Integrating multimodality into learning in dialogue

Jonathan Ginzburg
Université Paris-Diderot
yonatan.ginzburg@univ-paris-diderot.fr
Aix-en-Provence, May 24, 2011

Some issues in Language Learning

- Perennial question: inductive *or* interactive?
- Most grammar induction work ignores conversational interaction (Clark & Lappin, 2010).

Some desiderata from a model of Language Learning

- Carers use normal language to interact with novices.
- Minimize explicit teaching (?? lessons from the swimming pool)
- Learning relies on little data
  - cf. most existing machine learning approaches (e.g. reinforcement learning Henderson, Lemon, & Georgila, 2008)
  - One shot learning (Fei Fei 2006)
  - Active learning (Chao 2010)

A suggestive answer: Macura's simulation

- Macura's simulation (Macura, 2007; Macura & Ginzburg, 2006, 2008)
- The model contains an ALife environment in which the foraging efficiency and lexicon dynamics of populations that possess and lack MCI capabilities are compared.
- The environment contains:
  - *plants*: represent different meanings
  - *agents*: distributed randomly in the environment
- Agents forage for food and when proximate to one another engage in a brief conversational interaction concerning plants that are visible to them.

The Agents

Agent Properties

- An agent has the following parameters:
  - *vision*
  - *adulthood age*
  - *max age*
  - *vitality*
  - *hunger*
  - *memory*: location and plant type of last plant consumed
  - *private lexicon*: an association matrix which stores the association scores for every plant-word pair based on past experiences:

	m_1	$\dots$	m_N
w_1	S_{11}	$\dots$	S_{1N}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
w_M	S_{M1}	$\dots$	S_{MN}

The Agents

Agent Types

Three types of communicative agents:

- CR: capable of asking clarification questions,
- Introspective: lacking the capability of asking clarification questions,
- Hybrid: capable of both introspection and CR.

Communication Protocol

1 agent A_1 sees agent A_2, where $A_1 \longrightarrow$ *speaker* and $A_2 \longrightarrow$ *hearer*
2 A_1 looks around for plants in vision
   - **if** plants in vision a plant P_1 that is picked as topic of conversation depends on A_1's state of hunger ▸ Choose topic
   - **else if** there are no plants in vision, exit the dialogue if not hungry, otherwise ask for food location ▸ Ask for food location
3 A_1 chooses word w_i with the highest association score for the topic P_1
   - **if** no words associated with plant P_1 exit dialogue
   - **else if** two or more words with highest score chose one word at random
4 A_1 sends w_i to A_2

- The way that A_2 grounds w_i depends on A_2 type.

Communication Protocol

- Introspective strategy:
  - **if** no plants in vision $\rightarrow$ exit dialogue, **else**
  - A_2 hears w_i and for every plant in vision increases the association score of $(P_n \rightarrow w_i)$ by 1
- CR strategy:
  - **if** no plants in vision or w_i not in lexicon or mismatch between internal state and context $\rightarrow$ ask a CR "w_i?"
  - **else** if the perceived plant is in context, increase score by 1

Communication Protocol
Lexicon Update

- Only the hearing agent A_2 updates her lexicon after a conversational interaction.
- After the update, A_2 chooses the plant P_{perc} with the highest association score for the word heard w_i.
- If A_2 can see $P_{perc} \longrightarrow A_2$ steps towards P_{perc}.
- If the perceived plant P_{perc} matches with the speaker's intended plant $\longrightarrow$ communication successful.
- Neither agent given any feedback on this outcome.
- No lateral inhibition of other competing associations.

Generational Turnover

- Implementation of *natality* and *mortality*.
- For every agent that dies a new infant agent is born to an adult agent with the highest vitality that is not currently a parent to another infant.
- Infants have an empty lexicon and inherit their parent's type.
- Infants follow their parents and listen to their dialogues learning only by introspection.
- Upon reaching the *adulthood age* the infant stops following the parent and is able to communicate with other agents and have children.

Macura's simulation: main result

Multi-generational model

- The lexicon diverges at a faster rate for an introspective population, eventually collapsing to one single form which is associated with all meanings.
- This contrasts sharply with MCI capable populations in which a lexicon is maintained.

The Rest of Today's Talk

- Semantics and Dialogue in KoS
- Some missing ingredients for a model of language learning
- Incorporating multimodality
- Two simple examples

Where we are now

Interaction *in* the grammar?

- Grammar as pertaining to disembodied, decontextualized combinatorial system (Chomsky, Hauser, Chomsky, & Fitch, 2002)
- Grammar as an adaptation driven by communication.(Pinker and Bloom, 1990, Pinker and Jackendoff, 2005)
- Interaction is intrinsically built into grammar (Ginzburg, *The Interactive Stance*, Oxford University Press, 2011.)
- The meaning of words or constructions involves notions that irreducibly involve notions of interaction.

Interaction *in* the grammar: evidence from non-sentential utterances

- Evidence from NSUs:

(1) a. A: *Yes.*; meaning of ‘yes’: *p*, where *p*? is the current issue under discussion.

b. A:*Bye.*; meaning of ‘bye’: A seeks to disengage from a conversation with B which has involved at least some discussion.

c. A: *mmh.*; meaning of ‘mmh’: A acknowledges understanding of B’s latest utterance.

d. B: Did Jo leave? A: *Jo?*; intended content meaning of reprise fragment ‘u?’: A asks B what is the intended reference of B’s (sub-utterance) u under condition of phonological segmental identity.

e. B: Did Jo leave? A: *Why?*; meaning of metacommunicative ‘Why?’: A asks B of the cause of an utterance by B, an utterance the issue which it raises remains under discussion.

Dialogue Oriented Constructionism

- *Dialogue Oriented Constructionism* (DOC) (see e.g. Ginzburg & Sag, 2000; Ginzburg & Cooper, 2004; Schlangen, 2003; Purver, 2004; Fernández, 2006; Ginzburg, Fernández, & Schlangen, 2011), combines a view of grammar inspired by developments in construction grammar and HPSG (see e.g. Fillmore & Kay, 1999; Sag, 1997) and the modelling of dialogue context in the KoS framework (Ginzburg, 1996; Larsson, 2002; Cooper, 2005; Ginzburg, 2011).

Type Theory with Records as logical framework

- Type Theory with Records (TTR) (Cooper, 2005, 2011) maintains the insights of situation semantics—with simpler theory (no non-well-founded set theory and modulo the Liar)—but gain from the extra structure.
- Use Type Theory with Records to build the semantic ontology and to write grammatical and conversational rules.
- Type Theory with Records: a framework that allows
  —Dynamic semantic techniques à la DRT
  —Constraint-based Grammar à la HPSG
  —Formalization of Semantic Frame Theory—ability to deal with vagueness of word meaning.

Type Theory with Records as logical framework

- TTR notationally similar to Type Feature Structures, but substantively different
- TTR contains λ-calculus: crucial for doing semantics
- Types and tokens both first class entities in the ontology
- Crucial for uniform theory of illocutionary and metacommunicative interaction.
- Contrasts with both Typed Feature Structures used for CBGs and Discourse Representation Theory used for formal semantics

Fundamental notion

- The most fundamental notion of TTR is the typing *judgement* $a : T$ classifying an object a as being of type T.

(2) a. b : IND

b. t : TIME

c. $s : run(arg1_{IND} : b, arg2_{TIME} : t)$

d. s : run(b,t)

Records and Record Types

- The record $\begin{bmatrix} \text{runner} = \text{bo} \\ \text{time} = \text{2pm, Dec 20} \\ \text{place} = \text{batumi} \end{bmatrix}$ is of the type $\begin{bmatrix} \text{runner} : \text{Ind} \\ \text{time} : \text{Time} \\ \text{place} : \text{Loc} \end{bmatrix}$
- and of the type $\begin{bmatrix} \text{runner} : \text{Ind} \\ \text{time} : \text{Time} \end{bmatrix}$ and of the type $\begin{bmatrix} \text{runner} : \text{Ind} \end{bmatrix}$ and of the type $\begin{bmatrix} \; \end{bmatrix}$, the type that imposes no constraints.

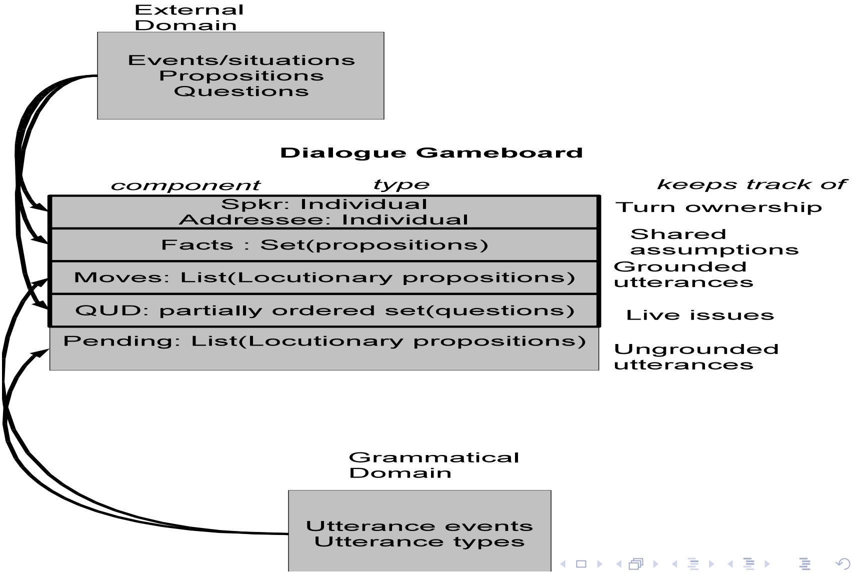
An event

- A situation with a woman riding a bicycle would then be a record

$$\begin{bmatrix} \ldots \\ \text{x = a} \\ \text{c1 = p1} \\ \text{y = b} \\ \text{c2 = p2} \\ \text{time = t0} \\ \text{loc = l0} \\ \text{c3 = p3} \\ \ldots \end{bmatrix} \text{ of type } \begin{bmatrix} \text{x: IND} \\ \text{c1: woman(x)} \\ \text{y: IND} \\ \text{c2: bicycle(y)} \\ \text{time : TIME} \\ \text{loc:LOC} \\ \text{c3: ride(x,y,time,loc)} \end{bmatrix}$$

such that: a:IND; c1: woman(a); b: IND; p2: bicycle(b); t0 : TIME; l0 : LOC;p3: ride(a,b,t0,l0);

KoS: Linking up the external world, grammar, and interaction



Simple assertion and querying: ingredients

- Querying: increment QUD with q (q becomes discourse topic)
- Assertion: increment QUD with p? (p? becomes discourse topic)
- Acceptance: decrement p? from QUD, increment FACTS with p

A simple example

(3) a. A: Hi
B: Hi
A: Who's coming tomorrow?
B: Several colleagues of mine (are coming).
A: I see.
B: Mike (is coming) too.

A simple example

Utt.	DGB Update (Conditions)	Rule
initial	MOVES = $\langle\rangle$ QUD = $\langle\rangle$ FACTS = cg1	
1	LatestMove := Greet(A,B)	greeting
2	LatestMove := CounterGreet(B,A)	countergreeting
3	LatestMove := Ask(A,B,q0)	Free Speech
	QUD : = $\langle q0\rangle$	Ask QUD–incrementation
4	LatestMove := Assert(B,A,p1) (About(p1,q0))	QSPEC
	QUD : = $\langle p1?, q0\rangle$	Assert QUD–incrementation
5	LatestMove := Accept(A,B,p1)	Accept
	QUD := $\langle q0\rangle$ FACTS := cg1 $\wedge$ p1	Fact update/QUD downdate
6	LatestMove := Assert(B,A,p2) (About(p2,q0))	QSPEC
	QUD : = $\langle p2?, q0\rangle$	Assert QUD–incrementation

Content construction for NSUs in a Dialogue–oriented constructionist grammar

- Yes—informal meaning: max-qud's proposition
- max-qud([ ])
- $\begin{bmatrix} \text{phon : yes} \\ \text{cat = adv : syncat} \\ \text{max-qud : PolarQuestn} \\ \text{cont = max-qud([]): Prop} \end{bmatrix}$

Content construction for NSUs in a Dialogue–oriented constructionist grammar

- Short answer—informal meaning: Function application of max-qud to fragment's content; syn parallelism with FEC
- max-qud(frag.cont)
- $$\begin{bmatrix} \text{cat} = \text{V[+fin]} : \text{syncat} \\ \text{hd-dtr} : \begin{bmatrix} \text{cat} = \text{max-qud.fec.cat} : \text{Syncat} \end{bmatrix} \\ \qquad \wedge\ \text{sign} \\ \text{max-qud} : \text{WhQuestn} \\ \text{cont} = \text{max-qud(hd-dtr.cont)} : \text{Prop} \end{bmatrix}$$

Genre specificity

- Relevance driven by the domain plays an important role, as emphasized by a vast literature in AI, going back at least to Cohen & Perrault, 1979; Allen & Perrault, 1980.
- In some cases the activity is very clearly defined and tightly constrains what can be said. In other cases the activity is far less restrictive on what can be said:

(4) a. **Buying a train ticket**: c wants a train ticket: c needs to indicate where to, when leaving, if return, when returning, which class, s needs to indicate how much needs to be paid

b. **Buying in a boulangerie**: c needs to indicate what baked goods are desired, b needs to indicate how much needs to be paid

c. **Chatting among friends**: first: how are conversational participants and their near ones?

d. **Buying in a boulangerie from a long standing acquaintance**: combination of (b) and (d).

Genre-based Relevance

- An account of genre-based relevance presupposes a means of classifying a conversation into a genre.
- One way of so doing is by providing the description of an information state of a conversational participant who has *successfully* completed such a conversation.
- Final states of a conversation will then be records of type T for T a subtype of $\text{DGB}_{\mathit{fin}}$, here Questions No (longer) Under Discussion (QNUD) denotes a list of issues characteristic of the genre which will have been resolved in interaction:

(5) $\text{DGB}_{\mathit{fin}} = \begin{bmatrix} \text{Facts} : \text{Prop} \\ \text{QNUD} = \text{list} : \text{list(question)} \\ \text{Moves} : \text{list(IllocProp)} \end{bmatrix}$

Some Genres

- CasualChat:

$$
\begin{bmatrix}
\text{A : Ind} \\
\text{B : Ind} \\
\text{t: Time} \\
\text{c1 : Speak(A,t)} \vee \text{Speak(B,t)} \\
\text{facts : Set(Prop)} \\
\text{qnud : list(question)} \\
\text{c2:} \left\{ \lambda P.P(A), \lambda P.P(B) \right\} \subset \text{qnud} \\
\text{moves : list(IllocProp)}
\end{bmatrix}
$$

Some Genres

- BakeryChat:

$$\left[\begin{array}{l} \text{A : Ind} \\ \text{B : Ind} \\ \text{t: Time} \\ \text{c1 : Speak(A,t)} \vee \text{Speak(B,t)} \\ \text{facts : Set(Prop)} \\ \text{qnud : list(question)} \\ \text{c2:} \left\{ \begin{array}{l} \lambda P.P(A), \lambda P.P(B), \lambda x.\text{InShopBuy(A,x)}, \\ \lambda x.\text{Pay(A,x)} \end{array} \right\} \subset \text{qnud} \\ \text{moves : list(IllocProp)} \end{array}\right]$$

What drives the dialogue?

- Activity relevance: one can make an initiating move m0 if one believes that that the current conversation updated with m0 is of a certain genre G0.
- Making move *m*0 given what has happened so far (represented in *dgb*) can be *anticipated* to conclude as final state *dgb*1 which is a conversation of type G0:

(6) m0 is relevant to G0 in dgb0 for A iff A believes that there exists dgb1 such that $dgb \sqsubset dgb1$ and such that dgb1 : G0

An utterance type

(7)

$$\text{IGH} = \begin{bmatrix} \textsc{phon}: \texttt{is georges here} \\ \textsc{cat} = \text{V[+fin,+root]} : \text{syncat} \\ \text{constits} = \left\{\text{is, georges, here, is georges here}\right\}: \text{set(sign)} \\ \textsc{c-params}: \begin{bmatrix} \text{spkr: IND} \\ \text{addr: IND} \\ \text{c1 : address(s,a)} \\ \text{s0: SIT} \\ \text{l: LOC} \\ \text{g: IND} \\ \text{c3: Named(g,‘georges’)} \end{bmatrix} \\ \text{cont= Ask(spkr,addr, ?}\begin{bmatrix} \text{sit = s0} \\ \text{sit-type = In(l,g)} \end{bmatrix}\text{) : IllocProp} \end{bmatrix}$$

A witness for IGH

(8)
$$\begin{bmatrix} \text{phon} = \text{izjorjhiya} \\ \text{cat} = \text{V[+fin,+root]} \\ \text{constits} = \left\{ \begin{array}{l} \text{u1(iz),u2(jorj),} \\ \text{u3(hiya),u4(izjorjhiya)} \end{array} \right\} \\ \text{c-params} = \begin{bmatrix} \text{spkr} = \text{A} \\ \text{addr} = \text{B} \\ \text{time} = \text{t0} \\ \text{s0} = \text{sit1} \\ \text{l} = \text{l0} \\ \text{g} = \text{g0} \\ \text{c3} = \text{pr1} \end{bmatrix} \\ \text{cont= Ask(spkr,addr, ?}\begin{bmatrix} \text{sit} = \text{s0} \\ \text{sit-type} = \text{In(l,g)} \end{bmatrix}\text{)} \end{bmatrix}$$

A witness for IGH

- Typing constraints: izjorjhiya : `is georges here`, cat = V[+fin,+root]
  A, B : IND; l0 : LOC . . .

Incorporating metacommunicative interaction

- Grounding: utterance type fully classifies utterance token
- CRification: utterance type calculated is weak (e.g. incomplete word recognition); need further information to spell out token (e.g. incomplete contextual resolution).

Pending: composition

- Utterances are kept track of in a contextual attribute PENDING in the immediate aftermath of the speech event.
- Given a presupposition that *u* is the most recent speech event and that T_u is a grammatical type that classifies *u*, a record of the form $\begin{bmatrix} \text{sit} = \text{u} \\ \text{sit-type} = \text{T}_u \end{bmatrix}$ (of type LocProp (*locutionary proposition*)), gets added to PENDING.

Grounding and Clarification Interaction

- Grounding (Clark, 1996), utterance *u* understood: update MOVES with *u*
- Clarification Interaction:
  1. *u* remains for future processing in PENDING;
  2. a clarification question calculated from u, CQ(u) updates QUD (CQ(u) becomes discourse topic)

Parameter Identification

- Parameter identification: the accommodated MaxQUD is the issue *What did spkr mean by u1?*
- Parameter identification: Input:

$$\begin{bmatrix} \text{Spkr} : \text{Ind} \\ \text{MaxPending} : \text{LocProp} \\ \text{u1} \in \text{MaxPending.sit.constits} \end{bmatrix}$$

Output: $$\begin{bmatrix} \text{MaxQUD} = \text{What did spkr mean by u1?} \\ \text{LatestMove} : \text{LocProp} \\ \text{c1: CoProp(LatestMove.cont,MaxQUD)} \end{bmatrix}$$

Parameter Identification

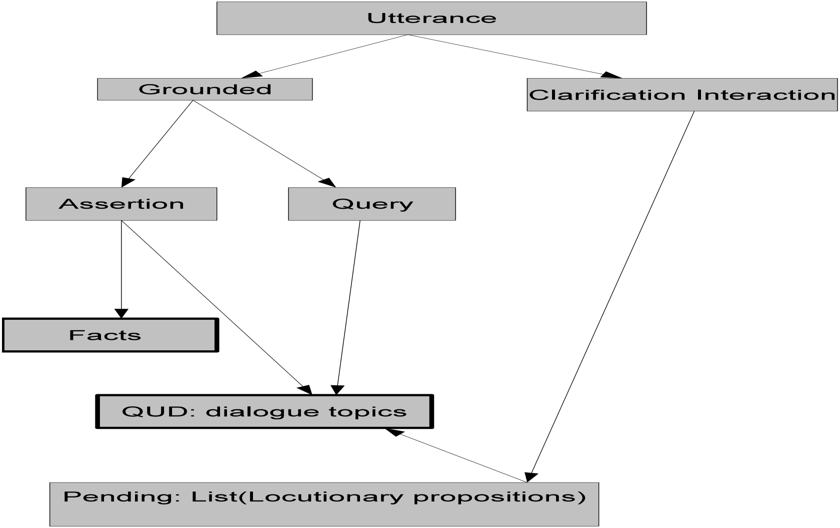
- Underpins CRs such as:

(9) A: Is Georges here?
B: Who do you mean 'Georges'?/WHO?/Georges? (= Who is 'Georges'?)
A: Georges Perec
B: Not any more.

- We can also deal with corrections, as in (10). B's corrective utterance is co-propositional with λxMean(A,u2,x), and hence allowed in by the specification.

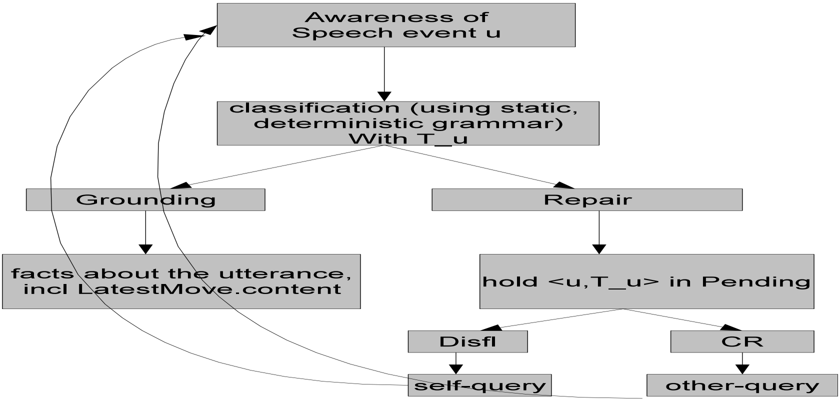
(10) a. A: Is Bo here?
b. B: You mean Jo.

Utterance processing in KoS



Where we are now

Basic picture of utterance dynamics



Utterance dynamics and language learning

- Clarification interaction here concerns problems with specific utterance *tokens*.
- Straightforward extension to incoporate dynamics at word type level. (See Purver, 2004, Cooper and Larsson, 2009)

Some missing ingredients for a model of language learning

- multimodality for bootstrapping
- accommodating uncertainty in classification that does not lead to interaction—grammar induction.

Madeline example 1

From the Colaje corpus (Parisse, C. and Morgenstern, A, 2011)

Madeline example 1 (1;07)

*OBS: un escargot *a snail*
*CHI: Sha
*MOT: berko !
*OBS: un escargot ! *a snail*
*MOT: et tu vois il est dans sa maison faut le laisser . *You see it's in its house, need to leave it.*
*MOT: faut le laisser lá . *need to leave it.*
*MOT: non c'est pas très propre . *it's not very clean*
*CHI: olaja
*MOT: attention ta tête . *watch it, your head*
*CHI: ajaja *MOT: non on va le laisser l'escargot c'est pas très propre tu sais . *need to leave it. it's not very clean*
*MOT: hop .
*MOT: c'est moi qui prends ? *I'll take it?*
*CHI: jajo
*OBS: l'escargot, t'intéresse ! *the snail interests you?*
*CHI: jajaja
*MOT: c'est un escargot Madeleine . *it's a snail, Madeleine.*

Back

Madeleine example 2

Madeleine example 2 (1;07)

*CHI: enlever . (*remove*)
*CHI: yy bébé . (yy *baby*)
*CHI: laver . (*wash*)
*CHI: l'enlever . (*remove it*)
*MOT: tu l'as lavé hier . (*you washed it yesterday*)
*MOT: oui . (*yes*)
*CHI: **bébé** . (*baby*)
*MOT: oui ah mais on n'a pas réussit à enlever ça . **Yes, oh but you didn't manage to remove it*).

*MOT: il est un peu sale ? (*it's a bit dirty*).

Back

Manchester corpus example

*CHI: wheelbarrow .
*CHI: in .
*MOT: you've got a wheelbarrow, haven't you ?
*CHI: yeah .
*MOT: it's outside .
*CHI: dolly Andy in .
*MOT: dolly and Andy were in it, weren't they ?
*MOT: but I've brought dolly in now .
*CHI: aah .
*CHI: big toe .
*MOT: pardon ?
*CHI: big toe .
*MOT: a big toe ?
*CHI: yeah .
*MOT: is that what she's got ?

Where we are now

Making the visual explicit in the DGB: some assumptions

Wolfe 2001:

- Vision before attention,
- Vision with attention,
- Vision after attention: quick decay and memory as entities, not via visual scene.
- **Visual Attention:**
  - The evidence suggests that focal attention can be directed to one or, perhaps, a few objects at any one time.
  - The number of possible targets for attention in a visual scene is usually many times that number.
  - Consequently, most of the visual stimuli that we see at any moment are represented either preattentively or postattentively.

Making the visual explicit in the DGB

- Add extra field VisField : RecType to DGB (cf. MSOA in Grosz 1987, Poesio and Rieser 2011)
- Represents dialogue participant's (view of) visual situation and attended entities.
- Basic structure: VisInf = $\begin{bmatrix} \text{VisSit} : \text{RecType} \\ \text{InAttention} : \text{Ind} \\ \text{c1} : \text{member(InAttention,VisSit)} \end{bmatrix}$
- A witness for VisInf :
  VisWitness = $\begin{bmatrix} \text{VisSit} = \text{vissit0} \\ \text{InAttention} = \text{focent1} \\ \text{c1} : \text{In(focent1,vissit0)} \end{bmatrix}$

Making the visual explicit in the DGB: an example

- Visual situation involving doll with spot on head, where the spot is attentional focus:
- : $$\begin{bmatrix} \text{VisSit} : \begin{bmatrix} \text{x: Ind} \\ \text{c1 : doll(x)} \\ \text{y : Ind} \\ \text{c2 : head(y,x)} \\ \text{z : Ind} \\ \text{c3 : spot(z)} \wedge \text{On(z,y)} \end{bmatrix} \\ \text{InAttention = VisSit.z : Ind} \\ \text{c1 : member(InAttention,VisSit)} \end{bmatrix}$$
- Just like we represent various dimensions of sign (phon, syn, sem, ctxt), potentially long term, could one do the same for the visual, viz. incorporate vector representation of scene entities (cf. Kelleher06, Chao 2010).
- Depends on ingredients needed for recovery from problems.

Making the visual explicit in the DGB

- VisInf dynamics independent of but coupled with utterance dynamics:
  - VisInf can remain static.
  - VisSit can remain static but InAttention change (‘What an ugly spot!’)
  - VisSit can change by utterance (‘hands up’)
  - Change in VisSit can cause utterance (‘Look!’)

Where we are now

Visually based non-sentential utterances

- Just about all adult utterances (e.g. (1), (3), (4),(5),(6), (8), (10)) *presuppose* shared visual context.
- The NSUs in (1), (3),(5),(7) involve multi-modally–resolved constructions.
- Canonical adult *Non*: negative simple answer to MaxQUD

$$\left[\begin{array}{l} \text{phon : non} \\ \text{cat = adv : syncat} \\ \text{dgb-params.max-qud : PolarQuestn} \\ \text{cont: Prop} \\ \text{c : NegProp} \wedge \text{SimpleAns(cont,max-qud)} \end{array}\right]$$

Back

Visually based non-sentential utterances

- Multimodal *Non*: A's objection to B's observable action

$$\left[\begin{array}{l}\text{phon : non}\\ \text{cat = adv : syncat}\\ \text{dgb-params : }\left[\begin{array}{l}\text{A: Ind}\\ \text{B: Ind}\\ \text{c : addressing(A,B)}\\ \text{v : VisInf}\\ \text{In(P(B), v.VisSit)}\end{array}\right]\\ \text{cont= }\neg\text{ Want(A,(P(B))) : Prop}\end{array}\right]$$

Visually based non-sentential utterances

- *Un escargot* an instance of the following construction: *entity InAttention has NP's descriptive property*
- Rough sketch:

(11)
$$\begin{bmatrix} \text{cat} = \text{V[+fin]} : \text{syncat} \\ \text{hd-dtr.cont} : \begin{bmatrix} \text{x} : \text{Ind} \\ \text{c1} : \text{P(x)} \end{bmatrix} \\ \text{dgb-params} : \begin{bmatrix} \text{v} : \begin{bmatrix} \text{VisSit} : \text{RecType} \\ \text{InAttention} = \text{hd-dtr.cont} : \text{Ind} \\ \text{member(InAttention,VisSit)} \end{bmatrix} \end{bmatrix} \\ \text{cont} = \text{P(hd-dtr.cont.x)} : \text{Prop} \end{bmatrix}$$

Visually based non-sentential utterances

- In terms of coherence, infelicitous as a conversation between two adults.
- Lack of deterministic grounding by child licences “redundant” repetitions.
- Instance of a *naming game* genre: *introduce referring terms to an entity InAttention*

Visually based non-sentential utterances

- Rough sketch of genre:

$$\left[\begin{array}{l} \text{A : Ind} \\ \text{B : Ind} \\ \text{c : child(B)} \\ \text{t: Time} \\ \text{c1 : Speak(A,t)} \vee \text{Speak(B,t)} \\ \text{facts : Set(Prop)} \\ \text{qnud : list(question)} \\ \text{v : } \left[\begin{array}{l} \text{VisSit : RecType} \\ \text{InAttention : Ind} \\ \text{member(InAttention,VisSit)} \end{array}\right] \\ \text{moves : list(IllocProp) c2: } \left\{ \text{Refer(A,B,v.InAttention)} \right\} \subset \text{moves} \end{array}\right]$$

Gesturally supported non-sentential utterances by children

- ‘Bébé’ has no appropriate linguistic antecedent (though clues provided.) $\mapsto$ ‘baby has the spot.’ (The baby’s spot needs removing?)
- Nor does it involve genre-specific exophoric resolution (e.g. In a boulangerie: *Une tradition*, *SVP*)
- Resolving property provided gesturally.
- Like exophoric NSUs this underdetermines content (e.g. Clark, 1996)

Gesturally supported non-sentential utterances

- Rough sketch of construction:

(12)
$$\left[\begin{array}{l}
\text{cat} = \text{V[+fin]} : \text{syncat} \\
\text{hd-dtr} : \text{sign} \\
\text{dgb-params} : \left[\begin{array}{l}
\text{a} : \text{Ind} \\
\text{b} : \text{Ind} \\
\text{addressing(a,b)} \\
\text{v} : \left[\begin{array}{l}
\text{VisSit} : \text{RecType} \\
\text{InAttention} = \text{hd-dtr.cont} : \text{Ind} \\
\text{member(InAttention,VisSit)}
\end{array}\right] \\
\text{c} : \text{Show(a,b,v.InAttention)}
\end{array}\right] \\
\text{cont} = \text{Poss(hd-dtr.cont,v.InAttention)} : \text{Prop}
\end{array}\right]$$

Back

Where we are now

Main Points

- Starting point: language viability requires interactive language learning
- Aspects of Language learning emergent from dialogue interaction.
- Enabled by uniform treatment of illocutionary and metacommunicative interaction in KoS.
- Rough sketch of extension that makes visual aspect explicit in Dialogue GameBoard.
- Application to describe visually based and gesturally supported non-sentential utterances.

Future Work

- Experimental work classifying visually based and gesturally supported NSUs.
- Pathway from visually based and gesturally supported NSUs using carer feedback to grammar induction.
- A learning theoretic account of the transition from visually based and gesturally supported NSUs to NSUs resolved with reference to linguistically produced semantic entities.
- Scaling up Macura's language simulation to NL.

Allen, J. & Perrault, R. 1980. Analyzing intention in utterances. *Artificial Intelligence*, *15*, 143–178.

Clark, A. & Lappin, S. 2010. *Linguistic Nativism and the Argument from the Poverty of the Stimulus*. Wiley Blackwells.

Clark, H. 1996. *Using Language*. Cambridge University Press, Cambridge.

Cohen, P. & Perrault, R. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, *3*, 177–212.

Cooper, R. 2005. Austinian truth in martin-löf type theory. *Research on Language and Computation*, *3*(4), 333–362.

Fernández, R. 2006. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King's College, London.

Fillmore, C. & Kay, P. 1999. Grammatical constructions and linguistic generalizations: the *what's x doing y?* construction. *Language*, *75*, 1–33.

Ginzburg, J. 1996. Interrogatives: questions, facts, and dialogue. In Lappin, S. (Ed.), *Handbook of Contemporary Semantic Theory*, pp. 359–423. Blackwell, Oxford.

Ginzburg, J. 2011. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

Ginzburg, J. & Cooper, R. 2004. Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, *27*(3), 297–366.

Ginzburg, J., Fernández, R., & Schlangen, D. 2011. Disfluencies as intra-utterance dialogue moves..

Ginzburg, J. & Sag, I. A. 2000. *Interrogative Investigations: the form, meaning and use of English Interrogatives*. No. 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.

Henderson, J., Lemon, O., & Georgila, K. 2008. Hybrid reinforcement / supervised learning of dialogue policies from fixed datasets. *Computational Linguistics*, *34*(4), 487–511.

Larsson, S. 2002. *Issue based Dialogue Management*. Ph.D. thesis, Gothenburg University.

Macura, Z. 2007. *Metacommunication and Lexical Acquisition in a Primitive Foraging Environment*. Ph.D. thesis, King's College, London.

Macura, Z. & Ginzburg, J. 2006. Lexicon convergence in a population with and without metacommunication. In Vogt, P. (Ed.), *Proceedings of EELC 2006*, No. 4211 in Lecture Notes in AI, pp. 100–112 Heidelberg. Springer.

Macura, Z. & Ginzburg, J. 2008. Dynamics and adaptiveness of metacommunicative interaction in a foraging environment. In Cooper, R. & Kempson, R. (Eds.), *Language in Flux: Dialogue Coordination, Language Variation, Change and Evolution*. College Publications, London.

Purver, M. 2004. *The Theory and Use of Clarification in Dialogue*. Ph.D. thesis, King's College, London.

Sag, I. A. 1997. English relative clause constructions. *Journal of Linguistics*, *33*, 431–484.

Schlangen, D. 2003. *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. Ph.D. thesis, University of Edinburgh, Edinburgh.