

MULTIMODAL INTERACTION IN DIALOGUE AND ITS MEANING

ESSLLI 2022 | LECTURE 1

JONATHAN GINZBURG
ANDY LÜCKING

LABORATOIRE DE LINGUISTIQUE FORMELLE
UNIVERSITÉ PARIS CITÉ

15TH AUGUST 2022



1. Multimodality and interaction: basic phenomena and requirements for a formal account; antecedents, TTR, KoS: basic notions
2. Quantification for dialogue
3. Applying the framework I: Pointing (deictic and discourse pointing); (some) iconic co-speech gestures
4. Applying the framework II: Laughter, head shake, mood, and dis-/association between speech and other tiers (illocutionary)
5. Relevance: from horizontal to vertical

Lecturers: Jonathan Ginzburg & Andy Lücking

Email: yonatan.ginzburg@u-paris.fr,
yonatanginzburg@gmail.com, luecking@em.uni-frankfurt.de

Website: <https://aluecking.github.io/ESSLLI2022/>

1. Develop a notion of (multi-)modality
2. Antecedents (speech act theory, language games, formal semantics, conversational analysis, ...)
3. KoS: some basics
4. Type Theory with Records (TTR): some basics

- '[i]t is not easy to define the notion of “modality”, or of “multimodality”' (Gibbon, Mertins and Moore, 2000, p. 102).
- Part of the difficulty is that there are two closely related terms, namely 'medium' and 'code' (the following is based on Lücking and Pfeiffer, 2012).
- In order to set the stage, let us simplistically assume that communication is the exchange of information between interlocutors—cf. Reddy's **conduit metaphor** (Reddy, 1979).
- The general scheme is the following:
Sender A transmits via M information X by means of sign Z to recipient B.

- *A* and *B* are uncontroversial
- Information *X* is an abstract entity that cannot be accessed directly by the human senses.
- Information has to be conveyed by means of a perceptible form that is open to interpretation – that is, by means of a *sign Z*.
- A sign is packaged into a physical container in order to be perceptible, it is transmitted via a **medium M**.
- The physical medium is accessible to the humans sense – the **modality** aspect of information.
- A sign is associated with ‘guidelines’ for interpretation – the **code** (Chandler, 2002, §. 9).

- Summing up: a sign is packaged into a 'sign container', and this is so in both a concrete and a virtual sense:
 1. In a concrete sense the container is the sign's medium, that is, a material device like ink or paper.
 2. Virtual containers are the kinds of information type or the sign systems the sign belongs to, that are, for instance, written or spoken language, figures, tables, paintings, gestures, music, or film.

- Modalities and media at our disposal:
 - ▶ *modality*: visual, auditory, chemical, tactile, kinaesthetic, vestibular
 - ▶ *medium*: sound waves, molecules, light waves
- Channels:
 - ▶ the optical channel
 - ▶ the acoustic channel
 - ▶ the chemical channel
 - ▶ the thermic channel
 - ▶ the tactile channel
 - ▶ and the electric and magnetic channels

- **Multimodality:** communication is *multimodal* iff the sign(s) Z exchanged between interlocutors A and B are perceived by the recipient via more than one sensory interface.
- **Multimediality:** communication is *multimedia* iff interlocutors A and B use more than one means M for transmitting Z .
- **Multicodality:** communication is *multicodal* iff the information X transmitted between interlocutors A and B is encoded in signs Z that belong to more than one sign system (i.e. are of more than one information type).
- Otherwise, communication is *unimodal*, *-media*, or *-codal*.

■ Questions:

1. newspaper page with text and image?
2. podcast?
3. speech and laughter?
4. speech and facial expression?
5. ...

■ Note that 'multimodality' is often used inconsistently...

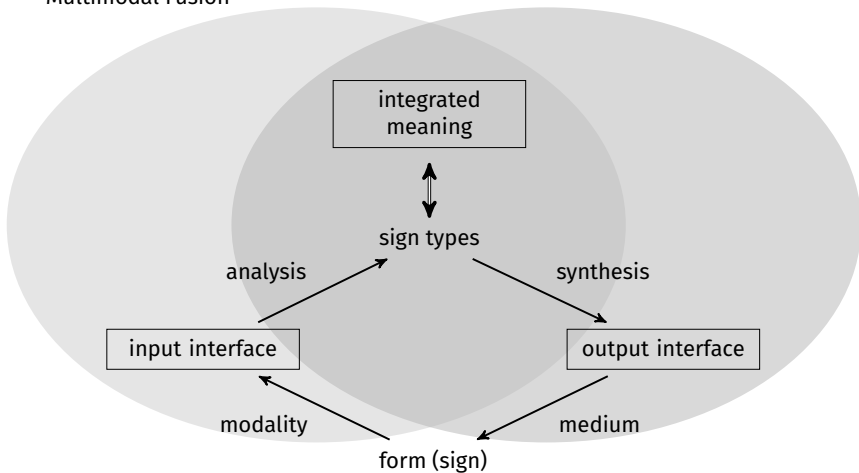
- We are dealing with human–human interaction as opposed to technical communication.
- We call a communication instance *technical*, iff
 - ▶ either one of the interlocutors *A* or *B* is an artificial device (the other being human), or both are artificial devices (say, robots or avatars); or
 - ▶ the transmission means *M* is an artificial device.
- The first distinction distinguishes *human-human* (HHI) from *human-computer* (HCI) and *computer-computer* (CCI) interaction.
- The second distinction separates *technically mediated* from *technically unmediated* communication

- Multimodality, multimediality and multicodality appear to be assymetric with regard to comprehension and production:
 - ➔ *multimodal fusion*, the integration of heterogeneously accessed information;
 - ➔ *multimedia fission*, the 'extegration' of unified information onto (possibly) different media.

FUSION/FISSION II

Multimodal Fusion

Multimedia Fission



OVERVIEW OF NONVERBAL COMMUNICATION MEANS I

- paralinguistic signals / speech phonation
- laughter, sighing / non-speech phonation (focal point of Lect. 4)
- manual gesture (focal point of Lect. 3)
- body signals
- facial signals
- gaze
- proxemics (Hall, 1968): spatio-social behavior connected to the interpersonal distances of interlocutors.
- tactile codes
- time behaviour ('chronemics'): length of laughter, respiratory pauses, or the duration of a visit (Poyatos, 1975); 'kairomics': at the right moment

OVERVIEW OF NONVERBAL COMMUNICATION MEANS II

- synchrony: temporal relations between different communication means (including the verbal one)
- Others, hitherto not much studied in communicative interaction: clothing, smelling (but see e.g., (Speed et al., 2021))

DIALOGICAL RELEVANCE: TURING STYLE

1. Dialogical **Relevance** in the sense of *conversational coherence* is the most fundamental notion for research on dialogue.
2. Some examples for relevant responses to a query and to an assertion are given in (1a,b) and irrelevant (indicated by '#') to both in (1c).
 - (1) a. A: Is that chair new? B: ✓ Yes/It's a Louis XIV replica/new?;
b. A: Jill arrived late last night. B: ✓ She did not./Why?/Jill?/To spite us.
c. B: # Tomorrow/Please insert your card/The train.
3. It is the cornerstone of theories of dialogue in the same way that *grammaticality* is to syntax.
4. Alan Turing: basic test for intelligence as a benchmark for theories of dialogue (Turing, 1950)

SPEECH ACTS: AUSTIN, AND SEARLE I

- 53 years since the publication of Searle's *Speech Acts* (Searle, 1969).
- Searle offers a systematic (but in some respects simplified) approach to a view of language initiated by his supervisor Austin.
- (Austin, 1962): language is a subspecies of *action*:
 - locutionary act (the linguistic action performed)
 - illocutionary act (the direct effect)
 - perlocutionary act (indirect effects)
- Austin: much of language is not merely *assertion*. (Though, as we shall see, Austin has very important things to say about assertion. (Austin, 1961))

SPEECH ACTS: AUSTIN, AND SEARLE II

- Searle: two dimensional view of content:
 - (2) a. Illoc-force(propositional-content)
 - b. Illoc-force includes { assert, threaten, promise, query, command }
 - (3) a. Emmanuel Macron will win the 2022 elections.
 - b. Will Emmanuel Macron win the 2022 elections?
 - c. (uttered to EM:) Win the 2022 elections!
- Problem: Speech act theory about isolated acts, no theory of context, in particular of relational dependencies (Question/Answer, Assertion/acceptance, Greeting/Counter-greeting etc)
- The lack of global structure is an issue for Speech Act Theory's contemporary version RSA (Goodman and Frank, 2016)

- Montague and his student Kaplan developed important initial analyses of context dependence: crucial distinction between meaning/character and content.
- Crucial for analyzing indexicals such as 'I', 'you', 'now'.
- Crucial for analyzing all words/phrases in conversation ...
- Problem: Generalized Quantifier Theory (Barwise and Cooper, 1981) one of the jewels of formal semantics relies on problematic denotations from a dialogical perspective.

- The Chomskyan view of language as a (disembodied, internal, non-communicative) biological endowment, emphasis on tight link between grammar and language acquisition.
- Problems:
 - ▶ interaction is crucial for understanding language acquisition (e.g., for explaining why wh-questions are acquired before polar questions, (Moradlou et al., 2021));
 - ▶ No opposition between I-language and E-language—both are needed, but the former brain-based.

CONVERSATION ANALYSIS I

- Everything so far has been arm chair theorizing...
- Real dialogues ('[' marks overlap):

(4)

1. Fri: They still haven't figured out, (.) how they're gonna get to the country: < who's gonna take care of huh m:othah while [they're- y'know 'p in the country. on the weekend. (**disfluency**)
2. Dav: [Mm (0.2 secs) (**non-sentential utterance**)
3. Fri: So: (.) you know, (0.8 secs)
4. Fri: an besides tha[:t,
5. Rub: [You c'n go any[way
6. Dav: [Don - Don git- don [get] (**disfluency**)
7. Fri: [they] won t be:
8. Dav: Y know there- there s no- no long explanation is necessary (**disfluency**)

9. Fri: Oh noon no: (**interjection**), (**disfluency**)
I'm not- I jus: : uh-wanted: you to know that you can
go up anyway.= (**overlapping turns**)
10. Rub: =Yeah:. (0.1 secs) (**non-sentential utterance**)
11. Fri: You know. (0.2 secs)
12. Fri: Because-ah (3.3 secs) (**disfluency**)
13. Rub: They don mind honey they're jus not gonna
talk to us ever again.= (**overlapping turns**)
14. Dav: = (laughter) / ri:(h)ight) (**non-sentential
utterance**)
- (From E.A. Schegloff, 2001)

CONVERSATION ANALYSIS III

- Conversation Analysis pioneered a different mode of theorizing based on taking seriously what happens in real conversations.
- Important insights include:
 1. Importance of *adjacency pairs* as markers of conversational structure.
 2. Disfluencies are not noise but acts of *self-repair* (Emanuel Schegloff, Gail Jefferson and Sacks, 1977) / *own communication management* (Allwood et al., 2005).
 3. Laughter is not a low-level emotional signal or a marker of jokes, but a conversational option akin to speech (G. Jefferson, 1979).
- Problems:
 1. No theory of context beyond adjacency.
 2. No semantics developed, which leads to explanatory poverty (case study: laughter, lecture 4).

3. CA is wedded to the idea that one speaker at a time is a fundamental *norm* of conversation. This is problematic once one considers multimodal interaction.

- The need for a synthesis—a synthesis that can strive to account for dialogical relevance:
 1. An *(inter)active* stance (Today)
 2. Compositional analysis of content that can deal with generalized quantification (Tomorrow)
 3. Conversational structure which underwrites the meaning of multimodal and non-sentential utterances. (Wednesday, Thursday)
 4. How content emerges from meaning, *if* it does.: need to develop approach where Self/other communication management is a natural option as “success”. (Friday)

- TTR (Cooper, 2005; Cooper, 2012; Cooper and Ginzburg, 2015): ontology for the world, for grammar, for interaction
 1. Semantics: Constructing an ontology for explicating semantic entities: events, propositions, questions,...
 2. Grammar: using this ontology to explicate speech events (utterances) and their types (Saussurean signs)
 3. Interaction: using the ontology to explicate what contexts are and how they change in interaction.
- TTR grounds KoS, a theory of cognitive states in interaction.

- Context in KoS (Ginzburg, 1994; Larsson, 2002; Purver, 2006; Fernández, 2006; Ginzburg and Fernández, 2010; Ginzburg, 2012)
- instead of a single context, analysis is formulated at a level of cognitive states, one per conversational participant.
- Each state has a private part and a part where publicized information is kept track of:

dialoguegameboard : DGBtype
private : Private
- Our focus is on understanding the structure of the publicized part, the dialogue gameboard (DGB).
- The simplest view of what this should consist of, going back to Montague (1974), is one which specifies the existence of a speaker, addressing an addressee at a particular time.

- One can represent that as follows (we will shortly explain what this amounts to formally):

$$\left[\begin{array}{ll} \text{spkr} & : \text{Ind} \\ \text{addr} & : \text{Ind} \\ \text{u-time} & : \text{Time} \\ \text{Cutt} & : \text{addressing}(\text{spkr}, \text{addr}, \text{u-time}) \end{array} \right]$$

- A really crucial point about the assumption that the DGB is not a shared entity (in other words rejecting talk of *the* context) is that there can be differences across participants in their view of the interaction.
- And this can be externalized in terms of clarification interaction, which can apply even to apparently shared information:
 - (5) a. (On the phone) A: Who's calling?
 - b. (In traffic) A: Are you honking at me?

DIALOGUE GAMEBOARDS III

- The need for DGBs to specify both shared information but also potentially information about which clarification is required is a point we will return to several times.
- Call it an *Interactive Stance*.
- We will see its impact on the theory of quantification tomorrow.

DIALOGUE GAMEBOARDS IV

- Since Montague and Kaplan there has been realization that the scope of publicized information is quite a bit wider than speaker, addressee, time.
- We assume the following structure for the DGB, which we will motivate extensively throughout the course:

DGBType =_{def}

spkr : Ind	<i>turn</i>
addr : Ind	<i>owner-</i>
utt-time : Time	<i>ship</i>
c-utt : addressing(spkr,addr,utt-time)	
Facts : Set(Proposition)	<i>shared assumptions</i>
VisSit : [InAttention : Ind]	<i>visual field</i>
Pending : list(locutionary Proposition)	<i>ungrounded utts</i>
Moves : list(illocutionaryProposition)	<i>grounded utts</i>
QUD : poset(Question)	<i>qs under disc</i>
Mood : Appraisal	<i>face</i>

- And here we would like to emphasize that two of these contextual parameters, *VisSit* and *Mood*, are probably **never** entirely identical across participants.
- Distinct interlocutors do not share the same pair of eyes—much of the time interlocutors have each other (or their phone) as their focus of attention.
- Nor do they register the same public ‘face’.
- But there are various devices such as pointing or the verbal *Look!* to effect alignment.

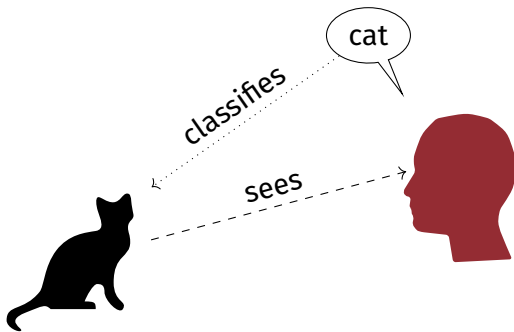
THE DIALOGUE GAMEBOARD

- The visual situation is a key component in interaction from birth (see Tomasello, 1999, Chap. 3), playing a major role in interlocutor *attention* (Mundy and Newell, 2007), itself a cornerstone for discourse participation and pointing, as discussed in Lect. 3.
- **FACTS** represents the shared knowledge conversationalists utilize during a conversation. More operationally, information a conversationalist can use embedded under presuppositional operators.
- **MOVES**: useful to single out LatestMove, a distinguished fact that characterizes the most recent move made.
- The main motivation—to segregate from the entire repository of presuppositions information on the basis of which coherent reactions could be computed.
- Later on see that keeping track of more than just the latest move can be useful.

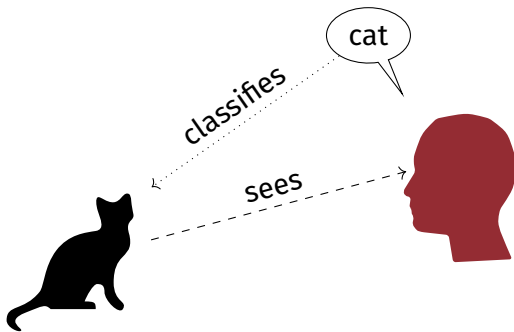
- **QUD:** (mnemonic for Questions Under Discussion)—questions that constitute a “live issue”. That is, questions that have been *introduced for discussion* at a given point in the conversation and not yet been *downdated*.
- There are additional, indirect ways for questions to get added into QUD, the most prominent of which is during clarification interaction.
- Being maximal in QUD (MAX-QUD) corresponds to being the current ‘discourse topic’ and is a key component in the theory.

- We begin to clarify what these representations we have been using are.
- We use Type Theory with Records (TTR) to build the semantic ontology (entities, events, propositions, questions,...), grammatical rules, and to write conversational rules.

PERCEPTION AS TYPE ASSIGNMENT



PERCEPTION AS TYPE ASSIGNMENT



individual x in
some situation

formally:

$$\begin{bmatrix} x : Ind \\ c_1 : \text{cat}(x) \end{bmatrix}$$

- The most fundamental notion of TTR is the typing *judgement* $a : T$ classifying an object a as being of type T .
 - (6) a. $s : \text{SIT}$
 - b. $b : \text{IND}$
 - c. $s : \text{run}(\text{arg1}_{\text{IND}} : b, \text{arg2}_{\text{TIME}} : t)$
 - d. $s : \text{run}(b, t)$

- A record is a set of fields assigning entities to labels, partially ordered by a notion of *dependence* between fields. Its general form is as in (7a):

(7) a.
$$\begin{bmatrix} l_1 = val_1 \\ l_2 = val_2 \\ \dots \\ l_n = val_n \end{bmatrix}$$

b.
$$\begin{bmatrix} x & = & 5 \\ e\text{-time} & = & 2\text{AM, Aug 15, 2022} \\ e\text{-loc} & = & \text{Nome} \\ c_{temp-at-in} & = & \text{sit1} \end{bmatrix}$$

- Together with records come record *types*. A record type is simply a record where each field represents a judgement rather than an assignment, as in (8).

$$(8) \quad \left[\begin{array}{l} l_1 : T_1 \\ l_2 : T_2 \\ \dots \\ l_n : T_n \end{array} \right]$$

- Record types allow us to place constraints on records.
- The basic relationship between the two is that a record r is of type RT if each value in r assigned to a given label l_i satisfies the typing constraints imposed by RT on l_i .
- More precisely,

(9) The record:

$$\begin{bmatrix} l_1 = a_1 \\ l_2 = a_2 \\ \dots \\ l_n = a_n \end{bmatrix} \text{ is of type: } \begin{bmatrix} l_1 : T_1 \\ l_2 : T_2(l_1) \\ \dots \\ l_n : T_n(l_1, l_2, \dots, l_{n-1}) \end{bmatrix}$$

iff $a_1 : T_1, a_2 : T_2(a_1), \dots, a_n : T_n(a_1, a_2, \dots, a_{n-1})$

■ The record:
$$\left[\begin{array}{lcl} x & = & 5 \text{ e-time} = 2:00\text{AM, Aug 15, 2022} \\ \text{e-loc} & = & \text{Nome} \\ c_{temp-at-in} & = & \text{sit1} \end{array} \right]$$

is of the type

$$\left[\begin{array}{lcl} x & : & \text{Ind} \\ \text{e-time} & : & \text{Time} \\ \text{e-loc} & : & \text{Loc} \\ c_{temp-at-in} & : & \text{temp_at_in}(\text{e-time}, \text{e-location}, x) \end{array} \right]$$

only if:

5 : Ind; 2:00AM, Aug 15, 2022 : Time; Nome : Loc; sit1 :
temp_at_in(2:00AM, Aug 15, 2022, Nome, 5)

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

...		x	: IND
x	= a	c1	: woman(x)
c1	= p1	y	: IND
y	= b	c2	: bicycle(y)
c2	= p2	time	: TIME
time	= to	loc	: LOC
loc	= lo	c3	: ride(x,y,time,loc)
c3	= p3		
...			

such that: a:IND; p1: woman(a); b: IND; p2: bicycle(b); to : TIME; lo : LOC;p3: ride(a,b,to,lo);

CONVERSATIONAL RULES I

- We characterize dialogue regularities in terms of *conversational rules*.
- Conversational rules are mappings between two cognitive states the *precond(ition)s* and the *effects*.
- Notationwise a conversational rule will be specified as in (10a). We will often notate such a mapping as in (10b):

$$(10) \text{ a. } r : \begin{bmatrix} \dots \\ \text{dgb1} : \text{DGB} \\ \dots \end{bmatrix} \mapsto \begin{bmatrix} \dots \\ \text{dgb2} : \text{DGB} \\ \dots \end{bmatrix}$$

$$\text{b. } \begin{bmatrix} \text{pre(conds)} : \text{RType} \\ \text{effects} : \text{RType} \end{bmatrix}$$

- An initiating greeting typically occurs dialogue initially.
- The primary contextual effect of such a greeting is simply providing the addressee with the possibility of reciprocating with a counter-greeting.
- A countergreeting simply grounds the original greeting, requires no response, nor has other contextual effects.

■ The conversational rule associated with greeting:

$$\left[\begin{array}{l} \text{pre : } \left[\begin{array}{l} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{moves} = \langle \ \rangle : \text{list}(\text{IllocProp}) \\ \text{qud} = \langle \ \rangle : \text{list}(\text{Question}) \\ \text{facts} = \text{commonground1} : \text{Prop} \end{array} \right] \\ \\ \text{effects : } \left[\begin{array}{l} \text{spkr} = \text{pre.spkr} : \text{Ind} \\ \text{addr} = \text{pre.addr} : \text{Ind} \\ \text{LatestMove} = \text{Greet}(\text{spkr}, \text{addr}) : \text{IllocProp} \\ \text{qud} = \text{pre.qud} : \text{list}(\text{Question}) \\ \text{facts} = \text{pre.facts} : \text{Prop} \end{array} \right] \end{array} \right]$$

PARTICIPANT SENSITIVE CONVERSATIONAL RULES I

- Conversational rules can come in two flavours, rules that each interlocutor applies in the same way to their cognitive state (*participant neutral*), as we have just seen.
- And rules that are specified only for particular interlocutors (*participant sensitive*).
- The latter kind of specification is, in principle, more general and is particularly important for an algorithmic perspective involving generation see e.g., (Larsson, 2002; Cooper, 2016).

PARTICIPANT SENSITIVE CONVERSATIONAL RULES II

- We exemplify a participant sensitive rule that relates to one of the most basic communicative interactions from infancy, namely visual attention directing, where A directs B to an object *o* (Lücking, 2018).
- This is a visual situation update rule, analogous to the MOVES update rules above.
- The sole difference is that in this case B needs to modify her visual situation so that it includes *o* as the visual focus, whereas A must already have updated his visual situation to effect such an act.
- The notation we use for such rules is exemplified in (11a), where the rule applies to the dialogue gameboard of current addressee, with the obvious change in the case where it applies to the current speaker. (11b) provides the specification for visual situation update rule:

(11) a.
$$\left[\begin{array}{l} \text{tcs} = \left[\begin{array}{l} \text{dgb} : \text{DGBType} \\ \text{private} : \text{Private} \end{array} \right] : \text{TCS} \\ \text{B} = \text{dgb.addr} : \text{IND} \\ \text{B.pre} = \text{T1} : \text{DGBType} \\ \text{B.effects} = \text{T2} : \text{DGBType} \end{array} \right]$$





b. Visual situation update:

$$\left[\begin{array}{l} \text{tcs} = \left[\begin{array}{l} \text{dgb} : \text{DGBType} \\ \text{private} : \text{Private} \end{array} \right] : \text{TCS} \\ \text{B} = \text{dgb.addr} : \text{IND} \\ \text{B.pre} : \left[\begin{array}{l} \text{o} : \text{Ind} \\ \text{LatestMove} = \text{DirectAttention}(\text{spkr}, \text{addr}, \text{o}) : \text{IllocProp} \end{array} \right] \\ \text{B.effects} : \left[\text{VisSit.InAttention} = \text{o} : \text{Ind} \right] \end{array} \right]$$




PLAN FOR THE REST OF THE COURSE

1. Lecture 2: QNPs in dialogue (Lücking and Ginzburg, 2022);
2. Lecture 3: Reference and pointing (Lücking, 2018); Discourse pointing (Ginzburg and Lücking, 2021); iconic co-speech gesture and property exemplification (perception as type assignment; Lücking, 2016)
3. Lecture 4: Head shaking (Lücking and Ginzburg, 2021); laughter, smiling, sighing (Ginzburg, Mazzocconi and Tian, 2020; Mazzocconi, Tian and Ginzburg, 2020)
4. Lecture 5: Towards vertical relevance (Lücking & Ginzburg, in rev.)




REFERENCES I

-  Allwood, Jens et al. (2005). 'Multimodality in Own Communication Management'. In: *Proceedings of the second Nordic WOrkshop on Multimodal Communication*. Gothenburg.
-  Austin, John L. (1961). 'Truth'. In: *Philosophical Papers*. Ed. by James Urmson and Geoffrey J. Warnock. Paper originally published in 1950. Oxford University Press.
-  — (1962). *How to do things with Words*. Harvard University Press.
-  Barwise, Jon and Robin Cooper (1981). 'Generalized Quantifiers in Natural Language'. In: *Linguistics and Philosophy* 4.2, pp. 159–219.
-  Chandler, Daniel (2002). *Semiotics: The Basics*. New York, NY: Routledge.




REFERENCES II

-  Cooper, Robin (2005). 'Austinian truth, attitudes and type theory'. In: *Research on Language and Computation* 3.4, pp. 333–362.
-  — (2012). 'Type Theory and Semantics in Flux'. In: *Handbook of the Philosophy of Science*. Ed. by Ruth Kempson, Nicholas Asher and Tim Fernando. Vol. 14: Philosophy of Linguistics. Amsterdam: Elsevier, pp. 271–323.
-  — (2016). *Type Theory and Language: From perception to linguistic communication*. Book Draft. URL: <https://sites.google.com/site/typetheorywithrecords/drafts/ttl161130.pdf?attredirects=0>.





REFERENCES III

-  Cooper, Robin and Jonathan Ginzburg (2015). 'Type Theory with Records for NL Semantics'. In: *Handbook of Contemporary Semantic Theory, Second Edition*. Ed. by Chris Fox and Shalom Lappin. Oxford: Blackwell, pp. 375–407. DOI: 10.1002/9781118882139.ch12. URL: <http://dx.doi.org/10.1002/9781118882139.ch12>.
-  Fernández, Raquel (2006). 'Non-Sentential Utterances in Dialogue: Classification, Resolution and Use'. PhD thesis. King's College, London.
-  Gibbon, Dafydd, Inge Mertins and Roger K. Moore, eds. (2000). *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. The Springer International Series in Engineering and Computer Science. New York, NY: Springer.





REFERENCES IV

-  Ginzburg, Jonathan (1994). 'An Update Semantics for Dialogue'. In: *Proceedings of the 1st International Workshop on Computational Semantics*. Ed. by H. Bunt. Tilburg: ITK, Tilburg University.
-  — (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.
-  Ginzburg, Jonathan and Raquel Fernández (2010). 'Computational Models of Dialogue'. In: *Handbook of Computational Linguistics and Natural Language*. Ed. by Alex Clark, Chris Fox and Shalom Lappin. Oxford: Blackwell.




REFERENCES V

-  Ginzburg, Jonathan and Andy Lücking (2021). 'I thought pointing is rude: A dialogue-semantic analysis of pointing at the addressee'. In: *Proceedings of Sinn und Bedeutung* 25. Ed. by Patrick Grosz et al. SuB 25. Special Session: Gestures and Natural Language Semantics. University College London (Online), pp. 276–291. URL: <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/937>.
-  Ginzburg, Jonathan, Chiara Mazzocchi and Ye Tian (2020). 'Laughter as language'. In: *Glossa: a journal of general linguistics* 5.1.
-  Goodman, Noah D and Michael C Frank (2016). 'Pragmatic language interpretation as probabilistic inference'. In: *Trends in cognitive sciences* 20.11, pp. 818–829.
-  Hall, Edward T. (1968). 'Proxemics'. In: *Current Anthropology* 9.2-3, pp. 83–108.





REFERENCES VI

-  Jefferson, G. (1979). 'A technique for inviting laughter and its subsequent acceptance/declination'. In: *Everyday language: Studies in ethnomethodology* 79, p. 96.
-  Larsson, Staffan (2002). 'Issue based Dialogue Management'. PhD thesis. Gothenburg University.
-  Lücking, Andy (Sept. 2016). 'Modeling Co-verbal Gesture Perception in Type Theory with Records'. In: *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*. Ed. by Maria Ganzha, Leszek Maciaszek and Marcin Paprzycki. Vol. 8. Annals of Computer Science and Information Systems. IEEE, pp. 383–392. DOI: 10.15439/2016F83.
-  — (2018). 'Witness-loaded and Witness-free Demonstratives'. In: *Atypical Demonstratives: Syntax, Semantics and Pragmatics* 568, p. 255.






REFERENCES VII

-  Lücking, Andy and Jonathan Ginzburg (2021). 'Saying and shaking 'No''. In: *Proceedings of the 28th International Conference on Head-Driven Phrase Structure Grammar*.
-  — (2022). 'Referential transparency as the proper treatment of quantification'. In: *Semantics and Pragmatics* 15, 4. DOI: 10.3765/sp.15.4.
-  Lücking, Andy and Thies Pfeiffer (2012). 'Framing Multimodal Technical Communication. With Focal Points in Speech-Gesture-Integration and Gaze Recognition'. In: *Handbook of Technical Communication*. Ed. by Alexander Mehler and Laurent Romary. In collab. with Dafydd Gibbon. Handbooks of Applied Linguistics 8. Berlin and Boston: De Gruyter Mouton. Chap. 18, pp. 591–644.





REFERENCES VIII

-  Mazzocconi, Chiara, Ye Tian and Jonathan Ginzburg (2020). 'What's your laughter doing there? A taxonomy of the pragmatic functions of laughter'. In: *Transactions of Affective Computing*.
-  Montague, Richard (1974). 'Pragmatics'. In: *Formal Philosophy*. Ed. by Richmond Thomason. New Haven: Yale UP.
-  Moradlou, Sara et al. (2021). 'Wh-Questions are understood before polar-questions: Evidence from English, German, and Chinese'. In: *Journal of Child Language* 48.1, pp. 157–183.
-  Mundy, Peter and Lisa Newell (2007). 'Attention, Joint Attention, and Social Cognition'. In: *Current directions in psychological science* 16.5. PMID: 19343102, pp. 269–274. DOI: 10.1111/j.1467-8721.2007.00518.x.

REFERENCES IX

-  Poyatos, Fernando (1975). 'Cross-Cultural Study of Paralinguistic "Alternants" in Face-to-Face Interaction'. In: *Organization of Behavior in Face-to-Face Interaction*. Ed. by Adam Kendon, Harris M. Richards and Mary Ritchie Key. The Hague: Mouton, pp. 285–314.
-  Purver, M. (2006). 'CLARIE: Handling Clarification Requests in a Dialogue System'. In: *Research on Language & Computation* 4.2, pp. 259–288.
-  Reddy, Michael (1979). 'The conduit metaphor'. In: *Metaphor and thought* 2, pp. 285–324.
-  Schegloff, E.A. (2001). 'Getting serious: Joke → serious' no". In: *Journal of pragmatics* 33.12, pp. 1947–1955.
-  Schegloff, Emanuel, Gail Jefferson and Harvey Sacks (1977). 'The preference for self-correction in the organization of repair in conversation'. In: *Language* 53, pp. 361–382.

REFERENCES X

-  Searle, John (1969). *Speech Acts*. Cambridge: Cambridge University Press.
-  Speed, Laura J et al. (2021). 'The sound of smell: Associating odor valence with disgust sounds'. In: *Cognitive Science* 45.5, e12980.
-  Tomasello, Michael (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
-  Turing, A.M. (1950). 'Computing Machinery and Intelligence'. In: *Mind* 59.236, pp. 433–460.