

Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus

Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, Stergos Afantenos

▶ **To cite this version:**

Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, Stergos Afantenos. Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus. 10th International Conference on Language Resources and Evaluation (LREC 2016), May 2016, Portoroz, Slovenia. pp.2721-2727. hal-02124399

HAL Id: hal-02124399
https://hal.science/hal-02124399

Submitted on 9 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
http://oatao.univ-toulouse.fr/22646

Official URL

http://www.lrec-conf.org/proceedings/lrec2016/summaries/339.html

To cite this version: Asher, Nicholas and Hunter, Julie and Morey, Mathieu and Benamara, Farah and Afantenos, Stergos *Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus.* (2016) In: 10th International Conference on Language Resources and Evaluation (LREC 2016), 23 May 2016 - 28 May 2016 (Portoroz, Slovenia).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus

[†]Nicholas Asher, [‡]Julie Hunter, [‡]Mathieu Morey, [‡] Farah Benamara, [‡]Stergos Afantenos
[†]IRIT, CNRS; [‡]IRIT, Université Paul Sabatier
118 route de Narbonne, 30162 Toulouse, France
{nicholas.asher; julie.hunter; mathieu.morey; farah.benamara; stergos.afantenos}@irit.fr

Abstract

This paper describes the STAC resource, a corpus of multi-party chats annotated for discourse structure in the style of SDRT (Asher and Lascarides, 2003; Lascarides and Asher, 2009). The main goal of the STAC project is to study the discourse structure of multi-party dialogues in order to understand the linguistic strategies adopted by interlocutors to achieve their conversational goals, especially when these goals are opposed. The STAC corpus is not only a rich source of data on strategic conversation, but also the first corpus that we are aware of that provides full discourse structures for multi-party dialogues. It has other remarkable features that make it an interesting resource for other topics: interleaved threads, creative language, and interactions between linguistic and extra-linguistic contexts.

Keywords: discourse structure, strategic conversation, multiparty dialogue

This paper describes the STAC resource, a corpus of multi-party chats annotated for discourse structure in the style of SDRT (Asher and Lascarides, 2003; Lascarides and Asher, 2009). The main goal of the STAC project[1] is to study the discourse structure of multi-party dialogues in order to understand the linguistic strategies adopted by interlocutors to achieve their conversational goals, especially when these goals are opposed. The STAC corpus is not only a rich source of data on strategic conversation, but also the first corpus that we are aware of that provides full discourse structures for multi-party dialogues. It has other remarkable features that make it an interesting resource for other topics: interleaved threads, creative language, and interactions between linguistic and extra-linguistic contexts.

1. The corpus

Our corpus comes from an online version of the game *The Settlers of Catan*—a win-lose, multi-player game in which players acquire and trade resources (ore, wood, wheat, clay, or sheep) in order to build roads, settlements, and cities and in turn score victory points. Resources are sometimes allocated automatically after a turn, but generally a player will have access to only a strict subset of the resources she needs, prompting her to trade with other players to achieve her goals. We collected our corpus by modifying an online, open-source version of *Catan* with a chat interface in which players could carry out trade negotiations.[2] Figure 1 is a snapshot of the board game, showing the chat window ("Chat"), chat history ("History"), and game history ("Game"), where the game history details all of the extra-linguistic events (e.g., dice rolls, card plays) from the game. This snapshot shows the perspective of the game administrator; normally, the type of resources that a player has is revealed only to that player. The moves from the "Game" window and those from "History" are automatically recorded and aligned in a game log, which allows us to replay an entire game.[3] The .soclog files, segmented files, and annotated files from our corpus, which we describe below, are available online.[4]

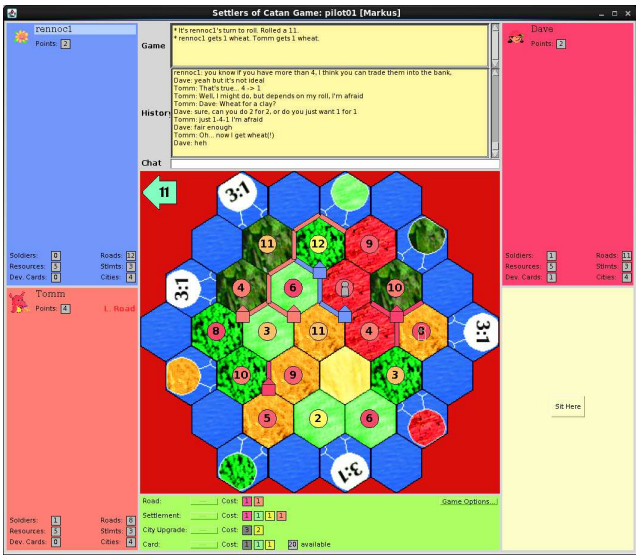


Figure 1: Snapshot of a *Catan* game board

Our data differed considerably from written single-authored text in other corpora that we have examined in earlier annotation campaigns (Péry-Woodley et al., 2009). In addition to negotiation dialogues, it features creative nouns (*dolly* for *sheep*), novel verbs (*as I alt tab back from the tutorial*), and V ellipsis without a surface antecedent (*I can wheat for clay*). The text is messy, requiring robust parsing to deal with ubiquitous misspellings, contractions, and missing punctuation.

Another feature of our corpus is the presence of subdialogues or threads that divide, merge, or get dropped as the dialogue proceeds. Example 1 contains at least 3 threads,

The authors gratefully acknowledge research support from ERC Advanced Researcher Grant n.269427.

[1] *Strategic Conversation*, ERC grant n.269427.

[2] http://homepages.inf.ed.ac.uk/mguhe/socl/

[3] This program is available upon request.

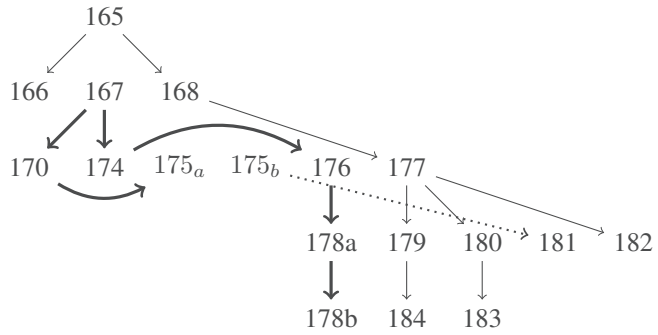
[4] https://aclweb.org/anthology/attachments/D/D15/D15-1109.Attachment.zip.

which we have represented with different fonts.

Example 1

165	lj	anyone want sheep for clay?
166	gwfs	got none, sorry :(
167	gwfs	**so how do people know about the league?**
168	wm	no
p 170	lj	**i did the trials**
174	tk	**i know about it from my gf**
175	gwfs	[**yeah me too,**]$_a$ [*are you an Informatics student then, lj?*]$_b$
176	tk	**did not do the trials**
177	wm	has anyone got wood for me?
178	gwfs	[**I did them**]$_a$ [**because a friend did**]$_b$
179	gwfs	lol wm, you cad
180	gwfs	afraid not :(
181	lj	[*no, I'm about to start math.*]$_a$ [*I just hang around appleton a lot*]$_b$
182	tk	sry no
183	gwfs	my single wood is precious
184	wm	what's a cad?

Note that turns can be complex, involving several discourse units (usually clauses) that serve distinct discourse functions. Moreover, the threads can cross over one another yielding discourse connections that challenge assumptions about the projectivity of discourse structure. Below is a picture of the dependencies we annotated in Example 1.



Our corpus also contains extra-linguistic turns; that is, moves that are made in the game but are not described linguistically by the players. Extra-linguistic moves in a real-life game of *Catan* would include dice rolls, card plays, resource distributions, building events, and so on. These are things that players *do* or *receive*, not things that they *say* or describe through language. In our virtual *Catan* game, we count these moves as extra-linguistic as well. Frequently, we see interesting linguistic dependencies on these events. In Example 2, GWFS plays a card (154.1) that allows him to steal from LJ (154.3); he then apologizes and explains himself (158-9). LJ in turn plays a soldier card and steals from GWFS (159.1-4), to which GWFS replies *touché*. The linguistic moves (158,159,163) only make sense if we interpret them in light of the extra-linguistic context.

Example 2

154.1	Server	gwfs played a Soldier card.
154.3	Server	gwfs stole a resource from lj
158	gwfs	sorry laura
159	gwfs	needed clay the mean way :D
159.1	Server	lj played a Soldier card.
159.4	Server	lj stole a resource from gwfs
163	gwfs	touché

Sometimes these extra-linguistic events provide antecedents for pronouns or serve as antecedents for elliptical or otherwise fragmentary utterances. In turn 280 from Example 3a, WM's use of the third-person pronoun *it* refers back to the resource that he stole from GWFS in turn 278.2. And in turn 57 of Example 3b, GWFS's utterance of *fast mover!* is clearly a response to WM's building success. (The extra-linguistic moves shown in 3b are some of the first such moves in the game; most of the previous turns were chat turns.) Thus *fast mover!* conveys the full content, *you're a fast mover!*, with *you* referring to WM.

Example 3a

278	Server	wm rolled a 6 and a 1.
278.2	Server	wm stole a resource from gwfs
279	gwfs	oucho
280	wm	you can have **it** back for some ore

Example 3b

51	Server	wm made an offer to trade 1 clay for 1 sheep.
52	Server	wm traded 1 clay for 1 sheep from gwfs.
53	gwfs	thx
54	Server	wm built a road.
55	Server	wm built a settlement.
57	gwfs	fast mover!

Extra-linguistic events can therefore support *subsentential* elliptical or anaphoric constructions, provide antecedents for anaphoric pronouns, and serve as terms of discourse relations, which means that they have something like a propositional type.

Our chat corpus differs from other extant dialogue corpora (Switchboard) in that its participants are involved in a shared, virtual activity that makes for a dynamic environment. Unlike the Mission Rehearsal Excercise (MRE) dialogues discussed in (Traum et al., 2008) or the "appointment discussions" of (Wahlster, 2000), our dialogues are spontaneous exchanges between human participants. There is no face to face communication, but participants exploit the virtual environment in communication as Examples 2 and 3 show, even if they can't demonstrate the visually presented elements in that environment.

Because the environment is dynamic (players' actions are changing the game board), the dialogue is largely reactive with few long distance attachments (greater than five turns away), although attachments of discourse moves to moves two to five turns away are relatively common as (Ginzburg and Fernández, 2005) noted for multi-party dialogue ellipses. Bargaining, for example, is a core of many of our dialogues with a pattern of offers and counteroffers, often leading to an explicit acceptance or rejection, which might be subsequently attached to the initial offer. Nevertheless,

discussions, especially those involving multiple threads between subgroups of participants do occasionally create intuitive, long distance and crossing dependencies between dialogue moves. Multiple threads are of course particular to dialogue, and something we have observed in group discussions like those in the classroom or a seminar.

Finally, like the MRE dialogues, the dialogues in our corpus have a strategic flavor. The dialogue participants in our chats are competing to win the game. This means that they have interests that are opposed; they sometimes don't reveal information that another agent asks for and even arguably lie or at least misdirect in the sense of (Asher and Lascarides, 2013; Asher et al., 2015). Consider Example 4 with the relevant hidden resources for each player in parentheses:

Example 4

846	tk	(ore=0)	anyone got some ore?
847	gwfs	(ore=2)	nope sorry

Our chat dialogues are thus not completely cooperative as, for example, are those discussed in (Grosz, 1979), and so do not follow a common, intended plan structure (Grosz and Sidner, 1986; Grosz and Sidner, 1990).

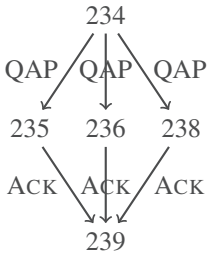
2. Theoretical background for the annotation model

There are several theories of discourse structure for texts: RST (Mann and Thompson, 1987), LDM (Polanyi et al., 2004), the graphbank model (Wolf and Gibson, 2005), DLTAG (Forbes et al., 2003), PDTB (Prasad et al., 2008), and SDRT (Asher and Lascarides, 2003). However, data from our corpus rule out DLTAG, LDM, and RST as candidate theories because they posit tree-based discourse structures. In particular, our corpus contains frequent acknowledgments or other moves directed at more than one player as shown in this excerpt from our corpus:

Example 5

234	gwfs	anyone got wheat for a sheep?
235	inca	sorry, not me
236	Ccg	$[\text{nope.}]_a$ [you seem to have lots of sheep!]$_b$
237	gwfs	yup baaa
238	dmm	i think i'd rather hang on to my wheat i'm afraid
239	gwfs	kk I'll take my chances then...

GWFS initiates an exchange in turn 234, and receives three negative replies (235, 236, 238). He responds in 239 with the acknowledgement "kk" (="okay cool"), which is intuitively addressed at all three negative replies. This exchange yields a "lozenge"-shaped graph, shown below (ignoring 236 and 237). The label QAP is short for "Question-Answer Pair" and types the connections between 234 and each of the three replies; Ack is short for "Acknowledgement" and types the connection between each reply and the acknowledgement in 239.



Exchanges like Example 5 motivate an annotation model that uses graphs rather than trees. We have chosen SDRT.

To define a discourse structure in SDRT we perform three tasks, common also to RST-style annotations. First, we segment the text into *elementary discourse units* (EDUs), which serve as arguments for discourse relations. Second, we determine which EDUs attach to which others. This sometimes requires constructing *complex discourse units* (CDUs) in which multiple EDUs and/or CDUs are grouped together to form a single argument to a discourse relation. We thus explicitly construct hierarchical structures both via discourse attachments and by constructing CDUs from smaller DUs in recursive fashion. Third, we label the edges of our graph. Formally, this means that an SDRT discourse structure is a graph, (V, E_1, E_2, ℓ), with a set V of discourse units; two types of edges, E_1 (relations) and E_2 (CDU membership), with $E_1, E_2 \subseteq V^2$; and ℓ a labelling function $\ell\colon E_1 \to \mathcal{T}$, where $\mathcal{T}$ is a set of discourse relation types.

While we have seen considerable overlap between the type of discourse relations needed for annotating multiparty dialogue and those used in many theories for the annotation of single-authored text, we have also noted differences. For instance, the most common relations in our *Settlers* corpus are QAP, linking a question and an answer (including partial or indirect answers), QUESTION-ELABORATION or follow up questions, and ACKNOWLEDGMENT. Another relation that is almost non-existent in annotation models for single-authored text is CORRECTION, where one speaker corrects a previous contribution (by herself or someone else). Other frequent relations are: ELABORATION, EXPLANATION, CONTINUATION, PARALLEL, CONTRAST, ALTERNATION, and CONDITIONAL, variants of which are also used in many models for the discourse annotation of single-authored text besides those based on SDRT. On the other hand, other relations, in particular temporal relations like NARRATION, TEMPORAL-LOCATION or BACKGROUND, are not at all frequent in the *Settlers* corpus.

Multi-party dialogue also requires modification of SDRT's constraints on discourse structure. Our annotations like discourse annotations in other applications, do not contain cycles, with E_1 and E_2 forming well founded relational structures. Also, once we adapt SDRT's Right Frontier Constraint modified for multi-party dialogue (Hunter et al., 2015), our structures obey this constraint, which is necessary for coherent discourse structures in SDRT, 95% of the time. In general, our discourse structures follow the SDRT constraints of having a unique root and of connectedness; however, the presence of multiple threads has shown that these assumptions are not always verified in multi-party dialogue. The most plausible explanation of this is that the

structure of a multi-party dialogue is a family of SDRSs, one for each independent thread. Another possibility is that we simply need to relax SDRT assumptions for discourse structures in multiparty dialogue.

3. Details on the Annotation Process

Our first step was to take the game log files (filename.soclog) and isolate from these the turns with linguistic contributions by the players (the result is a file with a .csv tag). We then segmented these turns into EDUs. EDU segmentation was initially performed automatically (with around 90% accuracy) (Afantenos et al., 2010) and then corrected by hand. We then used both naive and expert annotators with a series of adjudications to create the annotations. Annotation proceeds in two layers: dialogue act annotation (§3.1.) and discourse structure annotation (§3.2.). During the training phase, 4 naive annotators were trained on a pilot subset of our data. Two "seasons" of games were then organized—where players were grouped into several leagues and then competed to be the best Catan player—resulting in 59 games out of which 36 have been completely annotated and are part of the current release, while the remaining 23 are nearly complete and part of the next release. The original annotations by naive annotators and the versions revised and adjudicated by experts are part of the corpus.

Originally every game, each one of which contains between 100 and 900 dialogue turns, was to be split into negotiation dialogues (up to several dozen) in which one person is in charge of the bargaining after a roll of the dice. However, we noticed that, often, even if the person in charge of bargaining changed, a conversation that had begun during a previous bargaining episode, continued into the next episode. We therefore ended up splitting our games into "dialogues" that often put several rhetorically connected bargaining sections together.

3.1. Dialogue acts

In the dialogue act structure, each EDU is assigned a particular type; it is classified as either an *offer* (which is typically underspecified in the sense that no particular person is addressed, or no specific number of resources are mentioned, for example), a *counter-offer* (in which a player "replies" to a prior offer by another speaker with a refinement or alternative offer), an *acceptance*, a *refusal*, or *other*. (Sidner, 1994a; Sidner, 1994b) used similar dialogue act labels for a negotiation corpus they collected and annotated. While many of the turns in the chats involve negotiation about trade offers, there is also a significant portion of "other" types of dialogue acts that are not pertinent to negotiations, as in Example 1.

Many of the moves labelled as offers in our corpus do not express a standard offer as understood in Economics, in which an offer is a completely specified proposal to trade a particular number of resources for a specific quantity of some other resource. Our annotation model follows this outline by specifying for each offer a feature structure in which there is a good mentioned to be given by the speaker and a good to be received, as well as an addressee, but often these features must be left unspecified on a given turn. Here is an example.

Example 6

a.	player1	Anybody have any sheep?
b.		or wheat?
c.	player2	no.
d.	player3	$[\text{Sorry}]_i$ $[\text{Need my sheep}]_{ii}$
e.	player1	I have clay.
f.	player3	Still no.

In (a), player1 does not specify a particular addressee, a specific quantity of sheep, or a specific good being offered in exchange; as an offer in a non-cooperative bargaining game (Nash, 1951), therefore, (a) is less specific than required. We nonetheless label (a) as an offer, but lacking a specific addressee, mark the addressee as *all*. The speech act type is specified to be a question. The offer contains a resource, sheep, that is labelled *receivable*, but the quantity of sheep is left unspecified, and there is no *giveable* resource associated with (a). (b) is labelled as an offer with the same features; it too has a receivable resource, wheat, that forms part of the complex type *sheep or wheat*, which is specified by turns (a) and (b) together. We annotate complex types this way to allow other researchers to make their own decisions about how to put the two offers together: do they jointly specify an offer with a receivable resource of complex type, or are they two partial offers? The answer will depend on one's theoretical decisions. (c) is labelled as a refusal by player2 with player1 as the addressee; it is also labelled as an assertion. (d) contains two EDUs, (i) and (ii); (i) is a refusal by player3 directed at player1, while (ii) is labelled *other*. Its relation to (i) is specified by the discourse structure component—similarly for the relation between the underspecified offer in (a)/(b) and its specification in (e).

3.2. Discourse structure annotation

The annotation of discourse structure was a large effort carried out by 4 annotators without special knowledge of linguistics, but who received training over 22 negotiation dialogues with 560 turns. Using an exact match criterion of success, the inter annotator agreement score was a Kappa of 0.72 attachment on structures, 0.58 on labelling for doubly annotated dialogues. Bearing in mind that annotating full discourse structures is a very complex task, this is a relatively good score. Experts made several passes over the annotations from the naive annotators, improving the data and debugging it. We had at least 5 stages of revision. We used the Trello system to maintain our annotations and to keep track of the revisions (trello.com).

The statistics for our corpus are as follows, where EDUs are the elementary discourse units.

	Total	Training	Testing
Dialogues	1081	965	116
Turns	9160	8166	994
EDUs	10678	9546	1132
Relation instances	10513	9421	1092
CDUs	1284	1132	152

Our corpus thus is quite sizeable and has approximately the same number of EDUs and relations as the RST corpus

(Carlson et al., 2003), the only other large corpus with full discourse structures for texts. Here is a table of the absolute frequency of the relations in our corpus.

	Total	Training	Testing
Comment	1851	1684	167
Clarification_question	260	240	20
Elaboration	869	771	98
Acknowledgment	1010	893	117
Continuation	987	873	114
Explanation	437	407	30
Conditional	124	105	19
Question-answer_pair	2541	2236	305
Alternation	146	128	18
Q-Elab	599	525	74
Result	578	551	27
Background	61	58	3
Narration	130	116	14
Correction	212	189	23
Parallel	215	196	19
Contrast	493	449	44
TOTAL	10513	9421	1092

We compared the frequencies of our relations to those in a corpus of texts with a similar, SDRT-based annotation scheme like that described in (Muller et al., 2012) for a corpus of French newspaper and Wikipedia texts (ANNODIS). The most frequent relation in the ANNODIS corpus was CONTINUATION (20.3%) followed by ELABORATION (18.6%), ENTITY-ELABORATION (15.7%) and NARRATION (10.4%). CONTINUATION is the fourth most frequent relation in *Settlers* (9.4%), ELABORATION is the fifth (8.3%) and ENTITY-ELABORATION, used when the second EDU provides more description of an entity mentioned in the first, doesn't occur in our corpus. QAP, which doesn't occur in the ANNODIS corpus but is the most frequent relation in the *Settlers* corpus, occurred 24.1% of the time.

To compare the nature of discourse attachments in the *Settlers* corpus to attachments for texts, we used the measure "LAST", which is a method for predicting the attachment of an EDU to the discourse context. LAST is a baseline that attaches each EDU to the previous EDU in textual order. Over the *Settlers* corpus LAST has a global precision of 60.2%, a recall of 56.6% and an F1 measure of 58.4%. Within turns of a single speaker, LAST had a much higher value but between EDUs of differing speakers, it was lower: with 61.6% in precision, 51.4% in recall and 56.1% for an F1 score. The figures for the base line LAST given in (Muller et al., 2012) for the ANNODIS corpus was significantly higher: 62.4% F1 score. Thus our study of our multiparty chat corpus not only confirmed the observations of (Ginzburg and Fernández, 2005) concerning the frequency of long distance attachments compared to two party dialogue but also suggested that attachments of distance longer than one were more frequent in multiparty dialogue than in text.

Another interesting feature of discourse annotations for dialogue versus text has to do with the direction of attachments. In the *Settlers* corpus, the moves are sequential in the following sense: first one person talks and then others react to them or ignore them, but the discourse links that do occur between speaker turns are *reactive*. In other words, a turn n can't be anaphorically and rhetorically dependent on another speaker's turn that comes after n. Thus, the nature of dialogue, at least as we observed it in the *Settlers* corpus, imposes an essential and important constraint on the attachment process that is not present for monologue or single-authored text, where an EDU may be dependent upon any EDU, later in the ordering or not: in dialogue there are no "backwards" rhetorical links such that an EDU in turn n by speaker a is rhetorically and anaphorically dependent upon an EDU in turn $n+m$ of speaker b with $a \neq b$. (Afantenos et al., 2015) used this observation as constraint on discourse parsing. Within a turn, however, just as in monologue (as is evident from a study of most styles of discourse annotations of text), backwards links occur, as speakers can anticipate what their main point is but introduce it only after certain secondary considerations have been put forward. For instance, in the RST corpus (Carlson et al., 2003), "backwards" rhetorical links occur 16.3% of the time.

4. Experimentation

Our annotation efforts have led to experimental work already detailed in (Cadilhac et al., 2013; Afantenos et al., 2015), though this is the first time that we have described the corpus in detail. The dialogue act annotations have been used to train an automatic classifier for the dialogue act types of EDUs (Cadilhac et al., 2013). We have used the predicted dialogue acts as a feature in our discourse parsing experiments.

Our discourse parsing experiments detailed in (Afantenos et al., 2015), transform the SDRT graphs into dependency structures following (Muller et al., 2012). We did not want to include CDUs in our experiments because our prior experience on the ANR project ANNODIS did not make us confident that we could predict CDUs with any decent degree of accuracy. For a given discourse graph for SDRT of the form (V, E_1, E_2, ℓ), we have as yet no general and reliable method to calculate edges in E_2; and no such method has been presented in the literature. In order to exploit existing decoding methods over local probability distributions that yield tree-like structures, we transformed the graphs in the *Settlers* corpus using a "CDU to Head" transformation first presented in Muller et al. (2012) for SDRT. The strategy transforms hyper-graphs into dependency graphs. We transform our full graphs (V, E_1, E_2, ℓ) into dependency structures (V', E_1, ℓ), with $V' \subset V$ the set of EDUs in V by replacing any attachment to a CDU with an attachment to the CDU's head—the textually first EDU within the CDU which has no incoming links. Our transformation forces E_2 in our general definition of a graph to be $\emptyset$. In the case that we have a discourse relation between two EDUs, this relation is kept intact since it already represents a dependency arc. In case a discourse relation has one or two CDUs as arguments, the CDUs are replaced with their *recursive head*. In order to calculate the recursive head we identify all the DUs with no incoming links; if they are CDUs we recursively apply the algorithm until we get an EDU. If there is more than one EDU with no incoming links we pick the leftmost, i.e. the one firstly introduced in the text. Figure 2

shows an example of such a transformation.[5]

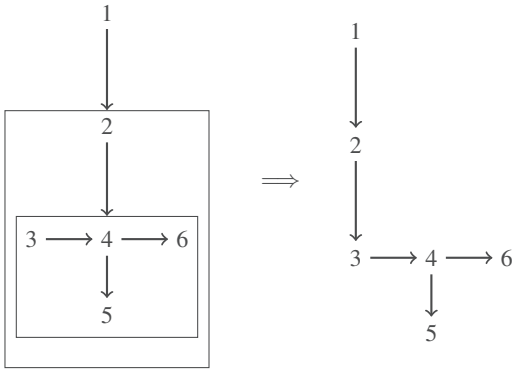


Figure 2: Translation of SDRT discourse graphs into dependency structures.

We then used a linear model (logistic regression) in order to get a probability distribution on attachments and discourse relations for all pairs of EDUs. We call this the *local model*. We then used this local model with a global decoding procedure to produce a full, labelled discourse structure for each dialogue. A tree decoding like algorithm like the Maximum Spanning Tree algorithm of (McDonald et al., 2005) produces results that are roughly comparable to the state of the art in discourse parsing for single-authored text. However, we have also shown that even when the local model is replaced with an oracle that gives the attachment and labelling of the gold annotations, the MST decoding algorithm still misses 9% of the correct attachments in the annotations, which shows the limit of algorithms that restrict predictions to tree-like discourse structures. This prompted us to investigate decoding strategies that can predict DAGs. Perret et al. (2016) shows how the use of integer linear programming to encode constraints on SDRT graphs produces better results on the *Settlers* corpus with the CDU to Head transformation. But our integer linear programming decoder also enabled us to investigate other transformations from SDRT's hypergraphs to dependency graphs such as a CDU Distribution strategy in which an arc between a CDU α and EDU β, $(\alpha, \beta) \in E_1$ in the original SDRS graph (V, E_1, E_2, ℓ) is replaced by a set of arcs $\{\alpha_i \rightarrow \beta : (\alpha_i, \alpha) \in E_2\}$. Carrying through this transformation recursively distributes arcs involving CDUs across their constituents to produce a different sort of dependency graph for an SDRS. And at least for many (though not all) discourse relations, the CDU Distribute strategy is semantically sound given SDRT's semantics for these relations. Figure 3 illustrates the CDU Distribute strategy at work.

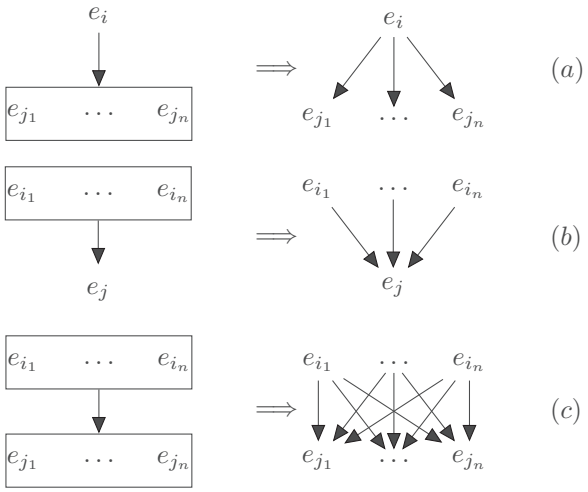


Figure 3: Distributing relations: (a) *right* distribution from an EDU to a CDU, (b) *left* distribution from a CDU to an EDU, (c) from a CDU to a CDU.distributive.

5. Future Work

Our future work will proceed along two lines. First on the corpus side, we are continuing to extend the data set and to revise the annotations. As noted in §1, our data contains rich information about the evolution of the game state itself—so much information that we can replay an entire game from our data files. This provides us an opportunity to study the interaction of the extra-linguistic context, the events that change the game state, with the linguistic context. Preliminary studies in (Hunter et al., 2015) show that the nonlinguistic context can affect discourse structure in several ways, but future work will involve a large annotation effort to determine the discourse relations involved in these dependencies and their effects on discourse structure. On the experimentation side, we plan to attack the problem of hierarchical structure and attempt to compute CDUs automatically. We believe that our corpus using the CDU Distribute transformation provides us the right kind of data to predict CDUs.

[5]Hirao et al. (2013) and Li et al. (2014) followed a similar strategy to that in (Muller et al., 2012) for the creation of dependency structures for RST.

Afantenos, S., Denis, P., Muller, P., and Danlos, L. (2010). Learning Recursive Segments for Discourse Parsing. In *Proceedings of LREC 2010*.

Afantenos, S., Kow, E., Asher, N., and Perret, J. (2015). Discourse parsing for multi-party chat dialogues. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 928–937, Lisbon, Portugal, September. Association for Computational Linguistics.

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

Asher, N. and Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, 6(2):2:1–:62.

Asher, N., Paul, S., and Venant, A. (2015). Conversational goals and achieving them in 'no-win' situations. submitted to *Journal of Logic Language and Information*.

Cadilhac, A., Asher, N., Benamara, F., and Lascarides, A. (2013). Grounding strategic conversation: Using negotiation dialogues to predict trades in a win-lose game. In *Proceedings of EMNLP*, Seattle.

Carlson, L., Marcu, D., and Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In Jan van Kuppevelt et al.,

editors, *Current Directions in Discourse and Dialogue*, pages 85–112. Kluwer Academic Publishers.

Forbes, K., Miltsakaki, E., Prasad, R., Sarkar, A., Joshi, A. K., and Webber, B. L. (2003). D-ltag system: Discourse parsing with a lexicalized tree-adjoining grammar. *Journal of Logic, Language and Information*, 12(3):261–279.

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

Grosz, B. and Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204.

Grosz, B. and Sidner, C. (1990). Plans for discourse. In J. Morgan P. R. Cohen et al., editors, *Intentions in Communication*, pages 417–444. MIT Press.

Grosz, B. (1979). Utterance and objective: Issues in natural language communication. In *Proceedings of the Joint International Conference in Artificial Intelligence (IJCAI 1979)*, pages 1067–1076.

Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., and Nagata, M. (2013). Single-document summarization as a tree knapsack problem. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1515–1520, Seattle, Washington, USA, October. Association for Computational Linguistics.

Hunter, J., Asher, N., and Lascarides, A. (2015). Integrating non-linguistic events into discourse structure. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS)*, pages 184–194, London.

Lascarides, A. and Asher, N. (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics*, 26(2):109–158.

Li, S., Wang, L., Cao, Z., and Li, W. (2014). Text-level discourse dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25–35, Baltimore, Maryland, June. Association for Computational Linguistics.

Mann, W. and Thompson, S. (1987). Rhetorical structure theory : a theory of text organization. Technical report, Information Science Institute.

McDonald, R., Crammer, K., and Pereira, F. (2005). Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting of the ACL*, pages 91–98, Ann Arbor. Association for Computational Linguistics.

Muller, P., Afantenos, S., Denis, P., and Asher, N. (2012). Constrained decoding for text-level discourse parsing. In *Proceedings of COLING 2012*, pages 1883–1900, Mumbai, India, December. The COLING 2012 Organizing Committee.

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, September.

Perret, J., Afantenos, S., Asher, N., and Morey, M. (2016). Integer linear programming for discourse parsing. In *Proceedings of the 2016 North American Chapter of the Association for Computational Linguistics*, pages 928–937, San Diego, CA, USA, JUne. Association for Computational Linguistics.

Péry-Woodley, M.-P., Asher, N., Enjalbert, P., Benamara, F., Bras, M., Fabre, C., Ferrari, S., Ho-Dac, L.-M., Draoulec, A. L., Mathet, Y., Muller, P., Prévot, L., Rebeyrolle, J., Tanguy, L., Vergez-Couret, M., Vieu, L., and Widlöcher, A. (2009). Annodis : une approche outillée de l'annotation de structures discursives. In *Actes de TALN 2009*, Senlis, June. ATALA, LIPN.

Polanyi, L., Culy, C., van den Berg, M., Thione, G. L., and Ahn, D. (2004). A rule based approach to discourse parsing. In Michael Strube et al., editors, *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 108–117, Cambridge, Massachusetts, USA, April 30 - May 1. Association for Computational Linguistics.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008). The penn discourse treebank 2.0. In N. Calzolari, et al., editors, *Proceedings of LREC'08*, Marrakech, Morocco, may. ELRA. http://www.lrec-conf.org/proceedings/lrec2008/.

Sidner, C. (1994a). An artificial discourse language for collaborative negotiation. In *AAAI*, volume 1, pages 814–819. MIT Press, Cambridge.

Sidner, C. (1994b). Negotiation in collaborative activity: a discourse analysis. *Knowledge-Based Systems*, 7(4):265 – 267.

Traum, D., Swartout, W., Gratch, J., and Marsella, S. (2008). A virtual human dialogue model for non-team interaction. In L. Dybkjaer, et al., editors, *Recent Trends in Discourse and Dialogue*, volume 39 of *Text, Speech and Language Technology*, pages 45–67. Springer Netherlands.

W. Wahlster, editor. (2000). *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer.

Wolf, F. and Gibson, E. (2005). Representing discourse coherence: A corpus based study. *Computational Linguistics*, 31(2):249–287.