

Situated Dialogue

Nicholas Asher and Julie Hunter

Meeting 2
July 10, 2016

Yesterday's class

- We looked at various ways that the extra-linguistic context in which a dialogue is situated can have semantic effects on linguistically expressed content.
- These effects come into the content and interpretation of the proposition itself: deictic elements, iconic gestures or other bodily actions.
- We looked at some of the ways that these phenomena have been modeled.

This morning's meeting

We're going to motivate and then propose revisions to the received view on extra-linguistic and intra-linguistic context dependence

But first

- question the strict separation between extra-linguistic and linguistic information
- and a classification of linguistic expressions into terms that either are insensitive to any context, sensitive to a linguistic context only or sensitive only to the extra-linguistic context.
- resolving the split requires a unified underlying mechanism. Look at presupposition as a candidate.
- see how using presupposition leads to a different account of indexicals and removes the strict separation while preserving its motivating virtues.

looking at intralinguistic context sensitivity

- discuss shortcomings with the treatment of linguistic context sensitivity of a simple dynamic semantic account.
- motivate the need for coherence relations.
- what does the intra-linguistic context look like when we add coherence relations?
- look at more recent theoretical efforts that argue for the need for coherence relations in the analysis of extra-linguistic context sensitivity

The split

Kaplanian view:

- primitive distinction between free variables and bound variables

Dynamic view:

- More complicated understanding of anaphora, but still a division between linguistic and extra-linguistic context sensitivity. (even in more recent work by, e.g., Kamp and Maier)

Concerns about the split

Intuitive similarity of deictically and anaphorically used pronouns, but also:

- tense and some 'pure' indexicals crossover the divide
- definite descriptions and proper names
- words like *local*, *ahead*, *enemy*
- resolution of certain presuppositions: *too*, *stop*

Free indirect discourse, but more mundane narrative environments as well:

- The letter is marked “personal and private” and is addressed to President Franklin D. Roosevelt’s secretary, Grace Tully... The writer was Lucy Mercer Rutherford, who decades before had been FDR’s mistress and who **now** was making arrangements for what would be their last fateful meeting...

Concerns about the split

Intuitive similarity of deictically and anaphorically used pronouns, but also:

- tense and some pure indexicals crossover the divide
- definite descriptions and proper names
- words like *local*, *ahead*, *enemy*
- resolution of certain presuppositions: *too*, *stop*

Definite descriptions and proper names

Crossover with definite descriptions

- (a) Pull in behind **the BMW**.
- (b) If I can choose between a Mercedes and a BMW, I'll take **the BMW**.

Bound uses of proper names

- (c) If a child is christened 'Bambi', and Disney Inc. hear about it, then they will sue **Bambi's** parents.
- (d) Mary is under the illusion that she has a son named 'John' and she believes that [her son/**John**] is the thief.

Examples (b)-(d) from Geurts (1997)

Concerns about the split

Intuitive similarity of deictically and anaphorically used pronouns, but also:

- tense and some pure indexicals crossover the divide
- definite descriptions and proper names
- words like *local*, *ahead*, *enemy*
- resolution of certain presuppositions: *too*, *stop*

Local

The Times had every reporter cover a **local** athlete

Nunberg (1993), p. 3

Concerns about the split

Intuitive similarity of deictically and anaphorically used pronouns, but also:

- tense and some pure indexicals crossover the divide
- definite descriptions and proper names
- words like *local*, *ahead*, *enemy*
- resolution of certain presuppositions: *too*, *stop*

Presupposition triggers

- A: I'd like some more water, please. B: I'd like some **too**.
- A: [hands glass over to C; C fills glass with water] B: I'd like some **too**.
- It **stopped** raining!

Towards a more unified account

Is there a single mechanism underlying the different uses of these expressions and constructions? If so, what does it look like?

- we've looked at dynamic models of tense and anaphoric binding
- but deictic expressions, proper names and presupposition triggers have a descriptive meaning not captured by variables alone
- what do we do with that?

Descriptive content

The donkey kicks the farmer.

- Russell: $\exists x(\text{donkey}(x) \wedge \forall y(\text{donkey}(y) \rightarrow y = x))$

If a farmer owns a donkey, **the donkey/it** kicks him.

- Russell's analysis is unmotivated here
- what is the role of the descriptive content?
- (similar questions arise for *it* and *him*)

Heim (1988), Strawson (1950)

Definites and presupposition

Definite descriptions and pronouns trigger familiarity presuppositions

- definite descriptions and pronouns contribute variables to logical form
- assignments inherited from the incoming context can be carried over to these variables (dynamic semantics)
- but under what conditions?
- the descriptive features of definites “impose constraints which help the addressee to identify the familiar discourse referent to which the speaker intends to refer” (Roberts 2004)
- these variables must be identified with (discourse) referents already available in, i.e. ‘familiar from’, the incoming context

Heim (1988), Roberts (2004)

The donkey

The donkey kicks the farmer.

- $\exists x(\text{donkey}(x) \wedge \forall y(\text{donkey}(y) \rightarrow y = x))?$
- a condition that must be *satisfied* by the incoming context; i.e. the incoming context must contain an accessible discourse referent z and condition 'donkey(z)'

Presupposition as anaphora

van der Sandt (1992)

- satisfaction replaced with binding
- general mechanism: all presuppositions analyzed as anaphors; presupposition projection modelled as anaphora resolution

Extension to the extra-linguistic context

Can we extend the mechanism of presupposition to deictically used expressions?

Can we model character as a kind of presupposition?

Character as a presupposition

Character: a semantic constraint on the incoming context

- character guides an interlocutor to the entity under discussion
- a defective context leads to semantic anomaly: a possibility for demonstratives at least
- the particular constraints imposed by a particular indexical follow from its meaning: *I* picks out the agent of the most recent utterance

Treating character as a particular type of familiarity presupposition in a dynamic account echoes Kaplan's treatment of character and context as something *prior* to Kaplanian content.

- also captures the fact that the descriptive (presupposed) meaning of an indexical does not show up in truth-conditional (asserted) content.

Zeevat (1999), Maier (2006), Hunter & Asher (2005), Hunter (2010);
cf. Geurts (1997)

A presuppositional account

Basic idea:

- the presupposition of, e.g., *I*, is always the same
- different tokens will bind to different discourse referents

But a presuppositional account requires a few amendments

- discourse referents for features of utterance events
- a distinction between utterance and linguistic contexts that nonetheless supports a single presuppositional mechanism
- an account of rigidity/reference (anchors)

Referents and contexts

An utterance event can be represented by a discourse referent e_U ; as can its features (agent, time, etc.)

Contexts: build on DRT's notion of structured discourse representations and add a super-global layer for information from the utterance context

- discourse is situated in the extra-linguistic environment
- captures asymmetry of discourse effects

Hunter (2010), (2013)

The semantics of indexicals

I has a presuppositional component:

- p : $\exists x \exists e_u (\text{agent}(x, e_u))$; a : $\lambda P.P(x)$

Shouldn't bind to the agent of just any utterance event, however

- typing of discourse referents

Not a general enough approach for expressions that crossover

- *now*, for example
- sometimes denotes utterance time; sometimes denotes another time
- other deictics, etc.

Hunter & Asher (2005), Hunter (2010)

Strategies

I adopts a particular strategy for finding an appropriate antecedent in a structured discourse context (representation)

- p : $\uparrow \exists x \exists e(\text{agent}(x, e))$; a : $\lambda P.P(x)$
- \uparrow forces binding at highest context

now has a similar, but slightly different strategy:

- $\uparrow \exists t \exists e(\text{time}(t, e))$
- \uparrow allows binding in lower contexts when binding at highest level is blocked and certain other conditions are met

Definites differ with regard to the strategies they adopt for finding an appropriate antecedent in a structured discourse context (representation)

- extra-linguistic/linguistic readings modelled not by the type of entity that an expression seeks, but about how the expression seeks it.

Hunter & Asher (2005); cf. van der Sandt (1992)

Rigidity and anchors

Put the burden of rigidity on the interpretation of discourse structures, not on the semantics of different expressions

- individuate utterance events by their spatio-temporal coordinates and add spatio-temporal conditions to the DRSs
- evaluate a DRS starting at the top level and carry assignments down (as usual)
- evaluate extra-linguistic layer relative to a single world taken as actual and an empty assignment function

Hunter (2013)

Loose ends

A lot left to do: how do these strategies work for different expressions and how should contexts be structured? But note,

- if pure indexicals functioned like Kaplan posited, then a dynamic, presuppositional model would be straightforward.
- what is difficult is not modelling character as a kind of presupposition, but modelling the actual behavior of indexicals, which is far more nuanced than Kaplan's picture suggests.

a presuppositional account that emphasizes structured contexts and different strategies gives us room to explore these questions

We've argued that at least deixis is better understood by generalizing a theory of intra-linguistic interactions. Can we stop there?

Problems with the intralinguistic account of dynamic semantics

We've looked at tense and temporal anaphora in a dynamic framework

Problem: tense features and adverbials alone do not determine temporal relations between events.

Example

- (i) John fell. Sam helped him up.
- (ii) John fell. Sam pushed him, (Lascares & Asher 1993)

Problems with the account of nominal anaphora

Problem 1: for reference to text contents, dynamic semantics alone has no account of what the appropriate propositional contents are (Asher 1993).

Problem 2: when more than one familiar discourse referent satisfies the conditions of a nominal anaphor, anaphoric reference depends on relations between sentences (Kehler et al 2008) that dynamic semantics does not have room for.

Example

- (i) John teased Sam. Pat tickled him as a result.
- (ii) John teased Sam. And (in addition) Pat poked him.

Coherence in discourse

What is it?

- intuition that some texts have an organization and that each clause plays a certain role in the whole
- if we think of a text being read, we can think of the reader assigning each of its clauses a temporal like moment, with text having come before and text coming after.
- A coherent text is one in which there are semantically important connections between what is said at any given moment in the text, what was said before and what will be said.

Example

Sam fell, because Pat pushed him. Pat got in trouble, as a result.

More on connections

Connections can be explicitly marked by *discourse connectives*—individual words or phrases, e.g., *as a result*—but need not be.

Example

Sam fell. Pat pushed him. Pat got in trouble

- The Penn Discourse Tree Bank, which assumes coherent text, has over 50% of the connections not marked.
- discourse connectives exist in all languages that I know of, though certainly not all semantic relations are important.
- EU Cost action on discourse coherence and its markers has documented discourse connectives for most European and some non-European languages.

Semantically important connections

Various kinds of relations

- causal relations (Result, Explanation), logical relations (Consequence, Defeasible Consequence), temporal relations (Narration, Flashback), relations of expanding on or summarizing content (Elaboration, Summary), comments on what was said or what happens in the extra-linguistic context.
- clarification questions, follow-up questions, question-answerhood,
- structural relations contrast, parallelism;
- other more genre specific structural relations—e.g. the structure of an academic psychology article or a computational linguistics paper in a major conference (RST's Topic-Organization).

Generalizations and theoretical differences

- different theories carve up the space of semantically important relations slightly differently, but large consensus at higher level
- theorists from major different frameworks (DLTAG, RST, SDRT) now largely agree on a general taxonomy, though the details are still a matter of active research (cf. EU Cost action)
- and discourse parsing experiments (automatically recovering discourse relations and a discourse structure from text, now an active field of research in computational linguistics) can move from one set of relations to another to compute discourse structures.

Relations and their semantic effects

- these relations have been documented to affect interpretation: psycholinguistic effects (cf. Sanders 1997), anaphora resolution (Fox 1987, Kehler et al. 2008), ellipsis resolution (Asher 1993, Kehler, Hardt), and temporal information (Lascarides & Asher 1993)
- for many of the relations mentioned, the semantic effects are straightforward, except perhaps for the structural relations. But...
- parallelism and contrast are often crucial for resolving ellipses (Asher 1993, but see also Dalrymple et al. 1991).

How to make this explicit?

Discourse coherence involves semantically informative relations between portions of discourse contents. How do we model the connections?

- more abstractly, a coherent text can be represented as a graph (V, E, ℓ) with V the set of discourse units, E the set of edges representing connections between units, and ℓ a function that gives the type of each edge (what relation it is).
- But what are discourse units?
- What are the types?
- Are there any general constraints on these graphs for coherent texts?
- What structure does an incoherent text give rise to?

The question of V , the set of discourse units

- Discourse units are “instances” of content (see Kaplan) relativized to the context (text+cotext); can also think of them as speech acts.
- The same content can have different, even incompatible semantic relations to other units (in different contexts)
- But how finely should we individuate discourse units? Sentences, clauses...?
- To some extent this is an empirical question to be resolved by the difficulties of annotation campaigns.

The question of V , continued

Embedded units:

Example

That man, [who has incited fear and hatred,] will never be President.

- the non-restrictive relative clause has a discourse function with respect to the main clause—Explanation or Background
- if we take such clauses to give rise to discourse units, we will have *discontinuous* units (the main clause).

The question of V , continued

Complex discourse units:

Example

(i) A: Will you give me a sheep?

(ii) B: For a wheat

(iii) B: oh, sorry, an ore.

(iii) is a correction of (ii); together they form a complete response to (i)

- (ii) and (iii) as a mini-discourse structure form a **complex discourse unit** that plays a rhetorical role of answering the question in (i).

The question of structure

- A fully coherent text is one where each unit bears some relation to some other constituent. Thus, coherent graphs are weakly connected.
- coherence then has a scalar value, depending on the degree of the approximation of the weak connectedness constraint.
- acyclicity: we don't have rhetorical functions where we have for $a, b \in DU, a \longrightarrow b$ and $b \longrightarrow a$.
- beyond that, theories differ: should the structure be a tree, a directed acyclic graph (DAG), or a set of DAGs?

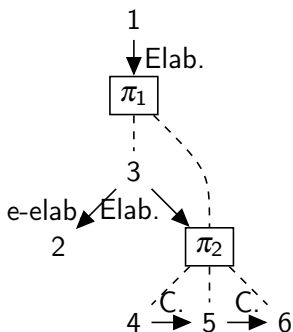
Some options for representing coherent structures

- Penn Discourse TreeBank (PDTB) uses directed edges, without structural constraints
- Rhetorical Structure Theory (RST) represents discourse as trees, which is too restrictive
- Segmented Discourse Representation Theory (SDRT) uses two-layered directed acyclic graphs

We have shown that SDRS structures can embed simpler structures like dependency trees or RST trees but SDRSs are strictly more expressive (Venant et al. 2013).

SDRT's DAGs with two sorts of arcs

[The principles of natural selection.]_1 [The theory of natural selection [as it was initially described by Charles Darwin,]_2 rests on three principles:]_3 [1. the principle of variation]_4 [2. the principle of adaptation]_5 [3. the principle of heredity]_6



The dotted lines indicate a special kind of link of parthood between CDUs and their constituent DUs.

Back to the extra-linguistic context

Interim recap:

- we've argued that presupposition can serve as a homogeneous underlying mechanism for expressions whose interpretations can come from either the extra-linguistic context or the linguistic context.
- we've pushed for a dynamic analysis of presupposition
- finally, we've argued that a standard dynamic account of intralinguistic dependence must be supplemented with rhetorical relations and structure.

What do rhetorical relations and structure have to do with the extra-linguistic context?

Definites and coherence

To the extent that we want a unified account of extra-linguistic and intralinguistic context dependence, we need an understanding of how extra-linguistic contexts and linguistic contexts fit together.

An example with *now*

The president's request that Mr. Kaine run for a senate seat caused commotion among Mr. Kaine's advisors.

- (a) When he agreed to run the DNC in 2009—even while finishing his last year as governor—his closest advisors were stunned and counselled him to renege.
- (b) **Now** Mr. Kaine was facing an unwanted repeat of the same, uncomfortable situation.

(a) provides background on the current situation; (b) returns the discourse to the “present” time or “now” of the discourse.

- interpretation of *now* is sensitive to discourse structure
- neither the time inherited from previous eventive predicate (despite modifying a stative predicate)
- temporal dynamics insufficient

- 1 If coherence can guide resolution to the right linguistic antecedent, does it also guide the choice between a linguistic and extra-linguistic antecedent?
- 2 Is coherence relevant to other aspects of extra-linguistic/linguistic interactions?

A more drastic modification to the Sellarsian view than one that involves only deixis

- 1 If coherence can guide resolution to the right linguistic antecedent, does it also guide the choice between a linguistic and extra-linguistic antecedent?

Demonstrations can override the salience of an entity mentioned in the discourse, but what happens when there is no demonstration?

Color balance



Demonstratives without demonstrations. No overt action like in the obvious cases of demonstratives and in the corpus work on demonstratives. No demonstration to tell you what the word connects to.

- the main contribution is to show that the semantics of a demonstrative is much more like the semantics of an indexical than one might assume.
- to do this they argue for a salience ordering of individuals in an EL situation or scene, where that salience ordering is determined at least to some extent by the coherence relations that the scene bears to linguistic moves.
- operationalizing this idea requires a rich ontology of situations (cf Kratzer, Barwise and Perry, Barwise and Etchemendy).

More on Stojnic et al.

- contexts are partially ordered sequences of entities, ordered according to prominence. The list seems to include both linguistic and EL entities.
- an utterance can change the ordering. *There's your omelette* should make the omelette in the scene most salient.
- the way this happens is by placing a particular coherence relation between the linguistic move and the scene it is describing, which they call *Summary*.
- Summary holds of a linguistic move ℓ and a scene s iff ℓ offers a good and correct answer to the question, "What is happening in s ?"

A worked example

Suppose Julia Child had said at the relevant point

Example

That's an omelette.

S, S & L's logical form:

$\langle \alpha 0 \rangle$; [*Summary*(s_0, e_0)]; $\langle \sigma 1 s_0 \rangle$; $\langle \pi 1 @that \rangle$; [*omelette*(e_0, x_1)]

- $\langle \alpha 0 \rangle$ places e_0 at the top of the stack after the utterance (state of being an omelette); [ϕ] is a test on e_0 .
- $\langle \sigma 1 s_0 \rangle$ places the most salient individual in scene s_0 next on the stack;
- $\langle \pi 1 @that \rangle$ forces the referent of *that* to position 1 (the real omelette).
- after the utterance: $\langle e_0, \text{the real omelette, other features of } s_0 \rangle$.

- What is the discourse relation doing? Is it just a test?
- SS&L say: The effect of Summary is that the central entity of a situation the summary is about is rendered most prominent.
- So we might think of Summary as entailing $\langle \sigma 1 s_0 \rangle$.

Assessment of S,S &L

- a simplified theory of discourse structure and a definition of dynamic actions over attentional states
- provides an account of how deixis can work even without demonstrations and how demonstrata are determined via linguistic constraints.
- on the other hand, there's no room for EL eventualities to play any role other than as an argument to the relations Summary and Assessment.
- the account is restricted to examining effects on a salience ordering for the purposes of determining referents for demonstratives.

Lascarides and Stone

- accounting for coverbal gesture as establishing deixis to events or objects in a real or imagined space as well as iconic action-signs that may themselves tell a story.
- interactions with scope bearing elements.

Example

And um I thought not too edgy and like a box, more kind of hand-held more um . . . not as uh computery [typing gesture] and organic, yeah, more organic shape I think.

- the overall message is not computery with a keyboard.
- Distinct but related content to the lexical item.
- this requires a theory in which the action falls under the scope of a linguistically given element.
- they also exploit a theory of underspecified meaning needed because gestures are iconically and deictically typically underspecified in meaning.

Another telling example from S & L

Example

You walk out the doors.

[The gesture is one with a flat hand shape and vertical palm, with the fingers pointing right, and palm facing outward.]

S & L: The linguistic component expresses an instruction. And intuitively, the interpretation of the gesture is also an instruction: “and then immediately turn right”.

More on Lascarides and Stone

- if gesture is just redundant or emphatic, a theory of unification coupling the targeted linguistic meaning and the action's meaning might suffice.
- But S & L argue that what is really important is the coherence of the linguistic message with the co-verbal gesture.
- They also argue that a variety of discourse relations can link gesture and linguistic message.

Particularities of coherent gestures

S & L add certain discourse relations to deal with particularities of gesture:

- *Depiction* linking redundant, iconic meaning with meaning that's linguistically expressed.
- *Overlay* linking gestures developing the same virtual space.

Example

Norris is like up here [The right arm is extended directly forward from the shoulder with forearm slightly raised; the right palm is flat and faces up and to the left]

And then the library is over here. [After returning the right hand to rest, the right hand is re-extended now to the extreme upper right periphery, with palm held left.]

- *Replication* reusing parts of a virtual space to refer to an antecedently, gesturally introduced object or eventuality. (cf. uses of *loci* in sign languages).

Formalizing the spatial parameters

Needed to handle deixis and spatial information about objects conveyed gesturally.

- distinguish between using the virtual space to convey information about some non deictically given spatial situation and deixis to an object in the immediate EL environment by using maps from the virtual space to the spatial situation.
- the identity map is used to give deixis
- e.g.: the location given by the pointing is identical to the one that the predication should apply to) whereas for Norris eg. the map from the physical space pointed to by the agent is not identity, because Norris Hall isn't in front of the speaker.
- constraints on maps μ that are not identity: no mirroring (—e.g., $\text{Right}(x, y)$ and $\text{Left}(\mu(x), \mu(y))$ is not allowed).

Formalizing the spatial parameters, continued

- a predicate for locating objects in a space.
- a predicate for classifying or relating a linguistic element to a location that stands for some object.

Example

We have this one ball, as you said, Susan. [The speaker sits leaning forward, with the right hand elbow resting on his knee and the right hand held straight ahead, in a loose ASL-L gesture (thumb and index finger extended, other fingers curled) pointing at Susan].

Some key assumptions

- gestures do not introduce new entities but always refer back to entities introduced linguistically or deictically present.
- So no coverbal gesture to introduce *a man* as opposed to an addressee *was walking in the woods*.
- use Bittner's foreground/background distinction amongst discourse referents; entities introduced only gesturally (backgrounded) can only be referred to gesturally, not with a pronoun.
- Contrast this with ASL *two Canadians were walking in the woods*, where there is no verbal accompaniment.
- gestured entities thus are predicted not to be anaphoric antecedents for simple pronouns or simple Corrections (No,...)

Some complications

“Filling in” gestures are different

Example

Kate was (like) [depicts snoring] during the lecture. It was so embarrassing.

And even some coverbal gestures seem to introduce entities that are prominent enough to enable coreference and correction.

Example

- (i) Kate and Julie walked towards each other[right and left hand making wiggly movements]. No, they walked straight towards each other.
- (ii) Kate and Julie walked towards each other[right and left hand straight towards each other]. #No, they meandered towards each other.

Deixis treated

- These things push up the pins. [The speaker points closely at the frontmost wedge of the line of jagged wedges that runs along the top of a key as it enters the cylinder of a lock.]
- $\pi_0 \exists s \exists p (things(s) \wedge pins(p) \wedge pushup(e, s, p)) \wedge [G] \exists w (exemplifies(w, s) \wedge loc(e, w, v_l(\vec{p}_w)))$

One discourse constituent that combines both verbal and co-gestural content.

Example

We have this one ball, as you said, Susan.

[The speaker sits leaning forward, with the right hand elbow resting on his knee and the right hand held straight ahead, thumb and index finger extended, other fingers curled pointing at his addressee]

- $\pi_1 : \exists w \exists b (we(w) \wedge have(e, w, b) \wedge one(b) \wedge ball(b))$
- $\pi_2 : \exists s \exists u (susan(s) \wedge said(e_0, u, s))$
- $\pi_3 : [G] \exists e'' classify(e'', u, v_m(\vec{p}_i))$
- $\pi : Depiction(\pi_2, \pi_3)$
- $\pi_0 : Elaboration(\pi_1, \pi)$

How do these coherence accounts differ from intention based accounts?

- There is no overt appeal to intentions.
- inferences to discourse structure are performed on the basis of cues
- these cues are features that automatic extractions of discourse structure using machine learning methods can exploit.

Some more details

- as communicating agents we are attuned to and learn quickly regularities about signals and what they signal.
- these signals may be words or expressions, but also actions or gestures.
- contextual features are also important
- from annotated corpora, we can induce to some extent such learned behavior in computers through machine learning algorithms.
- but so far we have only used rather superficial features,
- and everybody believes that to understand meanings of signals in context we need deeper features about lexical content
- it's just that nobody knows how to code that semantic information or rather extract it from a given context algorithmically.

The theoretical commitments of a rhetorical account

- If an agent uses a signal with a certain meaning, she makes a certain kind of public commitment that involves that meaning.
- E.g. In our example the gesture towards the drawing coupled with the preceding linguistic move *she was sent to her room* publicly commits the agent to giving a reason why her daughter was sent to her room.
- this commitment is something bystanders infer, and it can be defeasible.
- But the inference crucially does not depend upon whether the agent intended this consequence.
- It depends on regularities governing signal use and meaning.

The structure and semantics of gestures as developed so far

- gestures are encoded as a feature structure involving hand shapes, motion, and location (see also Kopp, Tesser and Cassell 2004)
- Each attribute–value element specifying a feature yields an underspecified representation for a predication, which must be resolved to a formula in the LF of gesture in context
- appeal to a hierarchy of possible specifications for each attribute value pair.
- these formulas can now enter into scope relations with operators like negation (cf. the ‘computery’ eg).

Where coherence comes in

- SDRT's glue logic can help in resolving underspecifications in gestural meaning.
- e.g. by inferring an Elaboration relation between π and π_1 above
- and then using the grammatical constraint that π_3 and π_2 have to be attached together since the stroke of π_3 is contemporaneous with the utterance of π_2 .

Why exactly is the content of gesture underspecified?

We don't have an answer to this question in this work.

Gesture can sometimes specify linguistic content

Example

Someone [speaker points to a particular person] was really embarrassing last night. He...

- Here the deixis accompanies an indefinite providing a witness.
- Though *someone* is typically thought of as introducing a novel discourse entity whose actual reference is not specified.
- gesture specifies the meaning of the indefinite as a specific indefinite.

Assessment of S&L

- EL eventualities enter into many more coherence relations here than in SS&L.
- a general analysis of gestural space and constrained mappings between gestural space and other domains.
- in S&L language disambiguates underspecified gestural meaning, but does not affect non gestured eventualities or scenes.
- could use SS&L's account of demonstratives to extend S&L's account of deixis (no longer constrained to involve a gesture)
- moving beyond co-verbal gesture to a more general account of EL eventualities in discourse.

More assessments

- Gestures have a simple nature. They are simple motion events, and so they simply concatenate together to form larger events via *Overlay*
- That's probably not right for EL events in general, and one might wonder whether that's sufficient for general iconic signs.
- and this view leads to a perhaps overly narrow view of how language can affect them—specifying underspecified meanings.
- As SS& L show, language can also manipulate (albeit still in a limited way) how the extra-linguistic environment is conceptualized, by making objects or eventualities in the scene more or less salient.

More generally, what is coverbal gesture doing?

Example

Do you want to switch places? [pointing successively to addressee and then to speaker]

- does the gesture “complete” the meaning of *Do you want to switch places?*
- or is it the EL situation (just 1 addressee as opposed to 2) that specifies who the addressee is supposed to switch places with?

Another example

Example

You walk out the doors.

[The gesture is one with a flat hand shape and vertical palm, with the fingers pointing right, and palm facing outward.]

- S & L: The linguistic component expresses an instruction. And intuitively, the interpretation of the gesture is also an instruction: “and then immediately turn right”.
- But does the gesture on its own really give us this information?
- consider two EL contexts: one where upon exiting the doors you can either go right or left, and one where there are many other route possibilities.

Coverbal gesture as a limited phenomenon

- one has the sneaking feeling that coverbal gesture isn't very autonomous from S&L's examples
- confined to the role of emphasis (maybe telling us what's salient in the linguistic message)
- non-coverbal gestures that complete linguistically incomplete expressions as in Davidson seem to give us a richer theory of gestural meaning, closer to that observed for sign languages.
- in particular predicates that are gesturally given— *she was like* [totally zonked out snoring gesture]— would seem to have to compose in the right way and have the right type. and more familiar than than Overlay or Depict.
- and they would have more content.

More assessment: why specifying iconic meaning seems like a hard semantic problem

- While feature structures currently used for co-verbal gesture might describe many (all?) co-verbal gestures, not clear that they do so for complex iconic signs.
- they seem to involve only a characterization of a pre-position, movement and post-position and each value for an attribute has a finite number of possibilities, while gestures are continuous paths in \mathbb{R}^4 .
- Given the amount of information loss, it's not clear that the features capture the semantically relevant features of the gestures.

Is iconic meaning different from linguistic meaning?

- At an abstract enough level, an incomplete iconic sign has the meaning of all its possible completions.
- This is analogous to linguistic meaning of subsentential constituents, think of a noun phrase as having a meaning of type $(E \rightarrow \text{PROP}) \rightarrow \text{PROP}$
- However, composition of linguistic meaning exploits predetermined syntactic categories that correspond to a rich system of semantic types
- This largely determines how composition should work by specifying how all the types have to combine to yield a propositional content (truth conditions).

Structured composition in gesture?

- it's not clear that there is any analogue to syntactic structure for gestures outside of Sign Languages.
- And the basic elements as described by feature structures don't seem to have any intrinsic meaning in and of themselves. (any of the strokes might just describe a nervous tick as well)
- Perhaps as S & L postulate, there's just a concatenation operation that eliminates possible continuations at each step...
- think of a space of all finite sequences in some finite vocabulary V ; any finite prefix $a \in V^*$ has a set of continuations $a.V^\infty$; if a is a subsequence of b then $b.V^* \subsetneq a.V^*$.

Settling on a particular gestural meaning

- a gesture determines a path in \mathbb{R}^4 and in general so is its target semantics, which we could take to be an eventuality type or set of eventualities
- To determine the meaning of an iconic sign, we must adjudicate between maps from the gesture to another path type or set of paths
- The problem of gestural meaning: What is the target event type (path) picked out by the gesture?

Why this problem seems hard to us

- The problem in general is unsolvable
- the general problem is to optimize a fit between pairs among the set of all maps $\mu: \mathbb{R}^4 \rightarrow \mathbb{R}^4$, which has cardinality $\aleph_1^{\aleph_1}$ or \aleph_2
- As in robotics, a semantics for iconic signs should work in an approximation of the full space,
- so that the gesture as a path over a finite array of pixels or discriminable regions and the target domain's representation is also a path over a finite array of pixels.
- even so...

What is a sign with an iconic meaning?

First Attempt:

- Let $\mathbb{P}(\alpha)$ be the the “natural” or “salient” part whole structure of α (similarly for $\mathbb{P}(\|\alpha\|)$). Then α is an iconic expression iff $\exists f : \mathbb{P}(\alpha) \rightarrow \mathbb{P}(\|\alpha\|)$, such that $\forall x \in \mathbb{P}(\alpha) \exists y \in \mathbb{P}(\|\alpha\|)$ such that $\|x\| = y$.
- potential problem: while in our examples of iconic expressions α it's intuitively clear what the natural or salient part whole structure of α and $\|\alpha\|$ are, it is difficult to say precisely what this is.
- potential problem: simply mapping parts of α to parts of $\|\alpha\|$ is not enough; the parts must stand in the right relations to each other; sequence is one such important relation, spatial relations are another, part whole structure is another.

Second attempt

- Non-linguistic objects, events and iconic expressions e can be represented as structures, $S_e = (X, R_1, \dots, R_n)$, with X a non-empty set of parts of e , and R_i the relevant relations for giving the structure of e . (a generalization of a feature structure)
- Then α is an iconic expression iff $\exists f : S_\alpha \rightarrow S_{\|\alpha\|}$, such that $\forall x \in X_{S_\alpha} \exists y \in S_{\|\alpha\|}$ such that $\|x\| = y$. and $f(R_i^{S_\alpha}) = R_i^{S_{\|\alpha\|}}$
- the nature and structure of $\|\alpha\|$ constrains the nature of α .
- question: is there a universal set of constitutive relations for S_e for all e ? Or do they depend on the type of e ?

Alternatives and accompanying linguistic content

- Optimize the fit by looking at an edit distance over maps from the gesture to a set of alternative domains.
- Perhaps gestural meaning requires linguistic input or at least some sort of top down cognitive expectations in order to fix the set of relevant alternatives

Continuing our speculative semantics

How would this work e.g. for the computery example from S & L?

- *computery* evokes a set of alternative specifications of computer components used in a certain way.
- the gesture maps best to the event type of using a keyboard given the alternatives (with a different alternative set, the gesture might have picked out a keyboard instrument).
- there are still a lot of holes to fill in here, conceptually and mathematically...

Recap of this morning

- the mechanism of presupposition can be used to provide a unified model of extra-linguistic and linguistic context sensitivity, at least for deictics and definites
- but a dynamic theory of presupposition/anaphora needs to be supplemented with rhetorical relations and structure
- rhetorical relations and structure are needed not only to handle intralinguistic interactions but also extra-linguistic/linguistic interactions

Looking ahead to this afternoon

A presentation of our *Settlers of Catan* corpus and its role in developing a more general account of extra-linguistic/linguistic interactions

Demonstratives

Demonstratives never fit comfortably within Kaplan's picture

- final proposal: speaker intentions

Inferences about denotations

How do we model this inference?

- recognition of speaker intentions?
- it would seem so if what is demonstrated is a matter of what the speaker intended to refer to.

Grice on communication

Meaning is a matter of intention.

Suppose you observe someone doing a demonstration.

You want to know what she wanted to convey what she meant by that, what was her intention.

Understanding a communicative act involves an inference to the communicative intention behind the act.

How the story would go

- In this particular case, you need to recover the author's communicative intention
- what did she intend to refer to with that demonstration?
- you do so on the basis of what the author or more generally people would normally intend by making such a demonstration in such a situation.
- you then infer an intention on the part of the gesture's author to communicate a particular meaning.
- and from this intention you infer the meaning of the gesture.

Why this might look appealing

Reference does seem to be a relation that requires a mind somehow.
A signal refers or means something in virtue of its being used by intelligent agents

Mind reading isn't necessary

- Inferring intentions is difficult, imperfect evidence about hidden states
- possibilities of deception, even self-deception.
- Why exactly do we need an intermediate conclusion about intentions in this reasoning?

Mind reading isn't necessary

Why isn't the explanation just involve a generalization of the sort:

- Making such a gesture in such circumstances normally implies that the author commits to some meaning about the most salient object (with respect to our expectations) in the general direction of the glance.

A fallacious inference?

- One can grant: Yes reference and meaning involve mental states,
- But it doesn't follow that inferring which particular reference was intended by the action necessarily a reconstruction of the speaker's intentions.
- the inference to a speaker's intention doesn't add anything in terms of predictive adequacy.
- rules about what constitute normal referring behavior are sufficient.

The explanation in terms of intentions seems backwards

- Rather than intentions serving a crucial role in inferring what a particular gesture meant,
- it seems that we infer an intention on the part of the gesture's author based on her referring behavior in the particular situation,
- what the gesture in fact means (or what it normally means).

Intentional explanations make the wrong predictions

- intentions are private mental states. An observer of a communicative act doesn't have access to those intentions.
- she has to infer the intentions, which she does on the basis of her beliefs about the person and the situation.
- in a case in which the observer misses the communicative intention, she should revise her beliefs.
- when we're not sure of what the rules are, this might indeed happen. But in most normal cases this doesn't happen.

Intentional explanations predict the wrong things

- Suppose the object the person in our example wanted to indicate lay in a direction orthogonal to the direction of her pointing.
- The observer does not revise her beliefs about what objects get picked out by the glance (a plain, immaculate wall);
- she concludes the author of the act was acting completely bizarrely and isn't making sense.
- The observer's beliefs about rules about normal behavior or even as applied to the referring behavior in this instance don't get revised.

The insufficiency of intentions

You can intend to refer to something, but if you use the wrong signals, you will fail.

- You suppose you want to refer to a person in the room whose name you think is 'Kate'. You're mistaken though; the person you want to refer to is named 'Julie'.
- You say *Kate is having a glass of wine* to your conversational partner
- Kate overhearing you says: *that's not Kate. I am, and I'm not having a glass of wine.*
- You communicated something false even if you didn't intend to.
- your intentions to communicate are different from what you actually communicated.

The same moral holds for demonstratives

As SS & L say: For example, even if you intend Sue to be the referent of your use of 'she', if she is not a prominent candidate referent, you will fail to pick her out. For example, if you are pointing at Ann, then the referent of 'she' is Ann, even if you intend it to refer to Sue.

Upshot: to the extent that a speaker expects her interlocutor to recognize her intended meaning, she must use cues normally associated with these intentions to get that intention across. But then, that combination of cues in the same context, will always yield the same result regardless of whether the speaker actually has those intentions or not. So let's just study the cues; no mention of intentions needed.