend on prior actions—e.g., '$\epsilon' : acc(a, \epsilon)$' depends on ϵ—with the notation, $\epsilon' \mapsto \epsilon$.

$g \| \epsilon$: $\mathit{offer}(a, c, d, b) \| g'$ iff $g.\|\epsilon\| = g'$ and $\forall g'' > g', g'' \in g'.V^* \rightarrow (g' \| E(acc(b, \epsilon)) \| g'' \vee g' \| E(rej(b, \epsilon)) \| g'')$

$g \| acc(b, \epsilon) \| g'$ iff ϵ: $\mathit{offer}(a, c, d, b) \in g$ and $\exists e \in g'(e \mapsto \epsilon$ and $g.e \models b\ gets\ c\ and\ a\ gets\ d)$

$g \| rej(b, \epsilon) \| g'$ iff ϵ: $\mathit{offer}(a, c, d, b) \in g$ and $\neg\exists e \in g'(e \mapsto \epsilon$ and $g.e \models b\ gets\ c\ and\ a\ gets\ d)$

where $\mathit{offer}(a, c, d, b)$ stands for 'a offers to give c to player b in exchange for d'. Thus (13) holds just in case all continuations of the game include the event of the offer specified in ϵ followed by an event of refusing that offer. We get this by noting that π^j entails that j can't give sheep to i because she has none. The relation QAP ensures the dependency of the linguistic refusal on the non-linguistic offer ($e \mapsto \epsilon$).

6 Related Work

The current work complements prior research on rhetorical dependencies between linguistically specified arguments and co-verbal gestures (Lascarides and Stone, 2009a,b) to model how non-linguistic events affect the overall architecture of discourse. Our project also goes beyond existing work on the non-linguistic context. Aside from semantic and philosophical work on indexicality, projects like TACoS (Regneri et al., 2013) and The Restaurant Game (Orkin et al., 2010) use non-linguistic events to refine event descriptions or to automatically learn scriptal information, but do not engage in the kind of theoretical investigation of the discourse interactions and their semantics that we have undertaken. Finally, corpus based studies of dialogue, e.g. Ginzburg (2012), posit that incomplete utterances (e.g. *Woo!* and *Yay!*) have as a part of their semantics a kind of anaphoric dependency like that studied in rhetorical theories—a requirement for interaction with the discourse situation. Yet this work does not look at how we interact with non-linguistic events in such situations. Our work complements research in these separate fields to give a more complete picture of how the semantic content of discourses depends on interactions with the situations in which our discourses take place.

There has been much recent progress in interpreting multimodal actions within the field of human robot interaction (Perzanowski et al., 2001; Chambers et al., 2005; Foster and Petrick, 2014). The task is to map the outputs of 'low-level' signal processors into a representation of speaker meaning. This work uses planning or reinforcement learning and holds that reasoning about the cognitive states of the participants is a primary and irreducible source of information for parsing multimodal actions. We take a slightly different view: instead of always exploiting reasoning about cognitive states directly, we infer speaker meaning via constraints afforded by models of discourse coherence. Discourse coherence has proved useful for predicting anaphoric dependencies and implicatures in purely linguistic discourse. Our hypothesis is that it will play much the same role in resolving the meanings of multimodal actions.

7 Conclusion

Interaction between non-linguistic and linguistic events extends well beyond the context of a shared task like that in *Settlers*. Suppose a driver does something dangerous in passing you. Whether you respond by yelling at him or by giving him the finger, your response anaphorically depends on and rhetorically connects to the driver's action. The *Settlers* data give us an interesting insight into the seamless web of linguistic and non-linguistic events. Because non-linguistic events are already segmented and described in our corpus, we have been able to ignore the segmentation problem—a problem that in general renders the study of the non-linguistic context dauntingly complex—and examine how non-linguistic events effect discourse structure. We have developed a model of situated conversation to capture these effects, giving us a better understanding of how the non-linguistic context affects semantic content.

References

Afantenos, S., N. Asher, F. Benamara, A. Cadilhac, C. Dégremont, P. Denis, M. Guhe, S. Keizer, A. Lascarides, O. Lemon, P. Muller, S. Paul, V. Popescu, V. Rieser, and L. Vieu (2012). Modelling strategic conversation: model, annotation design and corpus. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (Seinedial)*, Paris.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.

Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Chambers, N., J. Allen, L. Galescu, and H. Jung (2005). A dialogue-based approach to multi-robot team control. In *Proceedings of the 3rd International Multi-Robot Systems Workshop*, Washington, DC.

Foster, M. E. and R. P. A. Petrick (2014, June). Planning for social interaction with sensor uncertainty. In *Proceedings of the ICAPS 2014 Scheduling and Planning Applications Workshop (SPARK)*, Portsmouth, New Hampshire, USA, pp. 19–20.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Ginzburg, J. and R. Fernández (2005). Scaling up from dialogue to multilogue: some principles and benchmarks. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 231–238. Association for Computational Linguistics.

Hobbs, J. R., M. Stickel, D. Appelt, and P. Martin (1993). Interpretation as abduction. *Artificial Intelligence 63*(1–2), 69–142.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes from Kaplan*. Oxford.

Lascarides, A. and M. Stone (2009a). Discourse coherence and gesture interpretation. *Gesture 9*(2), 147–180.

Lascarides, A. and M. Stone (2009b). A formal semantic analysis of gesture. *Journal of Semantics 26*(4), 393–449.

Mann, W. C. and S. A. Thompson (1987). Rhetorical structure theory: A framework for the analysis of texts. *International Pragmatics Association Papers in Pragmatics 1*, 79–105.

Orkin, J., T. Smith, and D. K. Roy (2010). Behavior compilation for ai in games. In *Proceedings of the 6th Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE)*.

Perzanowski, D., A. Schultz, W. Adams, E. Marsh, and M. Bugajska (2001). Building a multimodal human-robot interface. *Intelligent Systems 16*(1), 16–21.

Regneri, M., M. Rohrbach, D. Wetzel, S. Thater, B. Schiele, and M. Pinkal (2013). Grounding action descriptions in videos. *TACL 1*(2), 25–35.

Venant, A., N. Asher, P. Muller, and P. D. S. D. Afantenos (2013). Expressivity and comparison of models of discourse structure. In *Proceedings of Sigdial 2013*, Metz, France.