Multimodal Interaction in Dialogue and its Meaning

ESSLLI 2022 | Lecture 3

Jonathan Ginzburg
Andy Lücking

Laboratoire de Linguistique Formelle
Université Paris Cité

17th August 2022

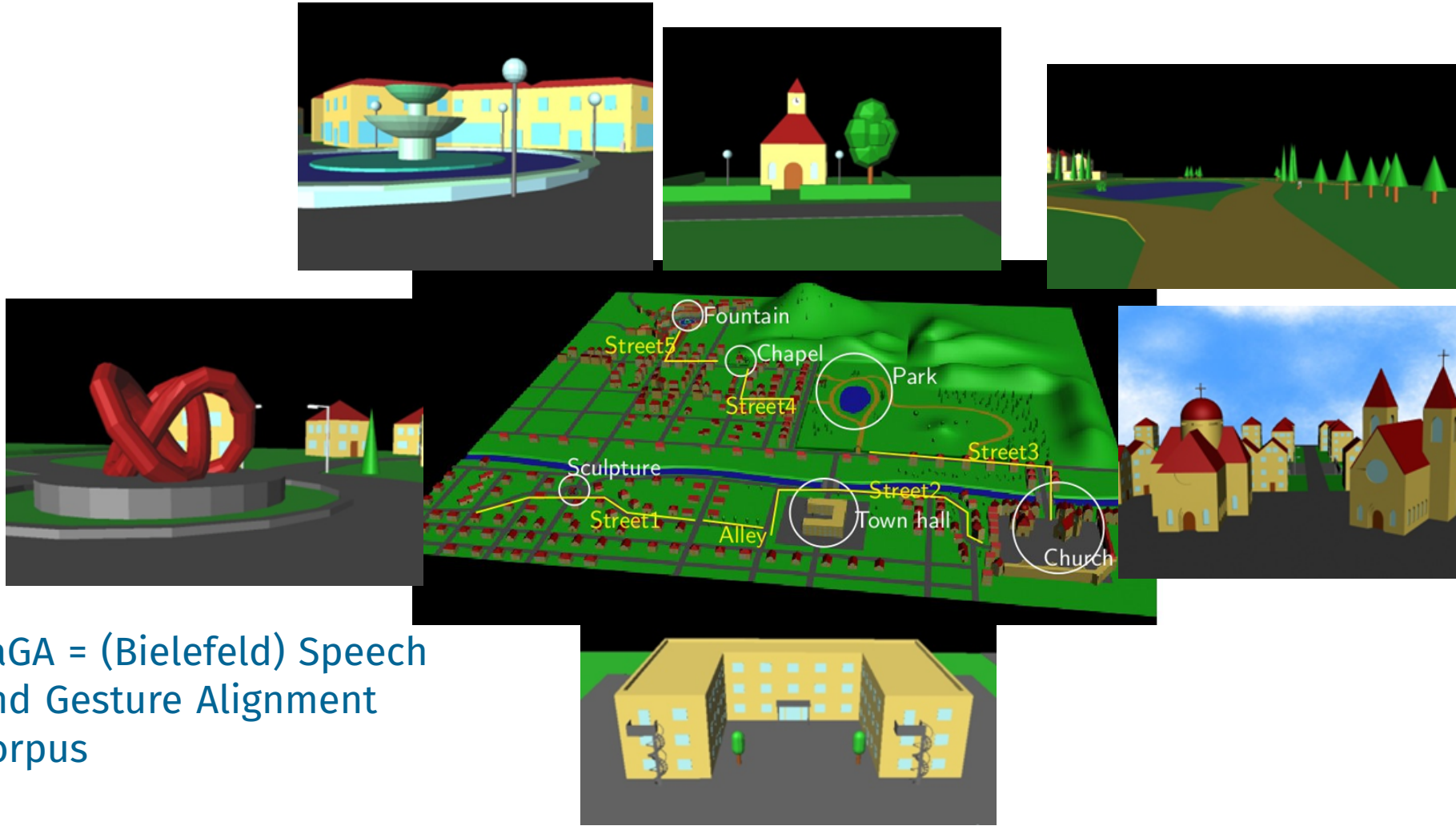
Yesterday's lecture

- dialogue-based approach to QNPs
- Referential Transparency Theory

Today's Lecture

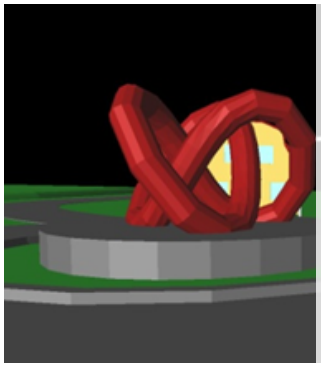
1. Pointing: from direct reference to visual attention (cf. DGB's VisSit)
2. Discourse pointing
3. Iconic gestures and exemplification (by example)

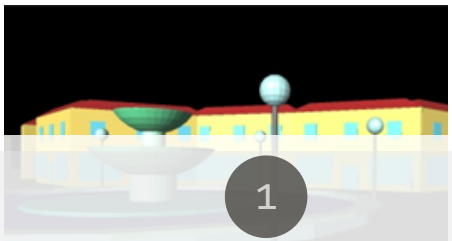
Some examples from SaGA (Lücking et al. 2010)



SaGA = (Bielefeld) Speech and Gesture Alignment Corpus

Some examples from SaGA (Lücking et al. 2010)



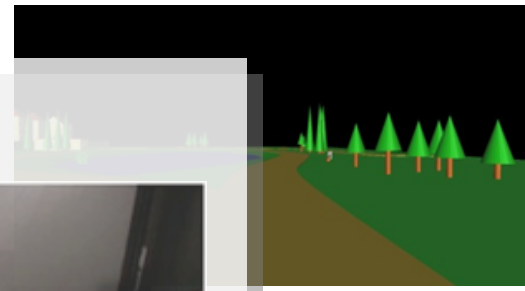


1

Virtual bus ride through SaGA town

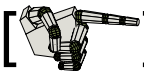
2

Giving directions, describing sights

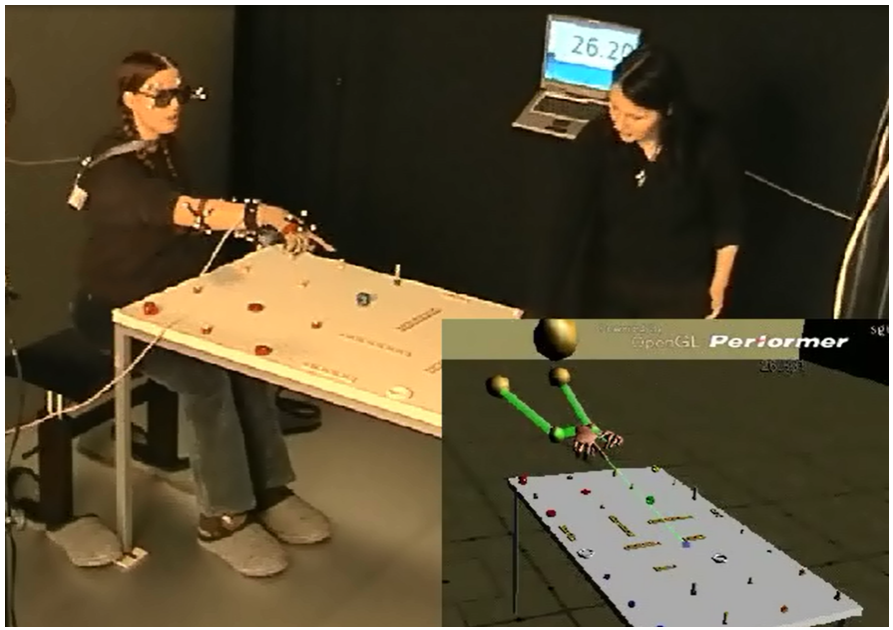


SaGA = (Bielefeld) Speech and Gesture Alignment Corpus

Look!

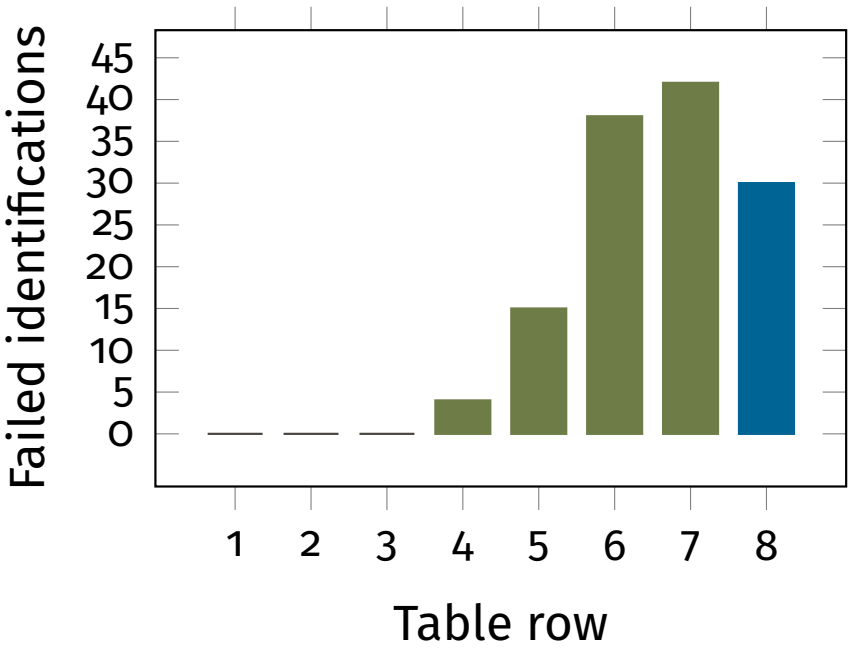
- ‘Look! [] All the dogs are barking.’
- According to RTT, the pointing gesture can point to a set of dogs, not to a property of set (of dogs).
- According to direct reference views (Kaplan, 1989) such a sentence is true if the entity provided by the pointing gesture is part of the denotation of barking things [NB: Kaplan does not deal with pluralities, but intuitively clear enough]
- But what does ‘entity provided by the pointing gesture’ mean? ➔ let us ask experimental pragmatics studies

Direct Reference?



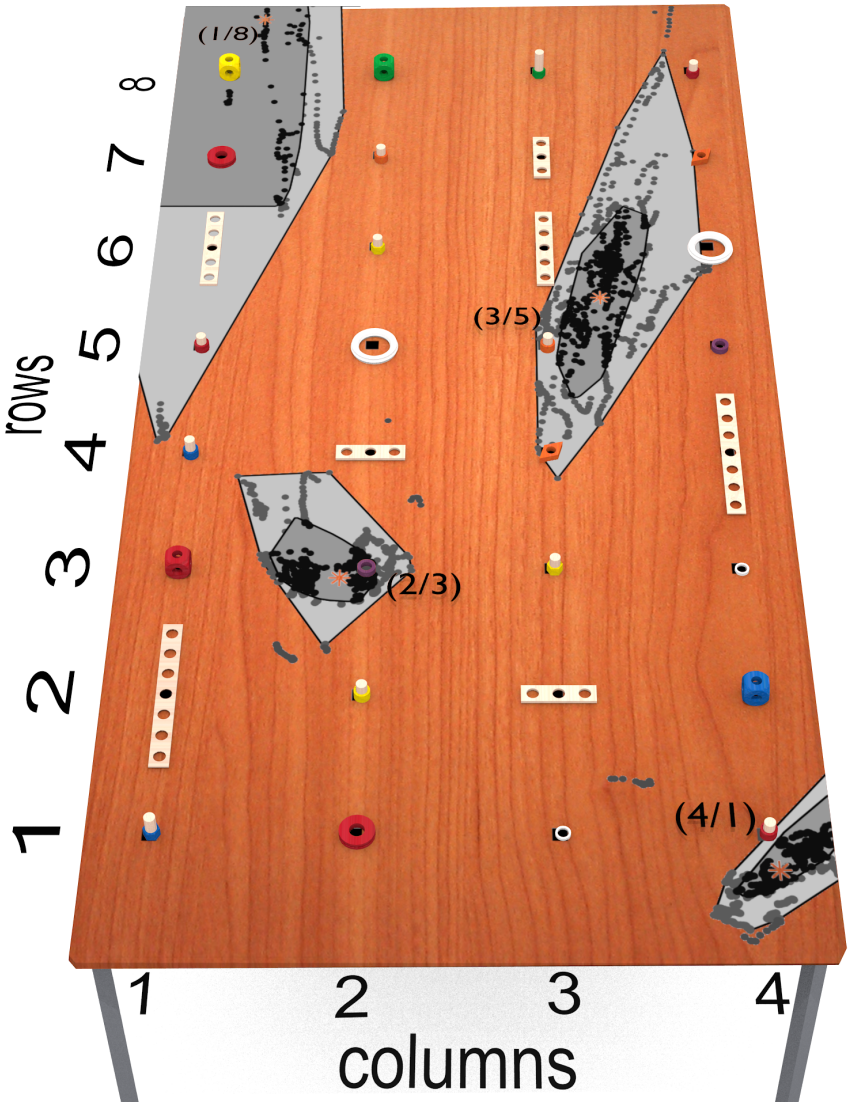
- *Experimental pragmatics* study (Kranstedt et al., 2006; Lücking, Pfeiffer and Rieser, 2015).
- Two runs: with speech and without speech.
- *Tracking of pointer*: simulate and 'measure' pointing.

Identification Failures



- For the addressee, the identifying force of pointings ceases in distal area.
- Note: decrease in row 8 due to ‘gestural hyperbole’.

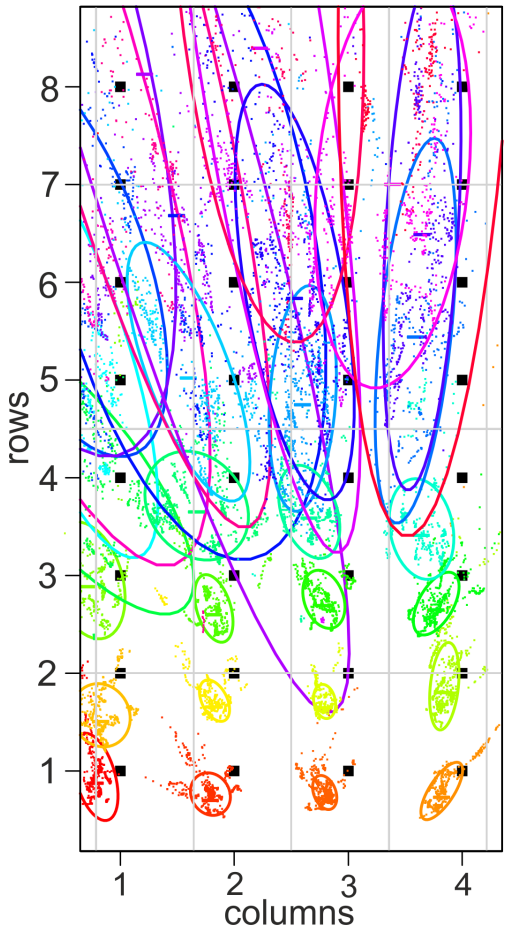
Measuring pointing



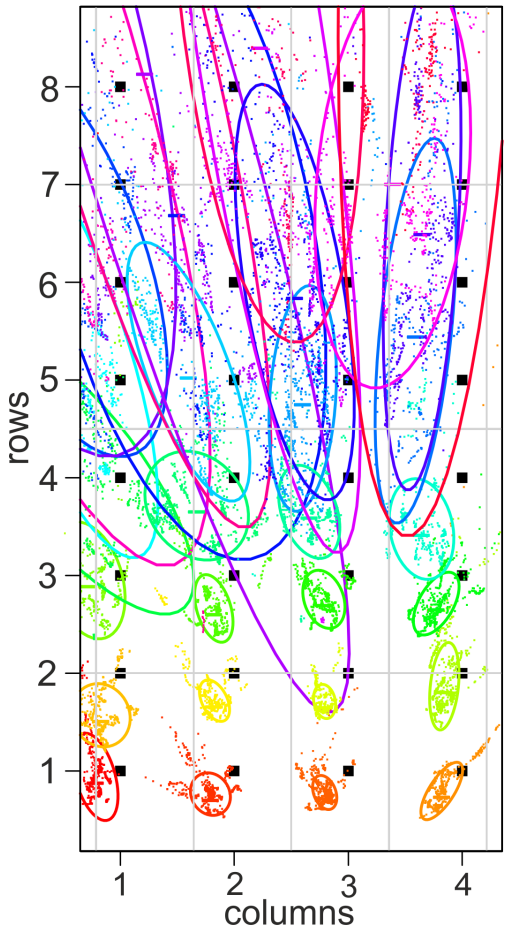
- Bagplots showing simulated pointing beams for four selected objects.
- Pointings usually do not hit their target.

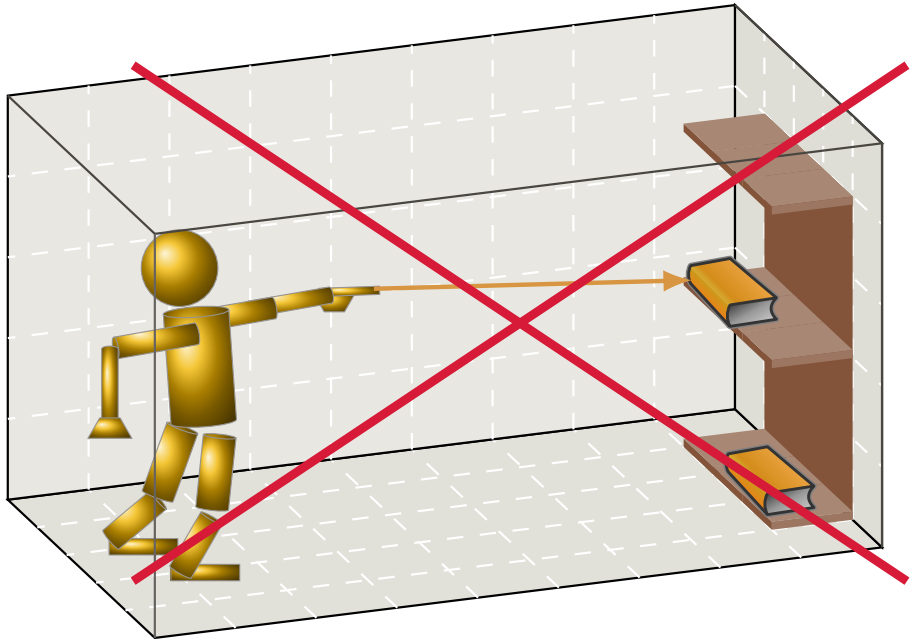
→ Demonstrative reference rests on a *pre-semantic pragmatic inference*: *take the object that is close(st) to the idealized pointing beam.*

Pointing Cone

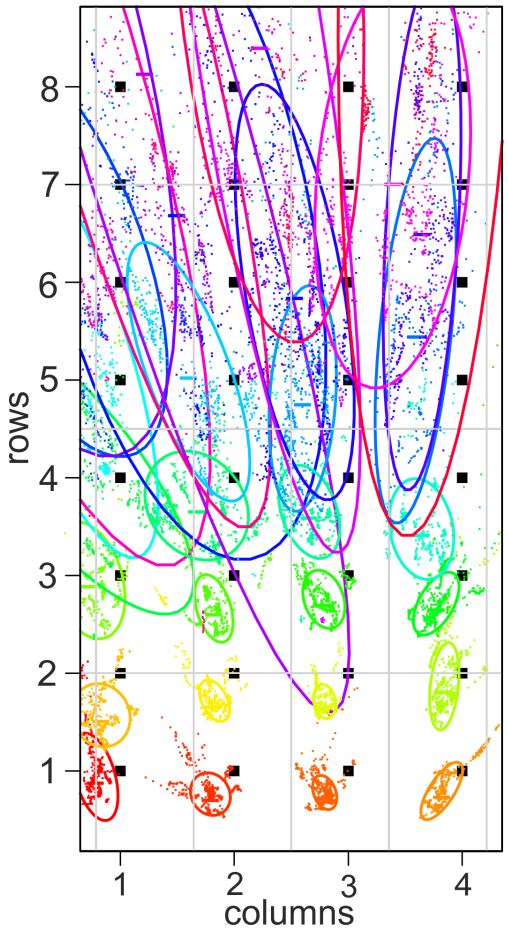


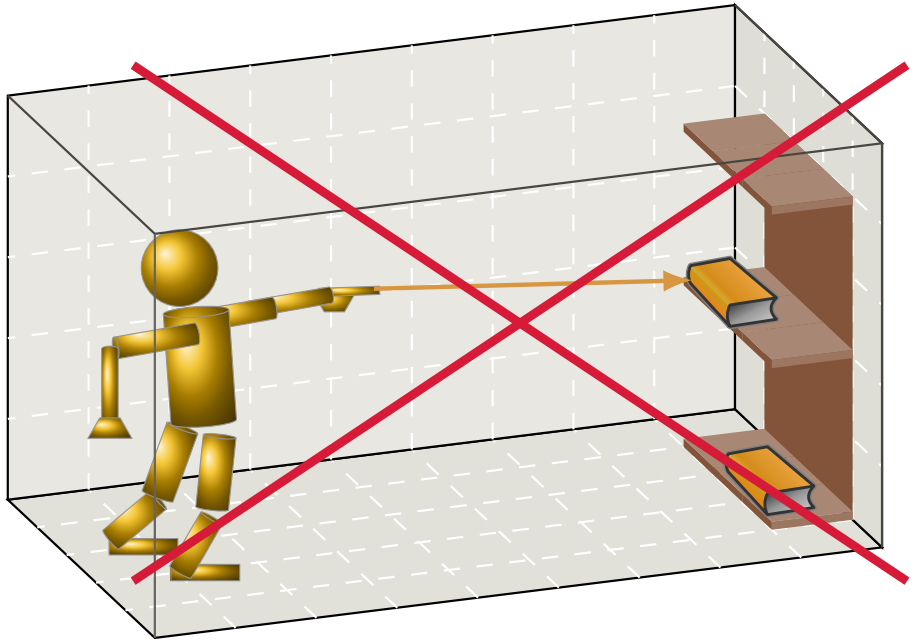
Pointing Cone

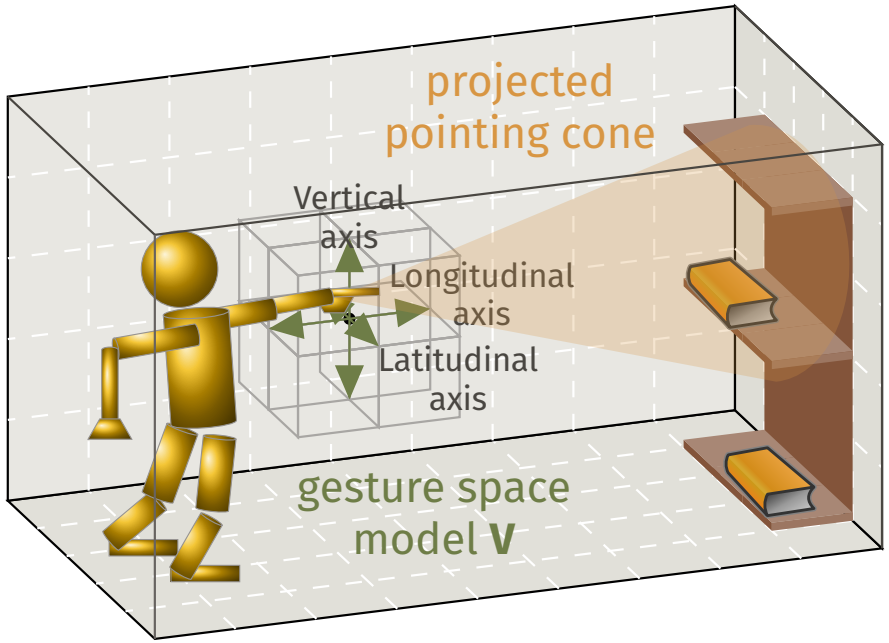




Pointing Cone

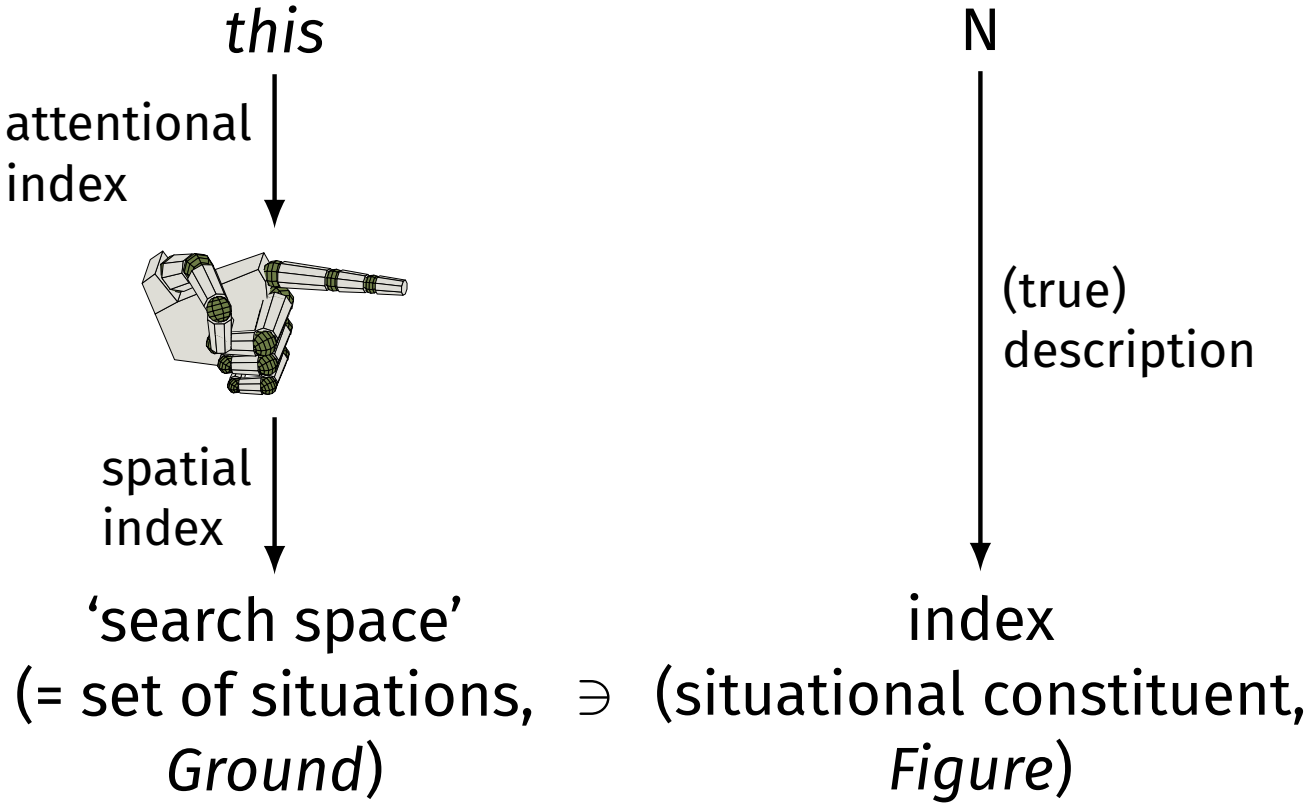




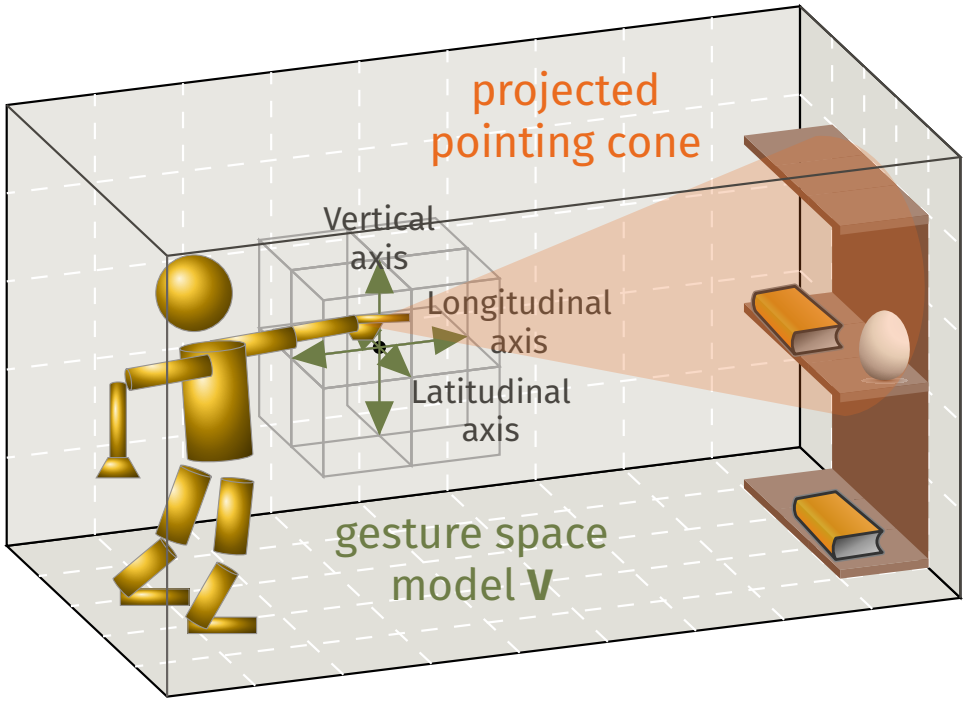


New proposal: figure-ground model

From reference to attention (Lücking, 2022)



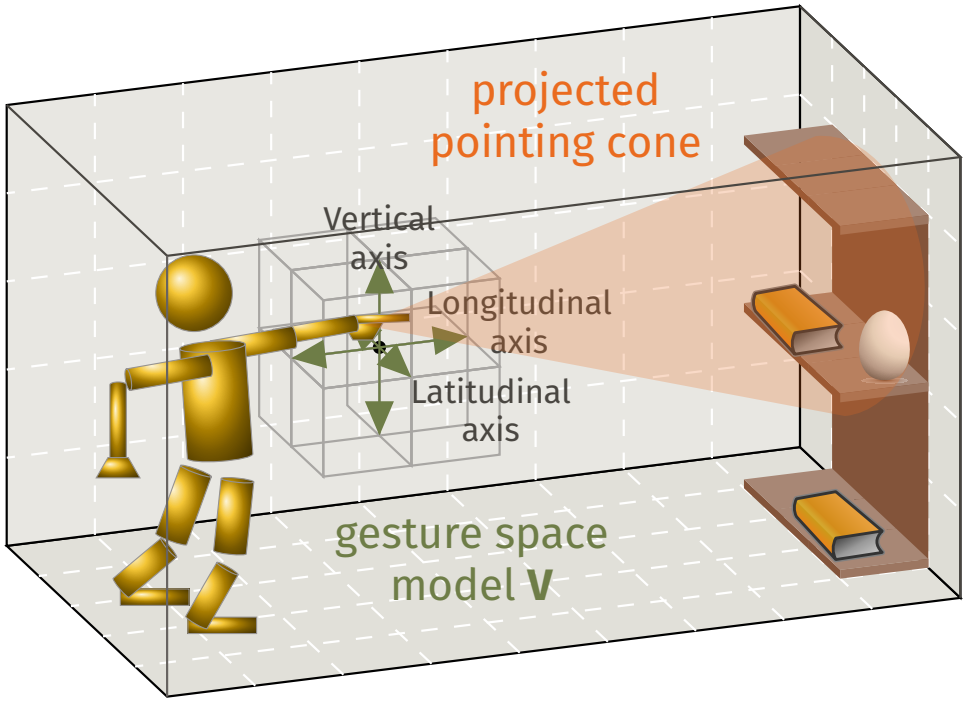
Spatial Semantics



Spatial Semantics:

Demonstrations *constrain situation variables.*

Spatial Semantics

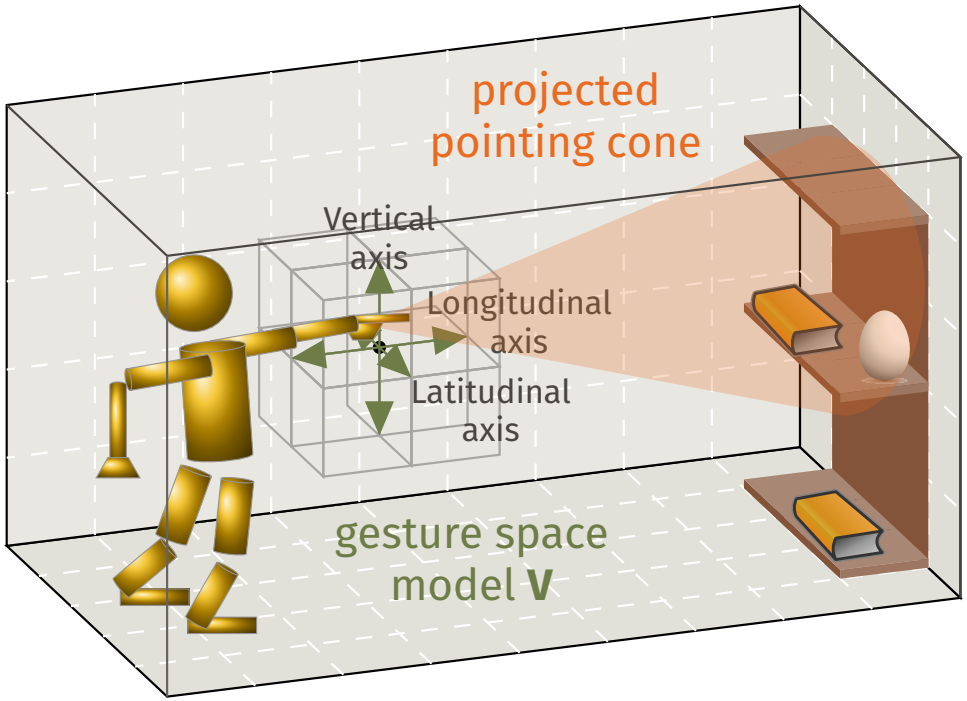


Spatial Semantics:

Demonstrations *constrain situation variables.*

- Pointing's *character* at u:
  $[\![\text{☞}]\!]^u = \lambda s.\, \text{region}(s) \cap \text{cone}(\text{☞})(u) \mapsto \text{relmax}$
  In short: ☞$(s) \mapsto \max_i$

Spatial Semantics



Spatial Semantics:

Demonstrations *constrain situation variables.*

- Pointing's *character* at u:
  $[\![\text{👉}]\!]^u = \lambda s.\, \text{region}(s) \cap \text{cone}(\text{👉})(u) \mapsto \text{relmax}$
  In short: $\text{👉}(s) \mapsto \max_i$
- *This[👉] book is great*:
  $\lambda s.\iota x x$ is a book in $s' \,\&\, \text{👉}(s') \mapsto \max_i$ is great in s.

Processing semantics of DemNPs

The dynamic semantics of DemNPs in dialog is governed by the retrieval question: 'Where to find the referent?' (Lücking, 2018)

Processing instructions for DemNPs

1. If there is a demonstration act, then the DemNP contributes to dgb-params and is witness-loaded in the focus of attention (*via* pointing cone).
2. If there is no demonstration, but a repetition of a constituent, the DemNP is interpreted anaphorically (also in dgb-params).
3. Otherwise, the DemNP contributes to q-params (but not to VisSit).

Lexical resource for pointing I

- The pointing device gives rise to a direction vector which indicates the direction into which the addressee of the pointing should turn its attention.
- $$\begin{bmatrix} \text{shape} : \texttt{pointing} \\ \text{dir=Vector(shape)} : \textit{Direction} \\ \text{dgb-params} : \begin{bmatrix} \text{spkr} & : \textit{Ind} \\ \text{addr} & : \textit{Ind} \\ \text{utt-time} & : \textit{Time} \\ \text{c-utt} & : \text{addressing(spkr,addr,utt-time,shape)} \end{bmatrix} \\ \text{content = DirectAttention(spkr,addr,dir))} : \textit{IllocProp} \end{bmatrix}$$
- triggers: Visual situation update (cf. Lect. 1)

Lexical resource for pointing II

- Visual situation update

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

- Interaction with exophoric demonstrative *this*:

Lexical resource for pointing III

- $$\left[\begin{array}{l} \text{phon=}\textit{this} : \textit{Phon} \\ \text{dgb-params} : \left[\begin{array}{ll} \text{spkr} & : \textit{Ind} \\ \text{addr} & : \textit{Ind} \\ \text{utt-time} & : \textit{Time} \\ \text{c-utt} & : \text{addressing(spkr,addr,utt-time,phon)} \\ \text{o} & : \textit{Ind} \\ \text{VisSit} & = \left[\text{InAttention} : \textit{Dir}\right] : \textit{RecType} \end{array}\right] \\ \text{cont=in(o,VisSit.InAttention)} : \textit{RecType} \end{array}\right]$$
- In sum: cognitively oriented, interactive modeling of spatial Figure–Ground model of deictic reference.

From reference to attention?

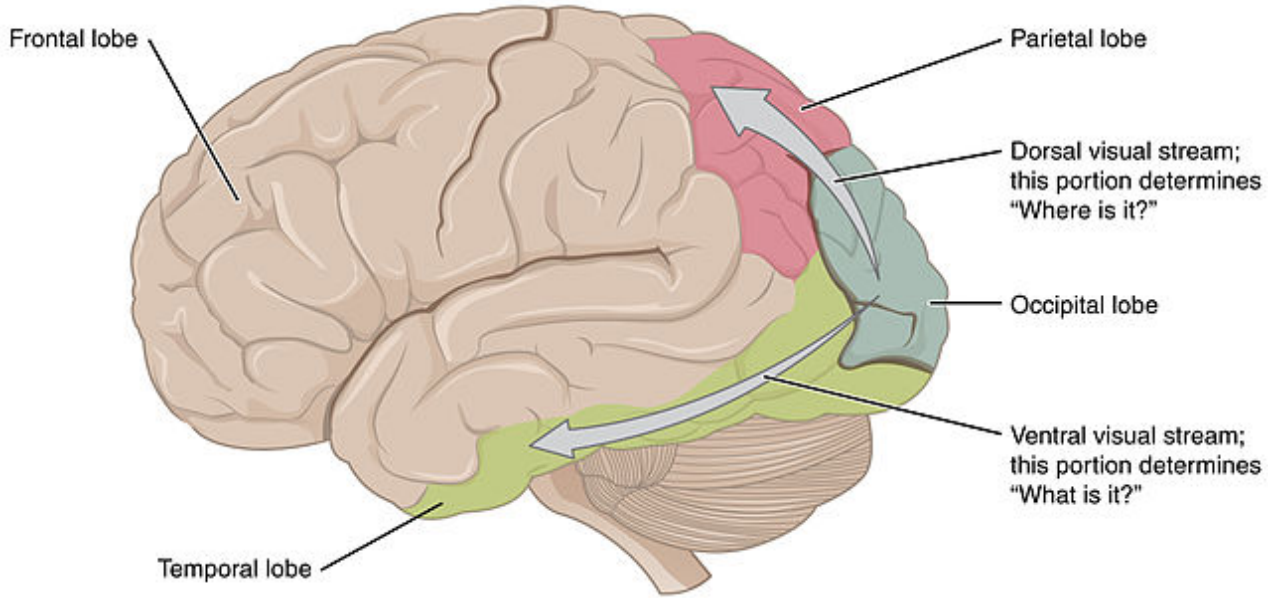
In some parts of *Conversation Analysis* (CA) attention seems to be derived from reference:

- '[…] a speaker introduces a new object by pointing at it and establishes the joint attention of the co-participants towards it' (Mondada, 2014, p. 95)
- 'In perhaps its barest form, referring consists of literally pointing to something in order for two people to share attention on that thing […]' (Enfield, 2013, p. 433)

From attention to reference?

- We conjecture that the mechanism for deictic reference is to be deduced from shared attention—not the other way round.
- Establishing pragmatic reference—that is filling the value of *InAttention* within the addressee's VisSit—is brought about by combining the ventral and dorsal processing streams (Connor and Knierim, 2017) such that an object becomes the unit of attention from a focused perceptual scene/direction (Scholl, 2001).
- Computationally, deictic reference is modeled in terms of a spatial semantics; procedurally, it employs two pathways of visual processing.

Ventral and dorsal pathway



Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013.
Author: OpenStax College, CC BY 3.0
https://commons.wikimedia.org/wiki/File:1424_Visual_Streams.jpg

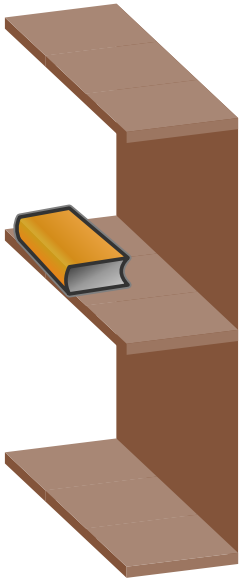
Uses of pointing gestures: deictic

© A. Lücking

‘Can you jump over this spout?’

locating referent

Uses of pointing gestures: deferred

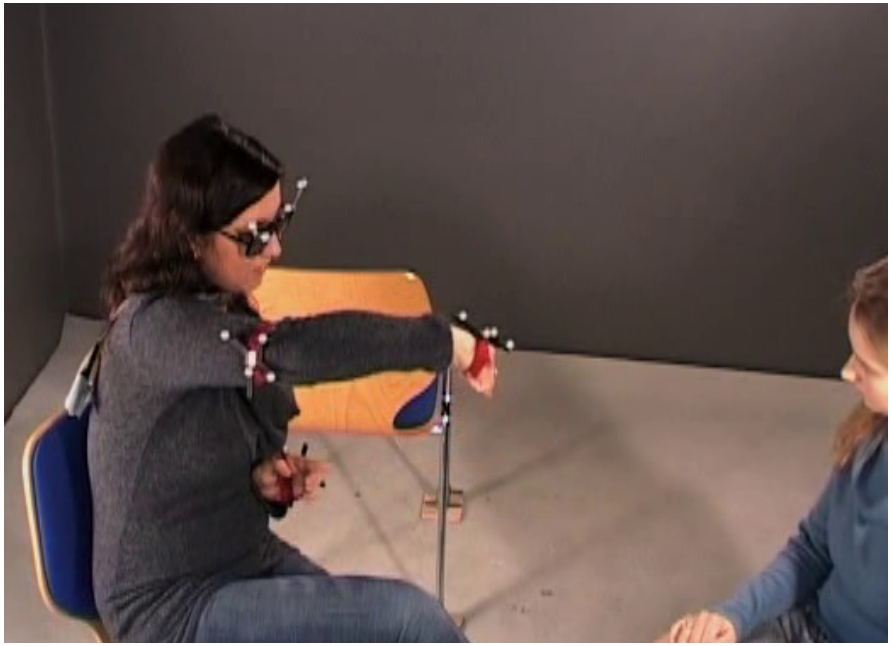


metonymic relation

‘This is my favourite author.’

(Nunberg:1993; Clark:1996)

Uses of pointing gestures: spatial proxy



'then you do not exit here [*index finger downwards*] (but there).'

(taken from SaGA V9, 6:56 (Lücking, Bergmann et al., 2010))

also called *abstract deixis* (McNeill, Cassell and Levy, 1993); projection from gesture space into described situation (cf. function $\vec{v}$ of Lascarides and Stone (2009))

Pointing at addressee

(context: *F [on the right] recaps route direction, hesitates*)

F: da steht die (.) die / T: there is the the
R: die SKULptur ((pointing at F)) / T: the sculpture

F: die skulptur drauf / T: the sculpture on top

(SaGA V5, 13:58)

Pointing at addressee

(context: *F [on the right] recaps route direction, hesitates*)

F: da steht die (.) die / T: there is the the
R: die SKULptur ((pointing at F)) / T: the sculpture

F: die skulptur drauf / T: the sculpture on top

(SaGA V5, 13:58)

R is pointing at the addressee (F), but:

- not locating addressee *F*
- no metonymic relation between *F* and the sculpture
- no spatial projection from *F*

→ what to do with the pointing gesture?

Rude pointing

‘Man zeigt nicht mit nacktem Finger auf angezogene Leute!’

(*It is bad manners to point at dressed people with naked fingers!*)

Informal analysis

Context of example:

- *F* recaps a route direction he got from *R*
- *F* has difficulties to recall a certain landmark
- *R* jumps in and supplies the landmark (i.e. ‘sculpture’)

→ the gesture emphasizes known material

Shared information gestures …

‘[…] mark material that the addressee probably already knows—information that is part of their common ground. They mean, essentially, “As you know”.’ (Bavelas et al., 1992, p. 397)

cf.: *marker of common ground* (Holler, 2010)

Lexicalizing CG pointing

Using KoS, the informal analysis of *common ground pointing* or *shared information gesture* can be made more precise in the following way:

$$\begin{bmatrix} \text{Preconds} : \begin{bmatrix} \text{Pending} : \textit{LocProp} \\ \text{u} : \textit{sign} \\ \text{c1} : \text{In(u,Pending.constits)} \end{bmatrix} \\ \text{Effects} : \begin{bmatrix} \text{R} : \textit{Rel} \\ \text{p} = \text{R(c)} : \textit{Prop} \\ \text{c} = \text{Preconds.u.cont} \\ \text{c1} : \text{In(FACTS,p)} \end{bmatrix} \end{bmatrix}$$

Further examples

F: ok_nochmal beim anfang dieses <<pointing at R> mit den säulen scheint ja irgendwie was komplizierter zu sein ja? (-)>

ok back to the start, the thing (*CG pointing*) with the pillars seems to be a bit more complicated, isn't it? (SaGA V2, 9:16)

F: auf jeden fall (.) DANN ((pointing at R)) muss ich in den park gehen?

anyhow, then (*CG pointing*) I have to go into the park?

(SaGA V4, 9:43)

Corpus survey

Survey of six SaGA dialogues: 13 instances of CG pointing.

But also other classes:

- UTT (*utterance anaphora*), 20
- SCTM (*something's coming to mind*), 9
- GrabTurn, 2

Utt (utterance anaphora)

indicating a DR of the actual utterance (difference to CG, which relates to grounded DR); occurs with topic (DR) introduction, affirmation of utterance of the other interlocutor, request clarification, or corrections; formally pointing at R/F, or index finger raising

R: °hh und dann kommen halt äh (-) die ((pointing at F)) BÄUme / and then there will just eh be the (*UTT pointing*) trees

(SaGA V2, 7:30)

SCTM (something's coming to mind)

pointing gesture associated with having an idea or recollection (in this case it is also CG); usually affiliated to expressives

R: da gehst du rein (-) °h da kommt n SEE: /
there you enter, and there is a lake
R: ah gut ((pointing at F)) (.) ich glaub

es kam doch erst der park
well (*SCTM pointing*) I guess there was the park first

(SaGA V4, 5:23)

SCTM DISCOURSE MEANING

$$\begin{bmatrix} \text{Preconds}: \begin{bmatrix} \text{spkr} : \textit{Ind} \\ \text{addr} : \textit{Ind} \\ \text{Pending.cont} : \textit{IllocProp} \\ \text{q} : \textit{Question} \\ \text{c1} : \text{About(Pending.cont,q)} \end{bmatrix} \\ \text{Effects} : \begin{bmatrix} \text{spkr} = \text{pre.spkr} : \textit{Ind} \\ \text{addr} = \text{pre.addr} : \textit{Ind} \\ \text{Pending.cont} : \textit{IllocProp} \\ \text{c2}: \neg\text{About(Pending.cont,Preconds.q)} \end{bmatrix} \end{bmatrix}$$

$\approx$ 'actual utterance pertains to a different question than the previous one'

Grab turn

usually index finger raising; affiliated to turn-taking expressions

R: du bleibst auf jeden fall auf der straße wo du bist und gehst geradeaus °h / in any case you stay on the street where you are and go straight ahead

F: <<index raised, repeated>ich frage nochmal kurz was nach> (.) also ähm / I have abrief clarification request ehm

(SaGA V4, 4:28)

GrabTurn discourse meaning

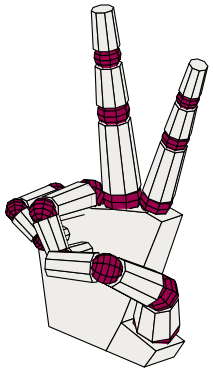
$$\begin{bmatrix} \text{Preconds} : \begin{bmatrix} \text{spkr} : \mathit{Ind} \\ \text{addr} : \mathit{Ind} \end{bmatrix} \\ \text{Effects} : \begin{bmatrix} \text{spkr} = \text{pre.addr} : \mathit{Ind} \\ \text{addr} = \text{pre.spkr} : \mathit{Ind} \end{bmatrix} \end{bmatrix}$$

- speaker change
- can be realised by finger-raising instead of addressee pointing

Conclusions on pointing

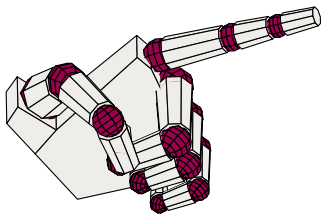
- Even ‘rude’ pointings have a discourse meaning
- The significance of pointing gestures not only consists in locating referents, but also in controlling the addressee’s attention and her view of the status of these referents in the incrementally emergent context

Kinds of gestures



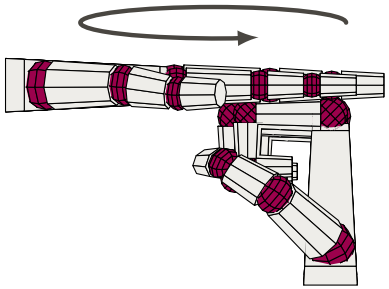
Emblems

lexicalized



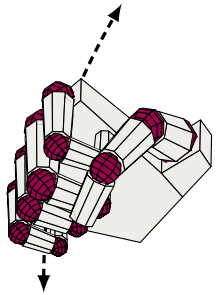
Deictic gestures

pointing



Iconic gestures

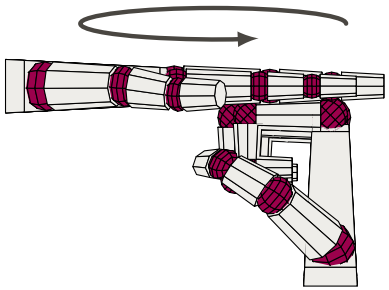
resemblance



Beats

rhythmic

Kinds of gestures



Iconic gestures

resemblance

Iconics: depiction methods

Drawing

draw outline

Shaping

form volume

Posturing

proxy

Placing

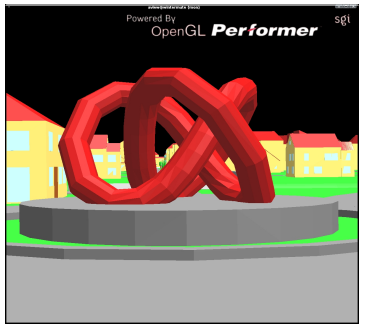
locate

…and more …

Why looking at (iconic) gestures? Image sources: SaGA/Lücking 2013

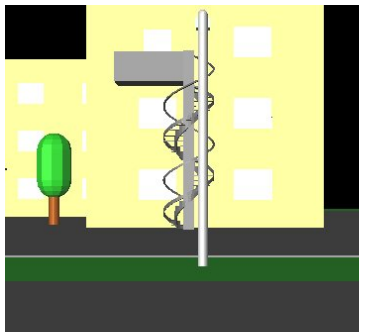
‘die Skulptur die hat ’n BETONsockel’

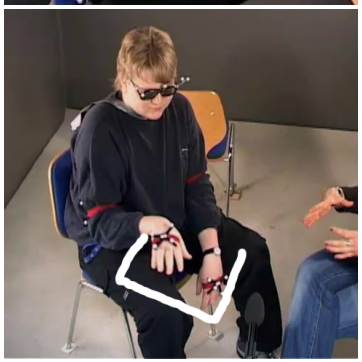
‘the sculpture it has a concrete base’



‘Ich glaube das sollen TREPPEN sein’

‘I think that should be staircases’





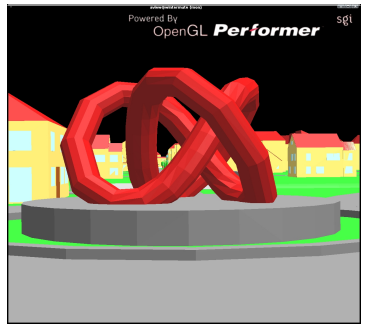
‘dann ist das Haus halt so’

‘then the house is like this’

Modifier + *good continuation*

‘die Skulptur die hat ’n BETONsockel’

‘the sculpture it has a concrete base’

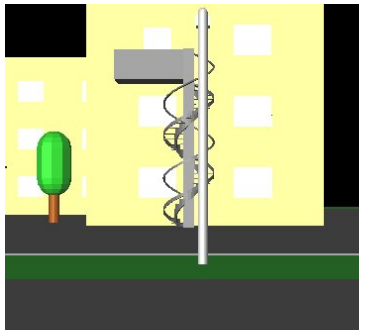


- Gestures modifies noun like an adjective.
- Gesture represents somthing like *cylindric*.
- But gesture performs only half of a cylinder $\rightarrow$ gestalt principle *good continuation*

Hyponym

‘Ich glaube das sollen TREPPEN sein’

‘I think that should be staircases’

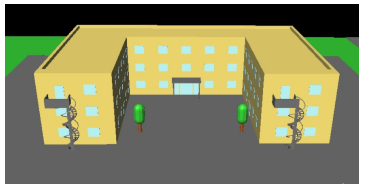


- The verbal description speaks of staircases.
- The gesture depicts a spiral.
- The composite utterance is about circular staircases, which is a *hyponym* of “staircases”.

Demonstration

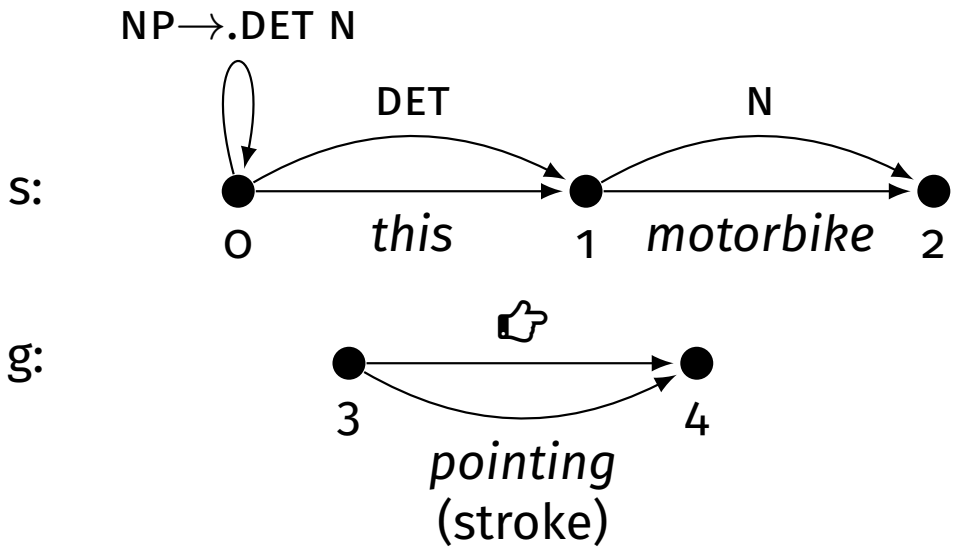
‘dann ist das Haus halt so’

‘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’.

Multimodal chart parser



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

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

References I

- Bavelas, Janet B. et al. (1992). 'Interactive Gestures'. In: *Discourse Processes* 15.4, pp. 469–489. DOI: 10.1080/01638539209544823.
- Connor, Charles E. and James J. Knierim (2017). 'Integration of objects and space in perception and memory'. In: *Nature Neuroscience* 20.11, pp. 1493–1503.
- Enfield, Nick J. (2013). 'Reference in Conversation'. In: ed. by Jack Sidnell and Tanya Stivers, pp. 433–454.
- Holler, Judith (2010). 'Speakers' Use of Interactive Gestures as Markers of Common Ground'. In: *Proceedings of Gesture Workshop 2009*. Ed. by Stefan Kopp and Ipke Wachsmuth. Lecture Notes in Artificial Intelligence 5934. Berlin and Heidelberg: Springer, pp. 11–22.

References II

- 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.
- Kaplan, David (1989). 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals'. In: *Themes from Kaplan*. Ed. by J. Almog et al. An earlier unpublished version exists as a UCLA Ms from ca. 1977. New York: Oxford University Press, pp. 481–614.
- Kranstedt, Alfred et al. (2006). 'Meaning and reconstructing pointing in visual contexts'. In: *Proc. of the 10th Workshop on the Semantics and Pragmatics of Dialogue*. Potsdam, pp. 82–89.

References III

- Lascarides, Alex and Matthew Stone (2009). 'A Formal Semantic Analysis of Gesture'. In: *Journal of Semantics* 26.4, pp. 393–449. DOI: 10.1093/jos/ffp004.
- Lücking, Andy (2018). 'Witness-loaded and Witness-free Demonstratives'. In: *Atypical Demonstratives: Syntax, Semantics and Pragmatics* 568, p. 255.
- — (2022). 'Aspects of Multimodal Communication'. Habilitation. Université Paris Cité.
- Lücking, Andy, Kirsten Bergmann et al. (2010). 'The Bielefeld Speech and Gesture Alignment Corpus (SaGA)'. In: *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. LREC 2010. 7th International Conference for Language Resources and Evaluation. Malta, pp. 92–98. DOI: 10.13140/2.1.4216.1922.

References IV

- Lücking, Andy, T. Pfeiffer and Hannes Rieser (2015). 'Pointing and reference reconsidered'. In: *Journal of Pragmatics* 77.2, pp. 56–79.
- McNeill, David (1992). *Hand and Mind – What Gestures Reveal about Thought*. Chicago: Chicago University Press.
- McNeill, David, Justine Cassell and Elena T. Levy (1993). 'Abstract deixis'. In: *Semiotica* 95.1-2, pp. 5–19. DOI: 10.1515/semi.1993.95.1-2.5.
- Mondada, Lorenza (2014). 'Pointing, talk, and the bodies. Essays in honor of Adam Kendon'. In: *From gesture in conversation to visible action as utterance*. Ed. by Mandana Seyfeddinipur and Marianne Gullberg. Amsterdam and Philadelphia: John Benjamins, pp. 95–124.

References V

- Müller, Cornelia (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Körper – Kultur – Kommunikation 1. Zugl. Diss. FU Berlin (1996). Berlin: Berlin Verlag.
- Scholl, Brian J. (2001). 'Objects and Attention: The State of the Art'. In: *Cognition* 80.1-2, pp. 1–46. DOI: 10.1016/S0010-0277(00)00152-9.
- Streeck, Jürgen (2008). 'Depicting by Gesture'. In: *Gesture* 8.3, pp. 285–301. DOI: 10.1075/gest.8.3.02str.