Multimodal Interaction in Dialogue and its Meaning

ESSLLI 2022 | Lecture 5

Jonathan Ginzburg
Andy Lücking

Laboratoire de Linguistique Formelle
Université Paris Cité

19th August 2022

- laughter: *Pleasant*, *Incongruent*
- head shake: ‘No’ uses, dissociated uses

1. horizontal relevance
2. vertical relevance and its consequences.
3. Iconicity in gesture.

Characterizing Dialogical Relevance: a programme

- Stage 1: RespSpace(Query,2Person): characterize relevance for queries and their responses = RespSpace(Query−specific,2Person) $\cup$ RespSpace(Utterance−general,2Person).
- Stage 2: Extend this characterization to assertions and their responses—RespSpace(Assertion,2Person): = RespSpace(Assertion−specific,2Person) $\cup$ RespSpace(Utterance−general,2Person).
- Stage 3, . . . , n: Extend this characterization to other moves (commands, exclamations, . . .) and their responses—RespSpace(2Person) = $\cup_{moves}$ RespSpace(moves,2Person) $\cup$ RespSpace(Utterance−general,2Person).

Characterizing Relevance: a programme (extensions)

- Multi-party dialogue (Ginzburg and Fernández, 2005; Ginzburg, 2012)
- Monological "text" (multi-genre: letters, press releases, etc):
  Text–Dialogicity Hypothesis:
  RespSpace(1Person) $\subsetneq$ RespSpace(2Person)

Query responses to queries I

- Starting point: the typology for responses in the form of questions provided in (Łupkowski and Ginzburg, 2017).
- A wide coverage taxonomy for question/question sequences tested on the BNC, CHILDES, BEE, AMEX , CornellMovie (English), Spokes (Polish) corpora ; formal modelling in the framework of KoS–TTR

- The study sample consisted of 1,846 query/query response pairs.
- 6 classes of questions (LG Classes) that a given query gives rise to.
  1. CR: clarification request:
     A:What's Hamlet about? B: Hamlet? [KPW, 945–946]
  2. **Dependent questions** (DP) constitute the case where the answer to the initial question (q1) depends on the answer to the query-response (q2), as in:

     A: q_1

     B: q_2

     $\mapsto q_1$ depends on q_2
     A: Does anybody want to buy an Amstrad? <*pause*> B: Are you giving it away? [KBo, 3343–3344]
     (cf. *Whether anybody wants to buy an Amstrad depends on whether you are giving it away.*)

Query responses to queries III

3. MOTIV: questions about an underlying motivation :
   A: What's the matter?
   B: Why? [HDM, 470–471]
4. CHT: questions aimed at Changing the Topic:
   A: Yeah what was your answer? B: What was yours? [KP3, 636–637]

   BBC INTERVIEWER: How did Singapore handle the pandemic so well?

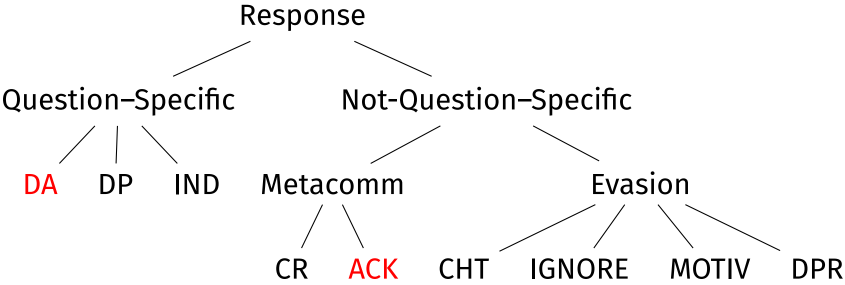
   SINGAPORE HEALTH OFFICIAL: The question should be "How did UK not handle it so well?".

   BBC INTERVIEWER: What do you mean?

   SINGAPORE HEALTH OFFICIAL: We followed 'UK Pandemic Response Protocol', the UK did not! (Twitter 24 May 2021)
5. IND: questions indirectly conveying an answer:
   A: Is your job safe? B: Well, whose job's safe? [G5L, 130–131]

6. IGNORE: responses ignoring the initial question, but addressing the same situation:
A: Well do you wanna go down and have a look at that now? <*pause*> While there's workmen there? B: Why haven't they finished yet? [KCF, 617–619]

(1) Main hypothesis: Resp(Query,2pers) = responses drawn from or concerning the 6 LG classes of questions, plus **direct answerhood** and **acknowledgements** exhaust the response space of a query.



- **Acknowledgement** (ACK)—a speaker acknowledges that s(he) has heard the question, e.g. *mhm*, *aha* etc.

  (2) A: that's about it innit?
  B: *Mm mm.*

- Propositional examples for these classes:
- MOTIV:

  REPORTER: Who did you back prime minister?
  THERESA MAY: As I said last week none of your business.
  (The Guardian, May 2019)

- CHT:

  (3) A: What's dolly's name?
  B: *It's raining.* **[BNC: KD4, 110-111]**

Hypothesis and Extended Taxonomy III

KAT: You're amazingly self-assured. Has anyone ever told you that?

PATRICK: *Go to the prom with me!* (Cornell Movie Corpus, m6, 839-840)

- DPR:

(4) A: When's the first consignment of Scottish tapes?

B: *Erm <pause> don't know.*

(5) A: Why?

B: *I'm not exactly sure.*

- Ignore:

(6) A: So does that mean that the ammeter is not part of the series, just hooked up after to the tabs?

B: *Let's take a step back.*

(7) DINO VELVET : Mister Welles . . . would you be so kind as to remove any firearms from your person?

WELLES: What are you... ?

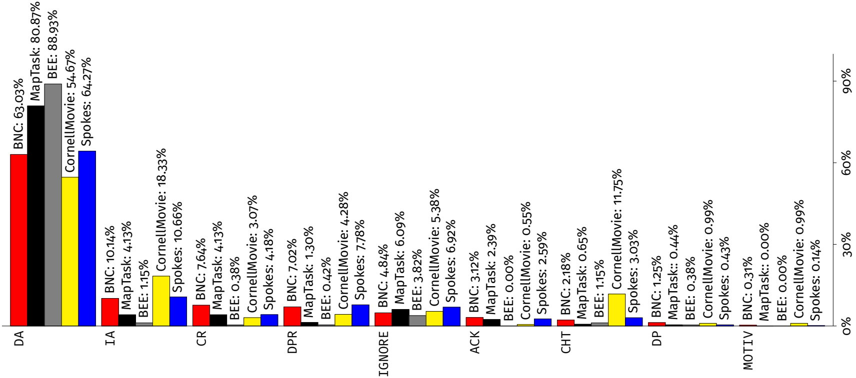
DINO VELVET : *Take out your gun!* (Cornell Movie Corpus, 6840-6842)

- The data for English comes from the BNC, BEE, the MapTask corpora, and the Cornell movie corpus (**bee**; **danescu2011chameleons**; Burnard, 2000; Anderson et al., 1991).
- The BNC data covers mainly topically unrestricted conversations; BEE contains contains tutorial dialogues from electronics courses; MapTask consists of dialogues recorded for a direction–providing task, Cornell movie corpus consists of fim scripts.
- 641 Q-R turns were taken from the BNC, 262 Q-R turns from BEE, 460 Q-R turns from the MapTask, and 911 from the Cornell movie corpus.
- Starting points: random turn selection of turn units ending with a '?'. tag questions and turns with missing text were eliminated.

- Data from Polish drawn from the Spokes corpus (**spokes**).
- The corpus currently contains 247,580 utterances (2,319,291 words) in transcriptions of spontaneous conversations.
- For the study 25 files were selected from the corpus (96,296 words, 1,424 turns)
- 694 Q-R pairs for the study.

The Corpus Study: Results I

- In all cases, the OTHER class is less than .5%, hence coverage is above 99%.
- The most frequent classes of responses in all corpora in English are direct answers (DA);
- in the BNC and and CornellMovie the next largest are indirect answers, whereas for BEE and MapTask the second largest are IGNORE.
- For Polish the two most frequent classes of responses for Spokes are answers: direct ones (DA=64.2%) and—much smaller—indirect ones (IA=10.66%).
- The next two most frequent classes are DPR (stating that it is difficult to provide an answer to the question, IDK=7.78%) and utterances ignoring the question asked (questions and declaratives, IGNORE=6.92%).



The most frequent responses types (BNC, n=506; BEE, n=256; MapTask, n=467; CornellMovie, n=678; Spokes, n=694)

Short digression on propositions and questions I

- We assume for a view of questions as propositional abstracts, for extensive motivation see (Ginzburg and Sag, 2000; Krifka, 2001)—this means that questions can be used to *underspecify* answerhood.
- Propositions are construed as typing relations between records (situations) and record types (situation types), or Austinian propositions (Austin, 1961; Barwise and Etchemendy, 1987), more recent linguistic motivation (Ginzburg, 2011).

(8) a. Propositions are records of type

$$\text{Prop} = \begin{bmatrix} \text{sit} & : \textit{Rec} \\ \text{sit-type} & : \textit{RecType} \end{bmatrix}.$$

b. p = $\begin{bmatrix} \text{sit} & = \text{s} \\ \text{sit-type} & = \text{T} \end{bmatrix}$ is true iff $p.sit : p.sit - type$ i.e., $s : T$

—the situation *s* is of the type *T*.

- Simple answerhood is the range of the propositional abstract, plus their negations.

Short digression on propositions and questions III

- We exemplify what this amounts to for some cases in (9), using as we do mostly in the sequel familiar λ-notation for wh-questions and p?-notation for polar questions, rather than the official TTR notation above:

(9) a. AtomAns(p?) = $\{p\}$
    b. AtomAns($\neg p$?) = $\{\neg p\}$
    c. AtomAns($\lambda x.P(x)$) = $\{P(a), P(b), \ldots, \}$
    d. NegAtomAns(q) = $\{p | \exists p_1 \in AtomAns(q), p = \neg p_1\}$
    e. SimpleAns(q) = AtomAns(q) $\cup$ NegAtomAns(q)

Formal Analysis: the classes DirectAns, DependentQuestion, IndirectAns I

- In fact, *simple answerhood*, though it has good coverage, is not sufficient.
- We suggest that the semantic notion relevant to direct answerhood is the relation *aboutness* (Ginzburg, 1995; Ginzburg and Sag, 2000).
- *Aboutness* must be sufficiently inclusive to accommodate conditional, weakly modalized, and quantificational answers, all of which are pervasive in actual linguistic use (Ginzburg and Sag, 2000):

(10) a. Christopher: Can I have some ice-cream then?
Dorothy: you can do if there is any. (BNC)

b. Anon: Are you voting for Tory?
Denise: I might. (BNC, slightly modified)

Formal Analysis: the classes DirectAns, DependentQuestion, IndirectAns II

c. How many players are getting these kind of opportunities to develop their potential? Not many. (The Guardian, Nov 2, 2018)

d. Dorothy: What did grandma have to catch?
Christopher: A bus. (BNC, slightly modified)

e. Elinor: Where are you going to hide it?
Tim: Somewhere you can't have it.

- Informally, Aboutness can be characterized by being a proposition that entails a disjunction of simple answers.
- and Direct answerhood by being a proposition entailed by either
  1. the conjunction of the positive atomic answers
  2. the conjunction of the negative atomic answers

- Given a notion of aboutness and some notion of (partial) exhaustiveness, one can then define question dependence (needed for the class DePQuest), though various alternative definitions have been proposed (Groenendijk and Stokhof, 1997; Wiśniewski, 2013):

(11) *q*1 depends on *q*2 iff any proposition *p* such that *p* resolves *q*2, also satisfies *p* entails *r* such that *r* is about *q*1. Ginzburg, 2012, (61b), p. 57

- With notions of aboutness and dependency in hand, one can define update rules licensing such responses.
- First, one adding a question as the maximal element of QUD following a query:

Formal Analysis: the classes DirectAns, DependentQuestion, IndirectAns IV

(12) Ask QUD-incrementation:

$$\begin{bmatrix} \text{pre} & : & \begin{bmatrix} \text{q} & : & \text{Question} \\ \text{LatestMove} = \text{Ask(spkr, addr, q)} & : & \text{LocProp} \end{bmatrix} \\ \text{effects} & : & \begin{bmatrix} \text{QUD} = \left\langle \text{q, pre.QUD} \right\rangle : & \text{poset(Question)} \end{bmatrix} \end{bmatrix}$$

- Then Gricean relevance:

(13) QSPEC: If *q* is the question under discussion, respond with an utterance *r* which is *q*–specific: About(r,q) or Depends(q,r)

(14)

QSPEC (Relevance)

$$\left[\begin{array}{l} \text{pre} : \left[\text{QUD} = \left\langle \text{q, Q} \right\rangle : \text{poset(Question)}\right] \\ \text{effects} : \\ \boxed{\dot\wedge} \left[\begin{array}{l} \text{r} : \text{Question} \vee \text{Prop} \\ \text{R} : \text{IllocRel} \\ \text{LatestMove} = \text{R(spkr, addr, r)} : \text{LocProp} \\ \text{c1} : \text{Qspecific(r, q)} \end{array}\right] \end{array}\right]$$

Formal Analysis: the classes ClarifReq, Acknowledgement I

- MetaCommunicative utterances, including acknowledgements, Clarification responses (CRs) and (metacommunicative) corrections are challenging for most existing frameworks for dialogue semantics.
- The deepest challenge is that the analysis/generation of metacommunicative utterances requires access to the entire sign associated with a given interrogative utterance.
- Any constituent, certainly down to the word level can be the object of an acknowledgement and a clarification response:

Formal Analysis: the classes ClarifReq, Acknowledgement II

(15) a. [George] Galloway [MP] is recorded reassuring his Excellency [Uday Hussein] that 'I'd like you to know we are with you 'til the end.' Who did he mean by 'we'? Who did he mean by 'you'? And what 'end' did he have in mind? He hasn't said. (From a report in the *Cambridge Varsity* by Jon Swaine, 17 February 2006)

b. Is The War Salvageable? That depends on what we mean by 'the war' and what we mean by 'salvage'. (Andrew Sullivan's Blog *The Daily Dish*, Sept, 2007)

- Moreover, there are a variety of parallelism constraints relating to the form of such utterances that require reference to the non-semantic representation of the utterance:

Formal Analysis: the classes ClarifReq, Acknowledgement III

(16) a. A: Do you fear him? B: Fear? (=What do you mean by 'fear' or Are you asking if I *fear* him) / #Afraid? / What do you mean 'afraid'?

b. A: Are you afraid of him? B: Afraid? (=What do you mean by "afraid"? or Are you asking if I am **afraid** of him) / #Fear?/What do you mean 'fear'?

- This issue rules out the lion's share of logic–based frameworks where reasoning about coherence operates solely at the level of content.
- For instance, in SDRT the semantics/pragmatics interface has no access to linguistic form, but only to a partial description of the content that is derived from linguistic form. This has been argued to be necessary to ensure the decidability of SDRT's glue logic (see e.g., Asher and Lascarides 2003, p. 77).

Formal Analysis: the classes ClarifReq, Acknowledgement IV

- In order to accommodate this class of utterances, it is crucial that the cognitive states keep track of the utterance associated with the question.
- In KoS this is handled via the field PENDING whose type (*LocProp*) is a proposition, one instantiated by an utterance token u, the other by an utterance type T_u (the sign classifying u).

(17)

$$\left[\begin{array}{l}
\text{sit} = \left[\begin{array}{l}
\text{phon} = \texttt{dijoliv} \\
\text{cat} = \text{V[+fin,+root]} \\
\text{constits} = \left\{\text{di,jow,liv}\right\} \\
\text{dgb-params} = \left[\begin{array}{l} \text{so} = \text{sito} \\ \text{to} = \text{timeo} \\ \text{j} = \text{jo} \\ \text{c3} = \text{c3o} \end{array}\right] \\
\text{cont} = ([\,])\left[\begin{array}{l} \text{sit} = \text{so} \\ \text{sit-type} = \text{Leave(j,to)} \end{array}\right]
\end{array}\right] \\
\text{sit-type} = \left[\begin{array}{l}
\text{PHON} : \texttt{did jo leave} \\
\text{CAT} = \text{V[+fin,+root]} : \text{syncat} \\
\text{constits} = \left\{\text{did, jo, leave}\right\} : \text{set(sign)} \\
\text{DGB-PARAMS} : \left[\begin{array}{l} \text{so: SIT} \\ \text{to: TIME} \\ \text{j: IND} \\ \text{c3: Named(j,jo)} \end{array}\right] \\
\text{cont} = ([\,])\left[\begin{array}{l} \text{sit} = \text{so} \\ \text{sit-type} = \text{Leave(j,to)} \end{array}\right] : \text{Questn}
\end{array}\right]
\end{array}\right]$$

Formal Analysis: the classes ClarifReq, Acknowledgement VI

- This allows *inter alia* access to the individual constituents of an utterance.

(18) $$\mathit{DGBType} \mapsto \begin{bmatrix} \text{spkr} & : \mathit{Ind} \\ \text{addr} & : \mathit{Ind} \\ \text{utt-time} & : \mathit{Time} \\ \text{c-utt} & : \text{addressing(spkr,addr,utt-time)} \\ \text{facts} & : \mathit{Set}(\mathit{Prop}) \\ \textbf{pending} & : \mathit{List}(\mathit{LocProp}) \\ \text{moves} & : \mathit{List}(\mathit{IllocProp}) \\ \text{qud} & : \mathit{poset}(\mathit{Question}) \end{bmatrix}$$

Formal Analysis: the classes ClarifReq, Acknowledgement VII

- Ginzburg and Cooper, 2004; Purver, 2004; Ginzburg, 2012 show how to account for the main classes of CRs using rule schemas of the form:
  "if *u* is the interrogative utterance and $u0$ is a constituent of *u*, allow responses that are *co-propositional* with the clarification question $CQ^i(u0)$ into QUD.", where '$CQ^i(u0)$' is one of the three types of clarification question (repetition, confirmation, intended content) specified with respect to $u0$.
- *CoPropositionality* for two questions means that, modulo their domain, the questions involve similar answers: for instance 'Whether Bo left', 'Who left', and 'Which student left' (assuming Bo is a student.) are all co-propositional.

Formal Analysis: the classes ClarifReq, Acknowledgement VIII

- Responses such as (15b) can be explicated in terms of the schema in (19):

  (19) if A's utterance *u* is yet to be grounded and u_0 is a sub-utterance of *u*, QUD can be updated with the question *What did A mean by u_0*

- Assuming a propositional function view of questions, CoPropositionality allows in propositions from the range of *Range*(q_0) and questions whose range intersects *Range*(q_0).
- Since CoPropositionality is reflexive, this means in particular that the inferred clarification question is a possible follow up utterance, as are confirmations and corrections, as exemplified in (21).

(20) Parameter identification:

$$\begin{bmatrix} \text{pre} & : & \begin{bmatrix} \text{MaxPENDING} = \begin{bmatrix} \text{sit = u} \\ \text{sit-type =T}_u \end{bmatrix} : \text{LocProp} \\ \text{A = u.dgb-params.spkr : IND} \\ \text{uo : sign} \\ \text{c1 : Member(uo,u.constits)} \end{bmatrix} \\ \text{effects} & : & \begin{bmatrix} \text{MaxQUD} = \lambda\text{xMean(A,uo,x) : Question} \\ \text{LatestMove : LocProp} \\ \text{c1: CoPropositional(LatestMove.cont,MaxQUD)} \end{bmatrix} \end{bmatrix}$$

(21) a. $\lambda x.Mean(A, uo, x)$

b. ?Mean(A,uo,b) (‘Did you mean Bo’)

c. Mean(A,uo,c) (‘You meant Chris’)

Formal Analysis: the classes Motiv, Difficult-ProvAns, ChangeTopic, Ignore I

- A quick reminder of the different evasion responses:
  - MOTIV:

    (22)

    REPORTER: Who did you back prime minister?
    THERESA MAY: As I said last week none of your business. [The Guardian, May 2019]
  - Difficult To Provide A Response:

    (23) A: When's the first consignment of Scottish tapes?
    B: *Erm <pause> don't know.* [BNC: FM2, 1061–1062]
  - Change the Topic:

    (24) A: When are you going to respond to the allegations?
    B: Anyway, when are we going to get credit for our world leading vaccination program?

- Ignore:

(25) DINO VELVET : Mister Welles ...would you be so kind as to remove any firearms from your person?
WELLES: What are you... ?
DINO VELVET : *Take out your gun!* [Cornell Movie Corpus, 6840–6842]

Formal Analysis: the classes Motiv, Difficult-ProvAns, ChangeTopic, Ignore III

- Łupkowski and Ginzburg, 2017 suggest that common to all classes of evasion utterances is a **lack of acceptance of** $q1$ **as an issue to be discussed**:
  - In MOTIV-type responses the need/desirability to discuss $q1$ is explicitly posed,
  - in CHT-type responses there is an implicature that $q1$ is of lesser importance/urgency than r_2 (expressing either a proposition or a question),
  - whereas for IGNORE type responses there is an implicature that q1 as such will not be addressed.

Formal Analysis: the classes Motiv, Difficult-ProvAns, ChangeTopic, Ignore IV

- Łupkowski and Ginzburg, 2017 also note that whereas q_1 is not accepted for discussion, it remains implicitly in the context. In (26), where move (2) could involve either a MOTIV query (2a), or a CHT query (2b), the original question has definitely *not* been re-posed and yet B still has the option to address it, which s/he should be unable to do if it is not added to his/her context before (26(2)):

(26)

A: Who are you meeting next week?
B(2)**:** (2a) What's in it for you? / (2b) Who are *you* meeting next week?
A: I'm curious.
B: Aha.
A: Whatever.
B: Oh, OK, Jill.

Formal Analysis: the classes Motiv, Difficult-ProvAns, ChangeTopic, Ignore V

A: When are you leaving? B: I don't know. A: Come on! B: Well, perhaps next week.

Formal Analysis: the classes Motiv, DifficultProvAns, ChangeTopic, Ignore VI

- This basic characteristic can be captured in the cognitive state architecture discussed above, given that QUD is assumed to be partially ordered;
- this is a crucial difference from a view of QUD as a stack or similar (Roberts, 1996; Farkas and Bruce, 2010).
- A bit more concretely, Łupkowski and Ginzburg, 2017 proposed to handle *metadiscursive* utterances such as MOTIV by viewing them as responses specific to the issue ?Wish-Answer(B,q) for a given question *q* which a conversational participant *B* can introduce as a response.

Formal Analysis: the classes Motiv, DifficultProvAns, ChangeTopic, Ignore VII

- This same approach can be applied to DPR, assuming that these involve responses specific to the issue

$$?\text{Know-Answer}(\{A, B, \ldots\}, q);$$

  we formulate this issue to address the knowledge of A, B and maybe others given that a possible response along these lines is ‘Sam knows’ or ‘Go ask Sam’.

- In both cases, in line with the fact that q remains accessible, as exemplified in (26), QUD is specified to include both q and the pertinent ‘metaquestion’.

Formal Analysis: the classes Motiv, DifficultProvAns, ChangeTopic, Ignore VIII

- In order to develop our analysis we will define a single type EvasiveResp that encompasses the commonalities between the four classes.
- Each class will then be specified by merging EvasiveResp with information specific to that particular class.
- In all cases, in line with the fact that *q* remains accessible, as exemplified in (26), QUD is specified to include both q and a pertinent 'metaquestion'.
- An additional commonality for all except DPR is turn change, underspecified for QSPEC given that for the latter it is not required, whereas in these cases it is more or less essential for coherence; this specification will be defused for DPR by using asymmetric merge.

(27) EvasiveResp=

$$\left[\begin{array}{ll}
\text{pre}: & \left[\text{QUD} = \left\langle \text{q1, Q} \right\rangle : \text{poset(Question)}\right] \\
\text{effects}: & \left[\begin{array}{l}
\text{spkr} = \text{pre.addr} : \text{Ind} \\
\text{addr} = \text{pre.spkr} : \text{Ind} \\
\text{r} : \text{Question} \vee \text{Prop} \\
\text{q2} : \text{Question} \\
\text{R: IllocRel} \\
\text{Moves} = \left\langle \text{R(spkr,addr,r)} \right\rangle \bigoplus \text{pre.Moves} : \text{list(LocProp)} \\
\text{c1} : \text{Qspecific(R(spkr,addr,r),q2)} \\
\text{QUD} = \left\langle \begin{array}{l} \text{Max} = \left\{\text{q2,q1}\right\}, \\ \text{Q} \end{array} \right\rangle : \text{poset(Question)}
\end{array}\right]
\end{array}\right]$$

- Given this, MOTIV and DPR are specified as follows:

Formal Analysis: the classes Motiv, DifficultProvAns, ChangeTopic, Ignore X

(28) a. MOTIV = EvasiveResp $\dot{\wedge}$

$$\begin{bmatrix} \text{effects}: \begin{bmatrix} \text{q2} = \text{?WishDiscuss(spkr,pre.MaxQUD)} : \text{Question} \end{bmatrix} \end{bmatrix}$$

b. DPR = EvasiveResp $\boxed{\dot{\wedge}}$

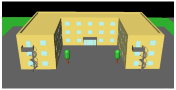
$$\begin{bmatrix} \text{effects}: \begin{bmatrix} \text{spkr} = \text{pre.spkr} \vee \text{pre.addr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{c}_{addr} : \neq\text{(addr,spkr)} \\ \text{q2} = \lambda \text{xKnow(x,pre.MaxQUD)} : \text{Question} \end{bmatrix} \end{bmatrix}$$

- for more details, see Ginzburg et al (forthcoming) 'Characterizing the response space of questions: data and theory.' *Dialogue and Discourse* .

Demonstration

‘dann ist das Haus halt so’

‘then the house is like this’



- The gesture is produced within the scope of a demonstrative ‘so’.
- The gesture contributes shape information on ‘how the house is’.

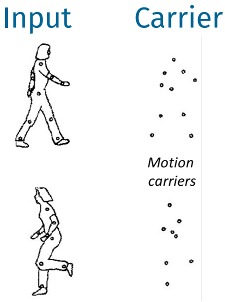
Gesture perception as type assignment

- TTR: linguistic processing as type assignment
- Schematically: $\begin{bmatrix} \text{sit} & : \text{speech event} \\ \text{sit-type} & : \text{grammatical sign} \end{bmatrix}$
- Extension to gesture: perceptual classification as type assignment

- Motion perception can be captured by means of a vector model (Johansson, 1973).
- Rotation and translation Carriers are the basis for the vector model.

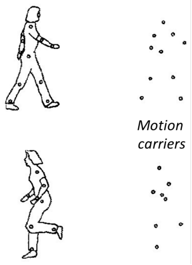
Input

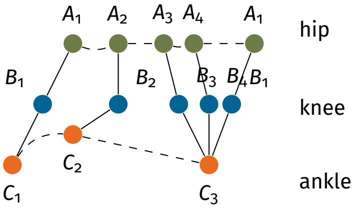
- Motion perception can be captured by means of a vector model (Johansson, 1973).
- Rotation and translation Carriers are the basis for the vector model.



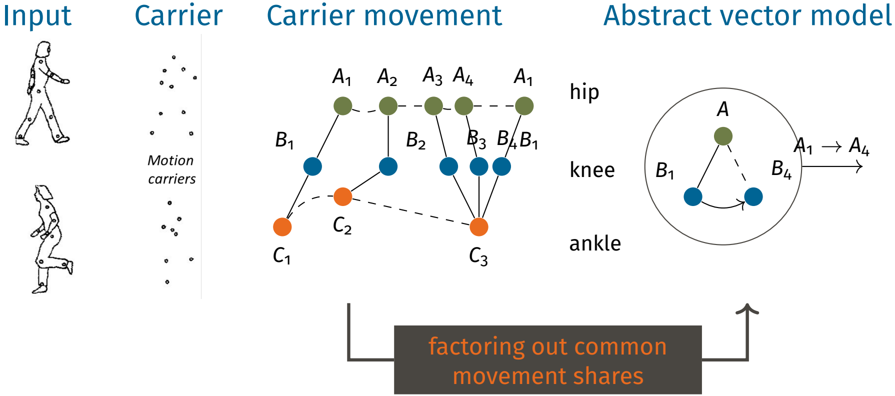
- Motion perception can be captured by means of a vector model (Johansson, 1973).
- Rotation and translation Carriers are the basis for the vector model.

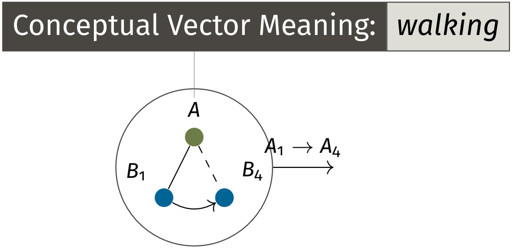
Input | Carrier | Carrier movement

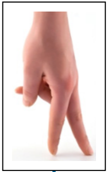




- Motion perception can be captured by means of a vector model (Johansson, 1973).
- Rotation and translation Carriers are the basis for the vector model.

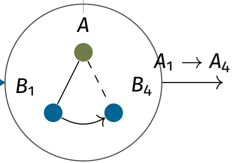



Conceptual Vector Meaning: walking
A
B_1
B_4
$A_1 \rightarrow A_4$

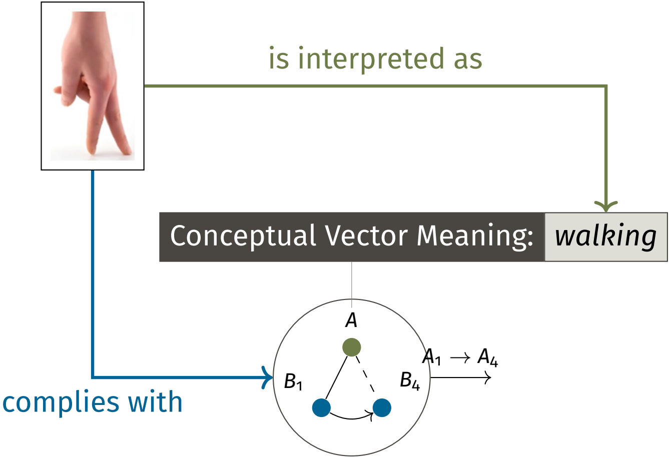


Conceptual Vector Meaning: *walking*

complies with



Gesture as vector model exemplifiers

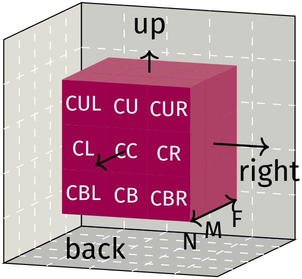


$$\begin{bmatrix} \text{hand} & = \text{right} \\ \text{hs} & = \text{claw} \\ \text{carrier} & = \begin{bmatrix} \text{boh} & = \text{none} \\ \text{plm} & = \text{none} \\ \text{wrst} & = \text{MR>MB>ML} \\ \text{move} & = \text{line>line>line} \end{bmatrix} \\ \text{sync} & = \begin{bmatrix} \text{sloc} & = \text{CBR-F} \\ \text{eloc} & = \text{CBR-N} \\ \text{stime} & = 2{:}32 \\ \text{etime} & = 2{:}33 \end{bmatrix} \\ \text{rel} & = \text{none} \end{bmatrix}$$

- Annotation format:
  - handedness (right, left)
  - handshape (modified ASL lexicon)
  - movement carrier (back-of-hand, palm or wrist; path of movement)
  - synchronized info (temporal, local)
  - relation to other hand
- The values of the features are of type *AP* (*annotation predicate*), e.g. [hs : *AP*]

Gesture Space Model

start and end locations of gesture movements are given in terms of three-dimensional **gesture space** (adapted from two-dimensional model of McNeill (1992))

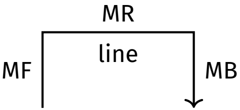


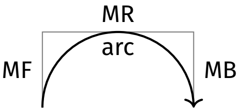
CBL:	center below left
CL:	center left
CUL:	center upper left
CB	center below
CC:	center center
…	…
N:	near
M:	middle
F:	far

Movements: lines vs. arcs

- A movement is captured in terms of a direction seen from the speaker (e.g. *move forward* (MF)) and
- a concatenation type which distinguishes straight (“line”) from roundish (“arc”) trajectories.
- Complex movements are built by combinations of directions (‘>’).

$$\begin{bmatrix} \text{wrst} & = \text{MR>MB>ML} \\ \text{move} & = \text{line>line} \end{bmatrix} \qquad \begin{bmatrix} \text{wrst} & = \text{MR>MB>ML} \\ \text{move} & = \text{arc>arc} \end{bmatrix}$$

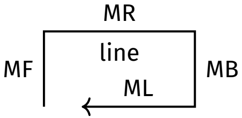




Open vs. closed paths

- Movements are underspecified with regard to the lengths of the movement parts.
- Closed and open paths are discriminated in terms of the *sync*-feature.

$$\begin{bmatrix} \text{wrst} & = \text{MF>MR>MB>ML} \\ \text{move} & = \text{line>line>line>line} \\ \text{sloc} & = \text{CC-M} \\ \text{eloc}\neq\text{sloc} & = \text{CR-M} \end{bmatrix} \qquad \begin{bmatrix} \text{wrst} & = \text{MF>MR>MB>ML} \\ \text{move} & = \text{line>line>line>line} \\ \text{sloc} & = \text{CC-M} \\ \text{eloc=sloc} & = \text{CC-M} \end{bmatrix}$$



Vector types

- Gesture annotations are mapped onto vector sequence representations **p** form spatial vector semantics (Zwarts, 2003): $\mathbf{p} : [0, 1] \mapsto \mathbf{V}$.
- Format:
  - **Type**: axis, place, outline, ... (Zwarts, 2005)
  - **Path**: description of contour (Zwarts, 2003)
  - **Shapes**: shape constraint (cf. Weisgerber, 2006)
- $Vec =_{\text{def}} \begin{bmatrix} \text{vt} : \text{Vtype} \\ \text{pt} : \text{Vpath} \\ \text{sh} : \text{multiset(Vshape)} \end{bmatrix}$
- Rule-based translation from gesture event to vector type: π_v and π_d.

Configuration	=	Vector π_v	→	Constraints π_d
Handshape $\in$ {C, 5, B, O, Y}	=	$\{\mathbf{u}\}$	→	volume
{MF, MR, MB, ML}	=	$\mathbf{u}$	→	translational
$\emptyset$	=	–	→	–
MF>MR + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MR>MB + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MB>ML + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MF>ML + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
MF>MR + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
...	=	...	→	...
MF + ...+ MB	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
ML + ...+ MR	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
sloc = eloc	=	$\mathbf{u}(0) = \mathbf{v}(1)$	→	closed
sloc $\neq$ eloc	=	$\mathbf{u}(0) \neq \mathbf{v}(1)$	→	open
lh.sloc = rh.sloc +	=	$\mathbf{u}(0) = \mathbf{v}(0)$		
lh.eloc = rh.eloc [two-handed]	=	$\mathbf{w}(1) = \mathbf{x}(1)$	→	closed
quadrant + quadrant + invers				semicircle
semicircle + semicircle + closed				circle
orthogonal + orthogonal + invers + open				rectangular
orthogonal + orthogonal + invers + closed				rectangle
...				...

Configuration	=	Vector π_v	→	Constraints π_d
Handshape $\in$ {C, 5, B, O, Y}	=	$\{\mathbf{u}\}$	→	volume
{MF, MR, MB, ML}	=	$\mathbf{u}$	→	translational
$\emptyset$	=	–	→	–
MF>MR + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MR>MB + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MB>ML + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MF>ML + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
MF>MR + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
...	=	...	→	...
MF + ...+ MB	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
ML + ...+ MR	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
sloc = eloc	=	$\mathbf{u}(0) = \mathbf{v}(1)$	→	closed
sloc $\neq$ eloc	=	$\mathbf{u}(0) \neq \mathbf{v}(1)$	→	open
lh.sloc = rh.sloc +	=	$\mathbf{u}(0) = \mathbf{v}(0)$		
lh.eloc = rh.eloc [two-handed]	=	$\mathbf{w}(1) = \mathbf{x}(1)$	→	closed
quadrant + quadrant + invers				semicircle
semicircle + semicircle + closed				circle
orthogonal + orthogonal + invers + open				rectangular
orthogonal + orthogonal + invers + closed				rectangle
...				...

Vectorizing our example

- $$\pi_v\left(\begin{bmatrix} \text{wrst} = \text{MR>MB>ML} \\ \text{move} = \text{line>line>line} \\ \text{sync} = \begin{bmatrix} \text{sloc} = \text{p1} \\ \text{eloc} = \text{p2} \neq \text{p1} \end{bmatrix} \end{bmatrix}\right) = \begin{bmatrix} \text{pt1} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \end{bmatrix}$$
- $$\pi_d\left(\begin{bmatrix} \text{pt1} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \end{bmatrix}\right) = \begin{bmatrix} \text{sh} : \left\{ \text{rectangular, open} \right\} \end{bmatrix}$$

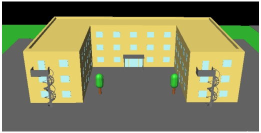
- The intensions of some predicates have a **Conceptual Vector Meaning (CVM)**, representing their perceptual impression in terms of **vector sequences** (Lücking, 2013).
- $[\![\text{U-shaped}]\!] =$

$$\begin{bmatrix} \text{x} : \text{Ind} \\ \text{c}_{\text{u}} : \text{U-shaped(x)} \\ \text{cvm} = \begin{bmatrix} \text{vt} : \text{axis-path(x, pt)} \\ \text{pt} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \\ \text{sh} : \left\{ \text{rectangular, open} \right\} \end{bmatrix} : \mathit{Vec} \\ \text{c}_{\text{shape}} : \text{shape(x, cvm)} \end{bmatrix}$$

Demonstration

‘dann ist das Haus halt so’

‘then the house is like this’



Annotation:

$$\begin{bmatrix} \text{wrst = MR>MB>ML} \\ \text{move = line>line>line} \\ \text{sync = } \begin{bmatrix} \text{sloc = p1} \\ \text{eloc = p2} \neq \text{p1} \end{bmatrix} \end{bmatrix}$$

Vector representation:

$$\begin{bmatrix} \text{pt1 : } \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \\ \text{sh : } \left\{ \text{rectangular, open} \right\} \end{bmatrix}$$

- Lexical entry: $[\![\text{house}]\!]$ =

$$\left[\begin{array}{ll}\text{bg} = \left[\text{x} : \text{Ind}\right] & \\ \text{f} & = \lambda r : \text{bg} \,.\, \left(\left[\begin{array}{l}\text{c}_{\text{hs}} : \text{house}\,(r.\text{x}) \\ \text{cvm} : \mathit{Vec} \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x}, \text{cvm})\end{array}\right]\right)\end{array}\right]$$

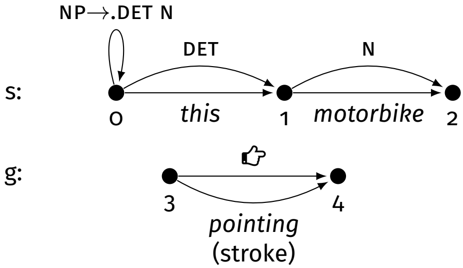
- Information state after processing the noun:

$$s_{t+1} = \left[\begin{array}{ll}\text{x} & : \text{Ind} \\ \text{c}_{\text{hs}} & : \text{house}\,(\text{x}) \\ \text{cvm} & : \mathit{Vec} \\ \text{c}_{\text{shape}} & : \text{shape}(\text{x}, \text{cvm})\end{array}\right]$$

- Gesture updates cvm of s_{t+2} and introduces additional predicate *U-shaped* via perceptual linking:

$$s_{t+2} = \begin{bmatrix} \mathrm{x} & : \mathrm{Ind} \\ \mathrm{c_{hs}} & : \mathrm{house}(\mathrm{x}) \\ \mathrm{cvm{=}dp} & : \mathit{Vec} \\ \mathrm{c_{shape}} & : \mathrm{shape}(\mathrm{x}, \mathrm{cvm}) \\ \mathrm{dp} & = \begin{bmatrix} \mathrm{pt} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \\ \mathrm{sh} : \left\{ \mathrm{rectangular, open} \right\} \end{bmatrix} : \mathit{Vec} \\ \mathrm{c_u} & : \mathrm{U\text{-}shaped}(\mathrm{x}) \end{bmatrix}$$

- $\approx$ ‘U-shaped house’



Possible **multicharts**, licensed by tier-crossing grammar rules (Johnston, 1998):

- $\{(s, 0, 1), (g, 3, 4)\}$,
- $\{(s, 1, 2), (g, 3, 4)\}$,
- $\{(s, 0, 2), (g, 3, 4)\}$

- 'Ensembles' (Lücking, 2013)

$$
\begin{bmatrix}
\textit{sg-ensemble} & \\
\text{PHON} & \boxed{12} \\
\text{CAT} & \boxed{2} \\
\text{CONT} & \boxed{3}\begin{bmatrix}\text{RESTR} & \left\langle \ldots, \boxed{5}\begin{bmatrix}\textit{pred} & \\ \text{CVM} & \boxed{1}\end{bmatrix}, \ldots \right\rangle\end{bmatrix} \\
\text{S-DTR} & \begin{bmatrix}\textit{verbal-sign} & \\ \text{PHON} & \boxed{12}\begin{bmatrix}\text{ACCENT} & \boxed{6}\end{bmatrix} \\ \text{CAT} & \boxed{2} \\ \text{CONT} & \boxed{3}\end{bmatrix} \\
\text{G-DTR} & \begin{bmatrix}\textit{gesture-vec} & \\ \text{AFF} & \left\langle\begin{bmatrix}\text{PHON} & \begin{bmatrix}\text{ACCENT} & \boxed{6}\,\textit{marked}\end{bmatrix}\end{bmatrix}\right\rangle \\ \text{TRAJ} & \boxed{1} \\ \text{CONT} & \begin{bmatrix}\text{MODE} & \textit{exemplification} \\ \text{EX-PRED} & \boxed{5}\end{bmatrix}\end{bmatrix}
\end{bmatrix}
$$

- other approaches:
  - assigning underspecified semantic descriptions to gesture morphology (instead of perceptual processing) (Alahverdzhieva, Lascarides and Flickinger, 2017)
  - speech and gestures as mutually communicating channels (instead of grammar) (Rieser and Lawler, 2020)
- various approaches needed since ensembles not appropriate for any kind of gesture ➔ e.g., dissociated uses of speech laughter or head shake

References I

Alahverdzhieva, Katya, Alex Lascarides and Dan Flickinger (2017). 'Aligning speech and co-speech gesture in a constraint-based grammar'. In: *Journal of Language Modelling* 5.3, pp. 421–464.

Anderson, Anne H. et al. (1991). 'The HCRC Map Task Corpus'. In: *Language and Speech* 34.4, pp. 351–366.

Asher, Nicholas and Alex Lascarides (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Austin, John L. (1961). 'Truth'. In: *Philosophical Papers*. Ed. by James Urmson and Geoffrey J. Warnock. Paper originally published in 1950. Oxford University Press.

Barwise, Jon and John Etchemendy (1987). *The Liar*. New York: Oxford University Press.

Burnard, L. (2000). *Reference Guide for the British National Corpus (World Edition)*. Oxford Universtity Computing Services.

References II

Farkas, Donka and Kim Bruce (2010). 'On Reacting to Assertions and Polar Questions'. In: *Journal of Semantics* 27:1, pp. 81–118.

Ginzburg, Jonathan (1995). 'Resolving Questions, I'. In: *Linguistics and Philosophy* 18, pp. 459–527.

— (2011). 'Situation Semantics and the Ontology of Natural Language'. In: *The Handbook of Semantics*. Ed. by Klaus von Heusinger, Claudia Maierborn and Paul Portner. Walter de Gruyter.

— (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Ginzburg, Jonathan and Robin Cooper (2004). 'Clarification, Ellipsis, and the Nature of Contextual Updates'. In: *Linguistics and Philosophy* 27.3, pp. 297–366.

References III

Ginzburg, Jonathan and Raquel Fernández (2005). 'Scaling up to multilogue: some benchmarks and principles'. In: *Proceedings of the 43rd Meeting of the Association for Computational Linguistics*. Michigan, pp. 231–238.

Ginzburg, Jonathan and Ivan A. Sag (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives*. CSLI Lecture Notes 123. Stanford: California: CSLI Publications.

Groenendijk, Jeroen and Martin Stokhof (1997). 'Questions'. In: *Handbook of Logic and Linguistics*. Ed. by Johan van Benthem and Alice ter Meulen. Amsterdam: North Holland.

Johansson, Gunnar (1973). 'Visual Perception of Biological Motion and a Model for its Analysis'. In: *Perception & Psychophysics* 14.2, pp. 201–211. DOI: `doi.org/10.3758/BF03212378`.

References IV

- Johnston, Michael (1998). 'Unification-based Multimodal Parsing'. In: *Proceedings of the 36th Annual Meeting on Association for Computational Linguistics – Volume I*. Montreal, Quebec, Canada, pp. 624–630.
- Krifka, M. (2001). 'For a structured meaning account of questions and answers'. In: *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow* 52, pp. 287–319.
- Lücking, Andy (2013). *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. Zugl. Diss. Univ. Bielefeld (2011). Berlin and Boston: De Gruyter.
- Łupkowski, Paweł and Jonathan Ginzburg (2017). 'Query responses'. In: *Journal of Language Modelling* 4.2, pp. 245–292.
- McNeill, David (1992). *Hand and Mind – What Gestures Reveal about Thought*. Chicago: Chicago University Press.
- Purver, Matthew (2004). 'The Theory and Use of Clarification in Dialogue'. PhD thesis. King's College, London.

References V

Rieser, Hannes and Insa Lawler (2020). 'Multi-modal meaning – An empirically-founded process algebra approach'. In: *Semantics and Pragmatics* 13.8, n/a. DOI: 10.3765/sp.13.8.

Roberts, Craige (1996). 'Information structure in discourse: Towards an integrated formal theory of pragmatics'. In: *Working Papers in Linguistics-Ohio State University Department of Linguistics*. Reprinted in Semantics and Pragmatics, 2012, pp. 91–136.

Weisgerber, Matthias (2006). 'Decomposing Path Shapes: About an Interplay of Manner of Motion and 'The Path''. In: *Proceedings of Sinn und Bedeutung*. Ed. by Christian Ebert and Cornelia Endriss. SuB 10, pp. 405–419. DOI: 10.18148/sub/2006.v10i2.741.

Wiśniewski, Andrzej (2013). *Questions, Inferences, and Scenarios*. London, England: College Publications.

References VI

Zwarts, Joost (2003). 'Vectors Across Spatial Domains: From Place to Size, Orientation, Shape, and Parts'. In: *Representing Direction in Language and Space*. Explorations in Language and Space 1. Oxford, NY: Oxford University Press. Chap. 3, pp. 39–68.

— (2005). 'Prepositional Aspect and the Algebra of Paths'. In: *Linguistics and Philosophy* 28.6, pp. 739–779. DOI: 10.1007/s10988-005-2466-y.