

Situated Dialogue

Nicholas Asher and Julie Hunter

Meeting 3
July 10, 2016

Last Class

- we criticized the received view on L and EL contexts from a philosophical and a linguistic perspective.
- we also showed that it was incomplete with a very limited view of L/EL interactions with a simplistic typology of context dependent expressions.
- Argued for a presuppositional account for indexicals exploiting unified L and EL contexts.

More on the last class

- We also criticized the received view of intralinguistic connections
- and introduced a richer view of intralinguistic dependencies that include discourse or coherence relations.
- We showed how these have been argued to play a role in resolving temporal and pronominal anaphora.

More on the Last Class

- We looked at prior work that uses discourse structure to constraint analyses of co-verbal gesture and deixis.
- We extended that idea to show how rhetorical connections might be important in accounting for an integration of iconic gestures that are not coverbal.
- Finally we introduced work on discourse structure which showed how EL actions by conversational participants might play a role similar to linguistic turns in structuring discourse. A different view of intra-linguistic relations.

More on the Last Class

- co-verbal gesture and gestural moves that complete linguistic utterances show that there is a lot more interaction between EL and L moves than the received view assumes (just deixis)

This Class

- Prior studies on co-verbal gesture and deixis are interesting but don't involve detailed corpus work.
- We will briefly look at some recent corpus work on gesture
- and then move on to our corpus on EL events that aren't co-verbal gestures but nevertheless have a rich interaction with the linguistic context.

Restaurant Game & Tacos

- just demonstrations, but virtual demonstrations!
- trying to get an automated agent to interact naturally with a human in a particular situation
- Tacos: a corpus of cooking descriptions (Jamie Oliver's assistants clips). And there's lots more see Vah-Chef's cooking videos, Jamie Oliver's cooking videos and the French TV show: Top Chef, where famous chefs disparage other chefs attempts to achieve some impossible goal (like making a meal whose main ingredient was bread).

Bielefeld SaGA corpus

- The Bielefeld Speech-and-Gesture-Alignment- corpus (SaGA), Rieser (*Semdia1*, 2011), Bergmann et al. *Sigdia1 2011*, Lücking et al. (2013))
- route descriptions and comparisons in 25+ dialogues with 2 or 3 people.
- over 6000 gestures total.

Gesture typology

Some gestures described both with a set of basic features but the novel aspect of this study is to divide gestures into several types based on their semantics:

- iconic (3165),
- deictic (1311)
- discursive (1223),
- mixed (900+)

More on discourse gestures

- discourse gestures related to turn taking —current speaker points to other selecting him as next.
- discourse gestures indicating acknowledgment of *A* by re-using an iconic gesture of *A*. (also just nodding!)
- discourse gestures indicating an assessment of evidence, denoting either the fit of a description (spreading hand and shaking of wrist)
- assessing the adequacy of a description's content for accomplishing the recipient's goal (lifting hand + headshake)

Assessment

- Bielefeld corpus uses a mix of different coherence relations + concepts from conversational analysis
- offer a more detailed typology than in previous work, though the authors aren't very explicit as to exactly what is content of the accompanying speech.
- could easily be added to SDRT's view of coherence relations (notice the gestural typology for Commentary).
- Have to do something about turn taking.

Corpora efforts at ICT, Los Angeles

- SASO, Tactical Questioning (TACQ), annotations multi-party multi-modal task oriented dialogue.
- with the aim to provide Mission Rehearsal Exercise a simulated environment for practicing negotiations with virtual agents (Traum et al. 2008, Swarthout et al. 2006).
- body position and facial expression (see also Justine Cassell's work)
- all of these can affect how we interpret a message

The Ami corpus (University of Edinburgh)

- Augmented Multi-party Interaction Project (I. McCowan et al.)
- 100 hours of meetings in which audio (multiple microphones), video (multiple cameras) slides (data projector) and textual information (associated papers, captured handwritten notes and whiteboard) are all recorded and time synchronized.
- looks downloadable from the web.
- detailed annotation including speech transcription, named entities, dialogue acts for decision making, some relations between acts, topic segmentation, head and hand gestures including deixis gestures, location of individuals in room and posture while seated, location of participant faces and hands within video frames, foci of attention.
- no annotation of what was pointed to in deixis.

Current state of empirical work on situated discourse structure

- Bielefeld corpus looks at some local coherence relations between gesture and linguistic input.
- restricted to gesture
- no work on non coverbal gestural events interacting with linguistic moves other than our own efforts as far as we are aware.

What are non-verbal gestural events?

- body stance seems to lie outside the realm of coverbal gesture.
- This is a difficult question and the concept has murky boundaries
- gestural actions that aren't performed roughly simultaneously with speech (some instances in the Bielefeld corpus?)
- actions by a dialogue participant that directly manipulate objects for some practical end—like Julia Child's making of the omelette, even though those actions occur with speech overlaid, describing what she's doing.
- perceived actions that aren't accomplished with gestures

Why is it so difficult to get empirical work on this topic?

One problem is that there is a lot of visual data at any given moment. Think of the difference of a movie vs. the information contained in its script.

But the problem isn't just one of quantity...

Two problems: individuation and conceptualization

- visual data (we'll think of it simplistically as an array of pixels) has to be arranged and conceptualized in a certain way in order for agents to be able exploit it.
- Objects and events have to be individuated as parts of the visual scene and given a conceptualization.
- What are the right or relevant concepts that should be applied in a particular situation?
- Part of this involves deciding what parts of the visual scene are relevant for the communication at hand.

Go back to the Julia Child video

Given our definition, the video includes a lot of non-verbal actions but are directly relevant to the narrative.

- we immediately apply concepts not linguistically expressed but that are relevant in the Julia Childs video: *pan, eggs, shaking, wrist,...*
- When the relevant part starts and stops (individuation): making the clip for the video
- Which concepts and combinations of concepts are relevant?
- Same video clip could have been used to advertise the non-stick pan; omelette would have been irrelevant.

Conceptualization continued

- Humans are very good at conceptualizing visual scenes that they process.
- indications from contemporary neurophysiological theories of vision are that vision is not a bottom up process (as was portrayed in the 80s, cf. David Marr)
- but rather a system in which bottom up processing interacts with top down cognitively induced expectations.

Language can affect top down expectations

- Re the video: we are told this is a cooking video, and Julia Child's words guide us to attend to certain parts of the visual scene as relevant.
- Prior linguistic context guides the cognitive expectations that are crucial.
- A linguistic description can allow us to see things in a visual scene that otherwise we would not.

A more radical view

- an action is a communicative action when I can embed it in a coherent discourse structure involving prior L and EL contexts that give rise to continuations from which one would infer communication has taken place.

Pushing the EL /L connection further

- So far we've been looking at co-verbal actions, where there is a direct grammatical tie between language and EL events
- or actions that verify or illustrate an accompanying verbal description.
- Situation gets trickier with EL L dependencies that aren't .
- Video of a real picasso like situation?
- Difficult because there is no co-present linguistic content that summarizes what is going on and tells you which concepts are important. Nevertheless, a particular grouping of concepts is necessary for understanding the discourse.

But there is language that's related

- prior linguistic moves can help in the conceptualization of an EL event.
- so can linguistic moves that are subsequent to the action.
- so can causally related other EL events that are linked to L moves.

What cues do we use?

- Hypothesis: extend the idea that we work it out as a matter of rhetorical reasoning.
- If we could get a hold on this, we could take a step forward in studying video-based corpora.
- Problem: co-dependency. no theory of how adding extra-linguistic information affects discourse structure.

A more detailed discussion of the Picasso example

Suppose Anne's husband Abner comes home to find his house abnormally quiet and Anne looking upset. He looks at her inquiringly and she says:

Example

(a) Our little Picasso has been sent to her room.

Just after she says this, she glances over her shoulder and her husband, taking her cue, spots graffiti on the living room wall behind her.

An informal analysis

- Anne can immediately interpret the inquiring look as asking a question as to why there is no children's patten.
- Her verbal move answers the question
- Her glance, together with the use of the NP *our little Picasso* points to the result of a particular EL event of drawing on the wall that explains why their daughter was sent to her room.

Continuing with the example

Call this drawing event 'e'. It can serve as an antecedent to a subsequent utterance:

Example

(b) I was cooking dinner.

The non-linguistic event *e* provides the temporal referent for the past tense in (b).

Interim conclusions

Intuitively, EL actions can stand in the same rhetorical or coherence relations to L discourse moves that other L moves do.

We can even imagine that two EL actions that could form a question answer pair:

- Abner gives an inquiring look.
- Anne gestures to the drawing on the wall.

So it looks like it makes sense to assign coherence relations EL actions as well as speech acts, at least in some cases.

Why this isn't crazy

- Coherence relations are very general conceptual relations
- Andy Kehler's use of Hume's relations of causality, resemblance and contiguity as a source for these relations.
- But perhaps a better source for ideas of conceptual coherence might be Kant.
- These relations are not only important for relating linguistic moves, but they are constitutive of experience, and hence relate EL events as well.

Place to start: a controlled but natural situation.

Settlers of Catan!

Extra-linguistic events

You guys are gonna hate me... [plays soldier card]

[Player builds a road and immediately after, builds a settlement] (You're a Fast mover!

Come on lucky 7! [rolls a 4 and a 2] Almost...

- Both EL and L content matter for ellipsis

Because we're in a virtual chat environment, there are no co-verbal gestures that are perceptible by all the players.

More relations and not necessarily demonstrative (though a lot of that too).

But, the advantage is that the game rules tell us a bit about what is important and how to think of the happenings in the extra-linguistic context.

The corpus

The game

Settlers of Catan

- multi-party, win-lose game
- players use resources (wood, clay, ...) to build roads and settlements
- board: multiple regions, each assigned a resource and number (2 - 12)
- players get resources by rolling dice, trading, or stealing
- robber: roll of a 7; discard, steal, move

The corpus

The board

Settlers of Catan Game: pilot01 [Markus]

rennoel
Points: 2

Soldiers: 5
Resources: 3
Dev. Cards: 0

Tomm
Points: 4
I. Road

Soldiers: 1
Resources: 5
Dev. Cards: 0

Game

* K's rennoel's turn to roll. Rolled a 11.
* rennoel gets 1 wheat. Tomm gets 1 wheat.

rennoel: you know if you have more than 4, I think you can trade them into the bank.
Dave: yeah but it's not ideal
Tomm: That's true. - 4 -> 1
Tomm: Well, I might do, but depends on my roll, I'm afraid
Tomm: Dave. Wheat for a clay?
Dave: sure. can you do 2 for 2, or do you just want 1 for 1
Tomm: just 1-4-1 I'm afraid
Dave: far enough
Tomm: Oh.. now I get wheat(!)
Dave: heh

Histor

Chat

Dave
Points: 2

Soldiers: 4
Resources: 3
Dev. Cards: 2

58 Here

Road:	Cost: 1 1
Settlement:	Cost: 1 1 1 1
City Upgrade:	Cost: 3 2
Card:	Cost: 1 1 1 1 available

Game Options...

Players were told to discuss their trades via a chat interface before confirming them non-linguistically

- 59 games, each with dozens of negotiation dialogues with 1-30+ turns,
- >1000 dialogues annotated with complete discourse structures for linguistic turns (11000 EDUs in total)
- in the style of SDRT (Asher & Lascarides 2003)

Players ended up discussing many other things as well.

- Non-linguistic events, temporally ordered with linguistic events

(3) 100 Server player i offers 1 wheat for 1 sheep from player j
101 **player i or an ore**

(4) 100 Server player i rolled a 2 and a 1
101 Server player j gets 2 sheep. player i gets 1 wheat
102 **player i Woo!**

The situated corpus

- Our game transcripts contain a complete log of the events both L and EL that happened during the game.
- The logs allow us to play back an entire game, should we so wish.
- Using these logs, we added EL events to our linguistic corpus, creating a situated STAC corpus.
- Many more turns were added to a game, reflecting the EL events that players engaged in that directly involved the game.
- The visual sequence of game states introduces many more elements that play a “discourse” function like in the Picasso e.g..
- Players often responded with actions in lieu of verbal moves to both L and EL moves.

Some sample stats for one given short game

Type	Non situated game	Situated Game
EDUs	340	970
Turns	281	828
Dialogues	30	67
CDUs	28	117

- the situated game contains almost three times as many eventualities as the non-situated or purely linguistic transcript.
- there are “EL dialogues”, where players propose exchanges and others accept or reject them and where other events can happen where no words are exchanged.
- the linguistic corpus contains around 11 000 EDUs and about the same number of relations.
- extrapolating from this game our situated corpus when completed will contain over 25K EDUs.

More on the character of EL events in our corpus

The EL events in our corpus are directly related to the game. They involve:

- game set up moves
- trades, requests for trades (which linguistically are almost always expressed as questions)
- responses to trade requests
- rolling dice, playing certain cards (soldier card), moving the robber
- building events (building roads, settlements, cities)

Objects and relations

- the objects that are actors in the events are the players, the bank, ports, resource amounts, and various cards. A pretty simple ontology.
- many of the non-linguistic events follow a set pattern.
- some are directly causally or logically related to each other
- a distribution of resources causally depends on the dice roll and the game state; and playing a soldier card implies that you must move the robber.

CDUs and conceptualization

- On our annotation scheme, simple events can form more complex events that are in our annotation framework CDUs. (Robber CDU or Distribution of Resources CDU)
- What are the natural larger or complex events?
- This is a conceptualization question, but the game context helps us here.

Nevertheless our corpus simplifies things

- Bypass the Individuation and Conceptualization problems for atomic events and of the objects that are actors in those events.
- Server and User Interface messages recorded in game log.
- the game log gives the basics of our conceptualization of this virtual world.

Non-linguistic?

Are these events really “non-linguistic”?

Yes.

- UI information not encoded in English, but accessible in code; allows replay of games (Robber placement, turn changes, etc.)
- Even Server messages are non-linguistic: all non-linguistic events must be conceptualized
- players do not rely on the messages; they are for the record

Non-linguistic?

Are these events really “non-linguistic”?

Yes.

- UI information not encoded in English, but accessible in code; allows replay of games (Robber placement, turn changes, etc.)
- Even Server messages are non-linguistic: all non-linguistic events must be conceptualized
- players do not rely on the messages; they are for the record

Going beyond the state of the art

- in co-verbal gesture, the physical events really are parasitic on the accompanying linguistic message (highly underspecified content)
- the actions in our game are not underspecified at all. Rather, often the linguistic messages are:
- E.g., *Does anyone want a sheep?* is a typical but underspecified offer to exchange a sheep for something else with someone.
- A corresponding EL event content: Tom made an offer to trade 1 sheep for 1 clay from Sam (along with a time stamp).
- Nevertheless our EL events enter into coherence relations with L and EL events, and we can study these and see how they differ from coverbal gestures.
- One difference: Result is a very frequent relation between EL events and between EL events and L moves. This relation doesn't occur with co-verbal gesture as far as we know and perhaps shouldn't because of the temporal overlap constraints on gesture and speech.

- Because we are looking at complete discourse structures with respect to the set of L and EL moves as specified by our conceptualization, we can also investigate how EL events change the overall discourse structure by comparing the situated annotations with non-situated annotations.
- Not possible with the simplified scheme of discourse coherence in SS&L.

Modelling interaction

Different types

Question-Answer-Pair:

- | | | | |
|-----|--------------|-------------|--|
| 630 | 18:23:16:042 | Server | rennoc1 made an offer to trade 1 clay, 1 sheep, 1 wood for 1 ore, 2 wheat. |
| 631 | 18:23:28:225 | Dave | don't have 2 wheat |

Broad range of interactions:

Alternations, Comments, Conditionals, Elaborations, Explanations:

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

Modelling interaction

Different types

Question-Answer-Pair:

- | | | | |
|-----|--------------|--------|--|
| 630 | 18:23:16:042 | Server | rennoc1 made an offer to trade 1 clay, 1 sheep, 1 wood for 1 ore, 2 wheat. |
| 631 | 18:23:28:225 | Dave | don't have 2 wheat |

Broad range of interactions:

Alternations, Comments, Conditionals, Elaborations, Explanations:

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

Modelling interaction

Different types

Question-Answer-Pair:

630	18:23:16:042	Server	rennoc1 made an offer to trade 1 clay, 1 sheep, 1 wood for 1 ore, 2 wheat.
631	18:23:28:225	Dave	don't have 2 wheat

Broad range of interactions:

Alternations, Comments, Conditionals, Elaborations, Explanations:

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

Modelling interaction

Interweaving

214	18:53:13:789	Server	dmm rolled a 2 and a 5.
214.1	18:53:13:789	Server	CCG needs to discard.
215	18:53:17:259	niko	ouch
218.4	18:53:48:461	Server	dmm stole a resource from CCG
219	18:53:57:782	dmm	equal opportunities attack this time
221	18:54:14:284	dmm	aaand just for good measure
222.1	18:54:15:958	Server	dmm played a Soldier card.
222.4	18:54:24:358	Server	dmm stole a resource from niko
225	18:54:27:697	dmm	sorry guys ;)
230	18:54:40:739	dmm	i still didn't get enough stuff to do anything with!

Modelling interaction

Adding eeus

Elementary discourse units (edus): $\pi_1^i, \pi_2^i, \pi_3^i, \dots$ (for player i)

Elementary event units (eeus): $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$

For every ε , there is a formula $\varepsilon: \phi$, for a 1st-order formula ϕ .

The interpretation of ϕ in the relevant model determines the conceptualization of ε .

Speaker interests determine space of preferred models (*Settlers* game)
See IWCS paper for more details.

Modelling interaction

Adding eeus

Elementary discourse units (edus): $\pi_1^i, \pi_2^i, \pi_3^i, \dots$ (for player i)

Elementary event units (eeus): $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$

For every ε , there is a formula $\varepsilon: \phi$, for a 1st-order formula ϕ .

The interpretation of ϕ in the relevant model determines the conceptualization of ε .

Speaker interests determine space of preferred models (*Settlers* game)
See IWCS paper for more details.

Modelling interaction

Adding eeus

Elementary discourse units (edus): $\pi_1^i, \pi_2^i, \pi_3^i, \dots$ (for player i)

Elementary event units (eeus): $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$

For every ε , there is a formula $\varepsilon: \phi$, for a 1st-order formula ϕ .

The interpretation of ϕ in the relevant model determines the conceptualization of ε .

Speaker interests determine space of preferred models (*Settlers* game)

See IWCS paper for more details.

Modelling interaction

Graphs

Discourse structure for d : the connected graph $G^d = (V^d, E_1^d, E_2^d)$

- V^d : the set of edus and cdus in d
- $E_1^d \subseteq V^d \times V^d$: the set of labelled discourse attachments
- $E_2^d \subseteq V^d \times V^d$: the part/whole relation for dus and cdus

Interaction calls for *situated discourse graphs*: $G^{sd} = (V^{sd}, E_1^{sd}, E_2^{sd})$

- V^{sd} : V^d plus eeus and cdus containing eeus
- $E_1^{sd}, E_2^{sd} \subseteq V^{sd} \times V^{sd}$: extended similarly

Modelling interaction

Graphs

Discourse structure for d : the connected graph $G^d = (V^d, E_1^d, E_2^d)$

- V^d : the set of edus and cdus in d
- $E_1^d \subseteq V^d \times V^d$: the set of labelled discourse attachments
- $E_2^d \subseteq V^d \times V^d$: the part/whole relation for dus and cdus

Interaction calls for *situated discourse graphs*: $G^{sd} = (V^{sd}, E_1^{sd}, E_2^{sd})$

- V^{sd} : V^d plus eeus and cdus containing eeus
- $E_1^{sd}, E_2^{sd} \subseteq V^{sd} \times V^{sd}$: extended similarly

Modelling interaction

Graphs

Discourse structure for d : the connected graph $G^d = (V^d, E_1^d, E_2^d)$

- V^d : the set of edus and cdus in d
- $E_1^d \subseteq V^d \times V^d$: the set of labelled discourse attachments
- $E_2^d \subseteq V^d \times V^d$: the part/whole relation for dus and cdus

Interaction calls for *situated discourse graphs*: $G^{sd} = (V^{sd}, E_1^{sd}, E_2^{sd})$

- V^{sd} : V^d plus eeus and cdus containing eeus
- $E_1^{sd}, E_2^{sd} \subseteq V^{sd} \times V^{sd}$: extended similarly

Modelling interaction

Graphs

Discourse structure for d : the connected graph $G^d = (V^d, E_1^d, E_2^d)$

- V^d : the set of edus and cdus in d
- $E_1^d \subseteq V^d \times V^d$: the set of labelled discourse attachments
- $E_2^d \subseteq V^d \times V^d$: the part/whole relation for dus and cdus

Interaction calls for *situated discourse graphs*: $G^{sd} = (V^{sd}, E_1^{sd}, E_2^{sd})$

- V^{sd} : V^d plus eeus and cdus containing eeus
- $E_1^{sd}, E_2^{sd} \subseteq V^{sd} \times V^{sd}$: extended similarly

Additions to discourse relation types

- While SDRT countenances causal relations (Explanation, Result) that can apply equally well to *eeus* and *edus*,
- the temporal relations (Background, Narration, Flash-back) come with topic constraints that *eeus* don't fit.
- So we have added a *Sequence* relation between *eeus* to model the temporal succession of *eeus*.

Modelling interaction

Sub-graphs

Given G^d and G^{sd} , is G^d a sub-graph of G^{sd} ?

- $V^d \subseteq V^{sd}$
- $E_1^d \subseteq E_1^{sd}$?
- $E_2^d \subseteq E_2^{sd}$?

Modelling interaction

Sub-graphs

Given G^d and G^{sd} , is G^d a sub-graph of G^{sd} ?

- $V^d \subseteq V^{sd}$
- $E_1^d \subseteq E_1^{sd}$? No.

449 20:50:09:097 niko now you're on 3sy street

454 20:50:19:376 william yay

- $E_2^d \subseteq E_2^{sd}$?

Modelling interaction

Sub-graphs

Given G^d and G^{sd} , is G^d a sub-graph of G^{sd} ?

- $V^d \subseteq V^{sd}$
- $E_1^d \subseteq E_1^{sd}$? No.

449	20:50:09:097	niko	now you're on 3sy street
452	20:50:29:293	Server	niko rolled a 4 and a 1.
453	20:50:29:294	Server	william gets 1 wood.
454	20:50:19:376	william	yay

- $E_2^d \subseteq E_2^{sd}$?

Integrated discourse structures

Sub-graphs

Given G^d and G^{sd} , is G^d a sub-graph of G^{sd} ?

- $V^d \subseteq V^{sd}$
- $E_1^d \not\subseteq E_1^{sd}$
- $E_2^d \not\subseteq E_2^{sd}$:

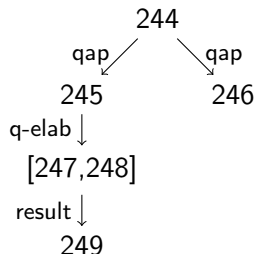
An example of a new CDU

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 1 sheep, 1 wheat.

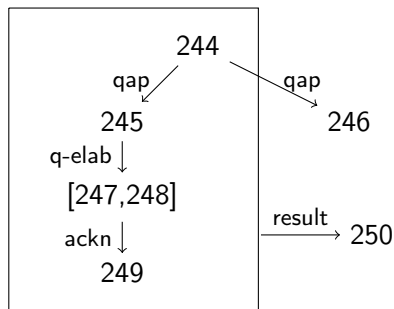
The presence of EEU 250 triggers the construction of a new CDU over linguistic elements.

The discourse structure with the new CDU

The "language only" SDRS



The situated SDRS with new CDU

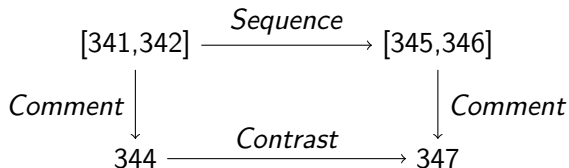


new types of discourse structures

341	19:05:26:615	Server	gotwood4sheep rolled a 6 and a 3.
342	19:05:26:616	Server	inca gets 2 wheat. dmm gets 1 wheat.
344	19:05:29:595	gotwood4sheep	9 nooo!
345	19:05:34:924	Server	inca rolled a 1 and a 3.
346	19:05:34:926	Server	gotwood4sheep gets 2 wood.
347	19:05:39:655	gotwood4sheep	4 better :)

- gwfs's comment in 347 on the cdu formed from eeus 345,346 also clearly attaches to 344,
- which was a comment on the eeus 341, 342.

A “box” structure



This simultaneous linking to elements in a sequence of events and to available linguistic moves is particular to situated dialogue.

- Extend rhetorical theories to situated dialogue
- Extension is non-trivial
- Linguistic clues were not hopeless in determining the discourse structure.
- Still, largely insufficient to get the full picture.

Non-linguistic events affect shape and change interpretation of discourse

ME: missing units; ML: missing links; MC missing CDUs; IL: incorrect links; BB: wrong dialogue breaks;
TE: total # errors;

game	ME	ML	MC	IL	BB
s1-league1-game3	122	162	26	44	6
s2-leagueM-game2	78	119	15	25	3
pilot14	72	115	6	25	2

game	TE	DU errors	link error	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: Error rates on Settlers games

Why our corpus an important stepping stone to understanding situated communication

- our corpus does not have co-verbal gesture; there is no virtual presence of the players in the virtual game space except as static avatars.
- so we can concentrate how already conceptualized EL events interact with the discourse structure.
- this has allowed us to pair the typically very rich EL context down to a few salient and relevant EL events.
- we have seen that these interactions even with these eventualities and linguistic discourse structure are very rich.

What are the limitations of our corpus?

- it bypasses the problem of conceptualization
- the ontology is fixed and overly simple.
- part of what can happen in situated conversation is that one can learn both about the meanings of expressions and about what events are taking place around us.
- Our corpus has only a limited number of such instances of learning —people learn how to make a trade in the virtual world by discussion.

What is the plan for extending our results to the problem more generally?

- more instances of situated discourse structures in different genres are important.
- corpus work might be boring and take a long time (the Stac corpus took 5 years) and is inherently imperfect. But it's necessary.
- studying the situated discourse structure of the AMI corpus, for instance, would probably yield a complementary corpus with additional insights.
- studying perhaps child/caregiver interactions (CHILDES corpus)
- moving from the conceptualized virtual environment to the real world environment.
- tackling the co-dependence between linguistic moves, event conceptualization and interactions between EL and L events.

Putting it all together

How does our study fit in with work on demonstrations and gesture?

- we have seen that there are many ways in which language exploits EL contexts
- deixis and the establishment of a denotation for a referring term is one way.
- the use of an EL event to complete, add to, or emphasize a linguistically given meaning in co-verbal gesture
- co-verbal gesture as introduced by McNeil 1992 and Kendon 2004 showed that some EL events were so closely tied to language that linguistic or grammatical constraints could apply to them.

More on the continuum between linguistic and non-linguistic means of communication

- semantic relations between EL events so that one can function as an iconic sign for another.
- quotation devices to change normal signs into iconic ones, and to change EL events into iconic signs.
- the use of EL events to stand in rhetorical relations to linguistically given discourse moves.
- there is a continuum as to how “linguistic” the events really are that interact with language.

What does it mean to be a language?

- Let us assume that every language L is defined by its grammar, consisting (at least) of a lexicon, a syntax and a semantics.
- At a most abstract level, the words of the lexicon are event types of some sort—e.g., the signs of SL or the words of a spoken language—that are realized in utterances or inscriptions as particular events.
- these event types are stable across language use of L over short periods of time.
- These event types/events \mathcal{E} have a semantics: a function sem from \mathcal{E} into a set of other (NL) eventualities \mathbb{E} (at this level of generality, we may consider objects as eventualities).
- typically if $\text{sem}(e) = e'$, then $e \neq e'$.

What does it mean to be a language?

- Given our definition above, languages with iconic expressions in their lexicons (SL, and spoken L) are languages
- But what of a system that makes use of EL events that would have been non-communicative actions in another situation of use?
- It's as though EL events on a particular occasion of use can have a communicative function.
- the “lexicon” of situated conversation in L is not stable.
- Also EL events in our corpus once conceptualized don't “denote” anything else (unlike gestural signs for the most part)

More on situated communication

- And we 've shown that typical situated conversations make use of such EL events.
- As such, situated conversations can go beyond our perhaps narrow definition of a language

Languages and communicative systems

Ultimately, how “linguistic” is the problem that we are studying?

- We've shown that EL events, that can interact with discourse moves.
- These moves aren't language exit or entrance moves in the sense of Sellars (*hello, good-bye*).
- They occur within and throughout an extended conversation.
- So they are intimately connected with linguistic moves and serve functions that linguistic moves also serve.

- Language is connected to the world in a variety of ways.
- one way that philosophers and semanticists have studied is through a referential or denotational semantics (19th 20th cents).
- but the way that project was formulated precluded a full view of L and EL interactions.
- this course offers a glimpse of what a fuller view of such interactions might look like.
- understanding language by understanding its connections to the world and how it makes use of EL events in a situated communication system.

Consequences of our problem for the analysis of conversation

- (real) conversations are embedded in an EL context
- and typically make use of various features of the EL contexts and events that transpire there.
- a full understanding of a conversation requires information of the EL context.
- EL events can play a variety of discourse functions.

Consequences for a theory of contexts that affect meaning

- an analysis of the structure of a conversation also requires an integrated representation of the EL and L context,
- or at least those parts of the EL context that are salient to the conversational participants and enter into aspects of the conversation.
- the received view of an EL context completely separate from the L context is incomplete and the wrong generalization.

Consequences of our problem for semantics and pragmatics

- the first consequence is iconicity; sometimes the relationship between a (simple) sign and its meaning is not always arbitrary
- A proper formal semantics for a full language must be one that can model such a relationship and compositionally build meaning involving multi-modal sources of information.
- Once we can assign iconic signs a formula whose interpretation has the appropriate content, then composition can then proceed as usual.

Incorporating EL events has a bigger impact on semantics

- A semantics for conversations has to also be able to explain how the language can be embedded in a wider communicative system in which L events and EL events interact.
- So the problem of conceptualization has to be addressed.
- Once again there is the question of compositionality, but a solution to the problem of conceptualization, we've argued, can provide a compositional analysis of conversational meaning involving both L and EL events.

Conclusions

- Language is connected to the world in a variety of ways.
- Philosophers and semanticists have studied and formalized this connection through a referential or denotational semantics (19th 20th cents).
- but the way that project evolved precluded a full view of L and EL interactions.
- this course offers a glimpse of what a fuller view of such interactions might look like.
- understanding language by understanding its connections to the world and how it makes use of NL events in a situated communication system.

Thank You!