

# I thought pointing is rude: A dialogue-semantic analysis of pointing at the addressee<sup>1</sup>

Jonathan GINZBURG — CNRS, Université de Paris, Laboratoire de Linguistique Formelle  
Andy LÜCKING — CNRS, Université de Paris, Laboratoire de Linguistique Formelle; Text  
Technology Lab, Goethe University Frankfurt

**Abstract.** A pilot corpus study on the use of pointing gestures in dialogue yielded 44 instances of pointing at the addressee. In none of these instances is the addressee the gesture’s referent, however. Rather, such discourse pointings are bound up with dialogue management: they control the addressee’s attention and her view of the status of these referents in the incrementally emergent context. We distinguish four classes of addressee pointings, descriptively glossed *utterance anaphora*, *common ground*, *something’s coming to mind*, and *grab turn*. We exemplify each class by means of empirical data and provide a dialogue semantics analysis. In this way, we extend the taxonomy of uses of pointings currently discussed in semantics and argue that the linguistic competence revealed by discourse pointings is inherently dialogical, adding evidence for extending the domain of grammar from well-formedness and truth conditions to include micro-level elements of conversational interaction.

**Keywords:** pointing, interactive gesture, discourse, common ground, utterance anaphora.

## 1. Introduction

That (manual) gestures belong in the grammar is by now well accepted (cf. Alahverdzhieva et al., 2017; Fricke, 2012; Lücking, 2013 from the view of formal grammar theory and Kendon, 1980; McNeill, 1985 from the field of communication psychology). Of all types of manual gestures, deictic or pointing gestures have received the most attention in semantics. This predominance is very likely due to the idea that pointing gestures are bound up with reference, a key notion in semantics and pragmatics. The standard account in this respect is still the incompleteness and direct reference view of Kaplan (1989), where pointing gestures act as demonstratum-donating means for otherwise referentially deficient demonstratives.<sup>2</sup> Closer to modern grammar-based approaches is the neo-Peirce-Wittgenstein-Quine view that both pointing gestures *and* their relation to language belong to the object language (Rieser, 2004).<sup>3</sup> These

<sup>1</sup>Many thanks to the insightful comments of three reviewers for the Gestures@SuB25 workshop, as well as to the virtual but very alert and warm audience at the workshop. We acknowledge support by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris — AN-1-IDE-0001. We also acknowledge a senior fellowship from the Institut Universitaire de France to Ginzburg.

<sup>2</sup>While Kaplan (1978) developed a particular technical account, the general idea can be traced back throughout the philosophical literature – consider, for instance, Frege (1918: 64): “In allen solchen Fällen [i.e., demonstrative expressions; the authors] ist der bloße Wortlaut, wie er schriftlich festgehalten werden kann, nicht der vollständige Ausdruck des Gedankens, sondern man bedarf zu dessen richtiger Auffassung noch der Kenntnis gewisser das Sprechen begleitender Umstände, die dabei als Mittel des Gedankenausdrucks benutzt werden. Dazu können auch Fingerzeige, Handbewegungen, Blicke gehören.” (*In all such cases [i.e., demonstrative expressions; the authors] the mere wording, as it can be recorded in writing, is not the complete expression of the thought, but for its correct understanding one still needs the knowledge of certain circumstances accompanying the speech, which are used thereby as means of the thought expression. To it also finger points, hand movements, looks can belong.*)

<sup>3</sup>Most notably, Rieser developed the notion of *region pointing*, where a deictic gesture indicates the spatial position of the value of a verbal reference marker. This view has been empirically tested and further developed by Lücking et al. (2015). A framework according to which speech and gesture occupy different informational channels and

referent-identification accounts focus on concrete index-finger pointings such as that in (1):

- (1) *Can you jump over this spout?*



© A. Lücking

The demonstration act in (1) is concrete since the thing indicated (pointed at) and the thing referred to (described) are identical.<sup>4</sup> However, this need not to be the case, as is evinced in cases of deferred reference (Nunberg, 1993), which exploit a metonymic relation between index and referent. An example is given in (2), taken from Clark (1996: 168), where the book is indicated, but its author is referred to:

- (2) *That man was a friend of mine.*



© A. Lücking, book icon from T. Tantau (L<sup>A</sup>T<sub>E</sub>X beamer class)

Deferred reference still involves a concrete index. But pointing gestures can also be used in an abstract way (McNeill et al., 1993). An example is given in (3), where the speaker virtually draws a map in gesture space and points to virtual locations on this map.

- (3) *Then you do not exit here, but there.* [Translated from German original]



Taken from dialogue V9, 6:56 (Lücking et al., 2010)

In formal semantics abstract deixis is modeled as a mathematical projection from gesture space into the described situation (Lascarides and Stone, 2009: 408).

In sum, the received view on pointing gestures in semantics/pragmatics is that they can be used in one of three ways:

---

cohere in a spatial model has been formulated by Kühnlein (1999). Further evidence has been provided by Grosz (2019) in terms of different kinds of pronoun-aligned pointing gestures (in addition to intonation).

<sup>4</sup>And since the gesture is produced in the context of a verbal demonstrative it contributes at-issue information (Ebert and Ebert, 2016).

- (4) a. *deictically* (i.e., locating the semantic value of a discourse referent (DR) in the perceptible environment); concrete deixis is prototypically affiliated to demonstratives like “this” or “over there” in speech;
- b. as *spatial proxy* (projection from gesture space to real world). Abstract deixis is well-known from gestures studies where a location in gesture space represents a location from the events talked about (Fricke, 2007);<sup>5</sup>
- c. in *deferred reference* (exploiting a metonymical relation between demonstratum and referent).

In what follows we argue that data from pointing gestures in natural interactions – most notably, pointing at the addressee – force us to extend this semantic taxonomy. Moreover, in order to provide a precise, semantic analysis of addressee pointing – as will become clear in its informal discussion in Sec. 2 – a detailed model of utterance context is needed. Such a model has been developed within KoS (Ginzburg, 2012), which is briefly introduced alongside some of its logical underpinnings in Sec. 3. KoS provides the analytical means to spell out the semantic significance of various forms of addressee pointing; the corresponding lexical entries are given in Sec. 4. Understanding discourse pointing involves attention managing and how it marks the grounding status of discourse referents in an incrementally unfolding conversational context. Accordingly, a grammar that includes addressee pointings must be intrinsically *dialogical* – a demand that probably applies to grammar in general (Ginzburg and Poesio, 2016).

Such an analysis is required all the more so because pointing at people is usually considered to be rude: in many cultures one finds rules of etiquette along the lines of *It is bad manners to point at dressed people with naked fingers!* For instance, the entry of the verb *point* in Harrap’s English Dictionary explicitly mentions that “You mustn’t point at people like that” (Higgleton and Seaton, 1996: 732). Accordingly, an explanation of the fact that discourse pointing at the addressee is nonetheless not offensive is needed.

## 2. Corpus study

The empirical evidence for the class of gestures called *discourse pointing* or *pointing at the addressee* is based on a corpus study. We surveyed six route-giving dialogues from the SaGA (*Speech and Gesture Alignment*) corpus (Lücking et al., 2010). A summary of this SaGA subset is collected in Tab. 1. It contains 2,192 gestures in total. The gestures have been assigned to the gesture classes *beat*, *deictic* (i.e. pointing), *discourse*, *iconic* and possible combinations thereof. Hand and arm movements that do not seem to constitute a gesture have been labeled *move* – see the annotation manual (Bergmann et al., 2014) for details. As can be seen in Tab. 1 the seventh part of the gestures observed in SaGA are deictic gestures, nearly the fifth part are discourse gestures. Within the six dialogues, we found 44 instances of discourse pointing. That is, about the seventeenth part of discourse or pointing gestures is addressee pointing. We looked through each of the dialogue recordings and collected and transcribed instances of discourse pointing (examples are presented below). Each occurrence was assigned to one of the following, bottom-up defined classes:

<sup>5</sup>Abstract deixis has already been discussed as “Deixis am Phantasma” by Bühler (1934: § 8).

	V2	V4	V5	V6	V7	V25	sum
beat	1	25	6	7	2	0	41
deictic	22	73	53	64	39	37	288
deictic-beat	1	9	8	7	4	2	31
discourse	43	74	107	210	10	13	457
iconic	27	185	344	154	124	57	891
iconic-beat	1	29	16	23	1	10	80
iconic-deictic	13	26	95	26	49	9	218
iconic-deictic-beat	0	6	3	2	0	0	11
move	5	77	25	26	15	23	171
unclear	3	0	0	0	1	0	4
sum	116	504	657	519	245	151	2,192

Table 1: Number of gestures (counted over both left and right hand/arm) in the six SaGA dialogues used in the corpus survey, summed up for both participants.

#### (5) Sub-kinds of discourse pointing

- a. “UTT” (*utterance anaphora*): indicating a discourse referent (DR) of the actual utterance (difference to CG); occurs with topic (DR) introduction, affirmation of utterance of the other interlocutor, clarification or information requests, or corrections; we found formal variation between pointing at interlocutor and index finger raising; corpus frequency: 20 (topic intro: 3, affirm: 12, self-correction: 3, request: 2)
- b. “CG” (*common ground*): shared information pointing; indicating a DR which already has been grounded; corpus frequency: 13
- c. “SCTM” (*something’s coming to mind*): pointing gesture associated with having an idea or recollection (in this latter case it is also CG); often affiliated to expressives (e.g., *ah!*) in speech; corpus frequency: 9
- d. “GrabTurn”: often realized by index finger raising; affiliated to turn-taking expressions (*wait; I have a question*); corpus frequency: 2

An instance of CG (common ground pointing) is shown in (6) – more examples for each of the classes just introduced are given in Sec. 4. (6) is taken from SaGA dialogue V5, starting at video time 13:58. The transcription follows a minimal transcription according to the GAT-2 (*Gesprächsanalytisches Transkriptionssystem 2*) standard for spoken German discourse (Selting et al., 2009). We use “R” in order to indicate the route-giver and “F” to indicate the route-follower. The original German transcription is given first, followed by a free English translation right of a slash “/”.

- (6) *Context: F is recapitulating the route which has just described to him by R. Now he is trying to recall the landmark at a certain point of the route (turn 1). Due to his*

hesitation (“die (.) die”), R completes the utterance (turn 2) while discourse pointing at F. The completion was successful since it got accepted by F (turn 3).

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



- 3 F: die skulptur drauf / the sculpture on top

Neither F nor the sculpture talked about is a plausible candidate for a located object, so the referent identification use (use 4a) has to be excluded. This holds literally as well as projectively, since the space F occupies has not been assigned to some outer-world scene. That is, abstract deixis (use 4b), mapping a region of gesture space to the physical space of the described situation) has to be excluded, too. Likewise a deferred interpretation (use 4c) licensed by a metonymic relation between the sculpture and F is not available: a reasonable contiguity relation between the index (F) and the referent (the sculpture) is simply lacking. Hence, CG pointing cannot be subsumed within the classes in (4).

In order to provide an analysis we have to take a different route. Following functional analyses from gestures studies, we take CG gestures to be a kind of a *shared information gesture* (Bavelas et al., 1992), which can be construed as *markers of common ground* (Holler, 2010). What then is the CG gesture’s contribution? Its affiliated expression in turn 2 in (6) is an NP, *die Skulptur*, which, due to its definiteness, either has to be linked to an already familiar DR, or to be accommodated. The CG gesture disambiguates this interpretation fork by cancelling the accommodation branch and signaling that the DR is indeed part of the CG of the interlocutors. Concretely, (6) indicates for some constituent of the current utterance (a contextual parameter of CG’s meaning) that it is the constituent of a grounded proposition. A more precise characterization of this informal interpretation is given in Sec. 4. Before we can spell out our semantic formalization of the discourse pointing subclasses, we have to provide background about the framework in which this formalization is couched.

### 3. KoS/TTR

We formulate our account within the framework of KoS (Ginzburg, 1994; Ginzburg and Cooper, 2004; Larsson, 2002; Purver, 2006; Fernández, 2006; Ginzburg and Fernández, 2010; Ginzburg, 2012).<sup>6</sup> KoS is a theory that combines an approach to semantics inspired by situation semantics and dynamic semantics with a view of interaction influenced by Conversation Analysis. In KoS instead of a single context, analysis is formulated at a level of cognitive states, one per conversational participant. Each cognitive state consists of two ‘parts’, a private part and the dialogue gameboard that represents information that arises from publicized interactions and on which we focus here. The structure of the dialogue gameboard is given in (7) – the *spkr*, *addr* fields allow one to track turn ownership, *Facts* represents conversationally shared assumptions, *VisSit* keeps track of the visual situation including the focus of visual attention, *Pending* and *Moves* represent respectively moves that are in the process of being or have been grounded, *QUD* tracks

<sup>6</sup>KoS is a toponym – the name of an island in the Dodecanese archipelago – bearing a loose connection to *conversation-oriented semantics*.

the questions currently under discussion, and *mood* tracks certain emotive aspects, important for the analysis of non-verbal signals such as laughter, smiling, and frowning (Ginzburg et al., 2020). Of these contextual parameters at least one, *VisSit*, is probably never entirely identical across participants since distinct interlocutors do not share the same pair of eyes, and moreover much of the time interlocutors have each other as their focus of attention. Nonetheless, there are various devices most prominently perhaps pointing to effect alignment.

$$(7) \quad \text{DGBType} =_{\text{def}} \left[ \begin{array}{l} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{utt-time} : \text{Time} \\ \text{c-utt} : \text{addressing}(\text{spkr}, \text{addr}, \text{utt-time}) \\ \text{Facts} : \text{Set}(\text{Proposition}) \\ \text{VisSit} : \left[ \text{InAttention} : \text{Ind} \right] \\ \text{Pending} : \text{list}(\text{locutionary Proposition}) \\ \text{Moves} : \text{list}(\text{illocutionary Proposition}) \\ \text{QUD} : \text{poset}(\text{Question}) \\ \text{mood} : \text{Appraisal} \end{array} \right]$$

To understand better the specification in (7), we offer a short digression concerning the logical underpinnings of KoS. KoS is formulated within the framework of Type Theory with Records (TTR) (Cooper, 2005, 2012; Cooper and Ginzburg, 2015; Cooper, 2021). TTR is a model-theoretic descendant of the by and large proof theoretic Martin-Löf Type Theory (Ranta, 1994; Betarte and Tasistro, 1998) and of situation semantics (Barwise and Perry, 1983; Cooper and Poesio, 1994; Seligman and Moss, 1997; Ginzburg and Sag, 2000). For current purposes, the key notions of TTR are the notion of a *judgement* and the notion of a *record*.

**The typing judgement:**  $a : T$  classifying an object  $a$  as being of type  $T$ . Examples are given in (8). (8a) and (8b) involve *basic* “atomic” types IND(ividual) and TIME. In (8c)  $\text{run}(\text{arg1}_{\text{IND}} = b, \text{arg2}_{\text{TIME}} = t)$  is a *p(redicate)-type*, that arises by assigning the entities  $b, t$ , respectively to the argument roles of *run*;  $\text{arg1}_{\text{IND}}$  requires its fillers to be *of type* IND, whereas  $\text{arg2}_{\text{TIME}}$  requires its fillers to be *of type* TIME; we will usually notate such types as (8d). Ranta (1994) proposed that elements such as  $s$  in (8d) be viewed as events or situations.

- (8) a.  $b : \text{IND}$   
 b.  $t : \text{TIME}$   
 c.  $\text{run}(\text{arg1}_{\text{IND}} = b, \text{arg2}_{\text{TIME}} = t)$   
 d.  $s : \text{run}(b, t)$

**Records:** A record is a set of fields assigning entities to labels of the form (9a), partially ordered by a notion of *dependence* between the fields – dependent fields must follow fields on which their values depend. A concrete instance is exemplified in (9b). This is a record with four fields  $x$ ,  $e\text{-time}$ ,  $e\text{-loc}$ , and  $c_{\text{temp-at-in}}$  to which are assigned respectively a number, a time, a location, and a situation *sit1*; the example is further discussed in (11). Records are used here to model events and states, including utterances, and dialogue gameboards.

$$(9) \quad \text{a.} \quad \begin{bmatrix} l_1 & = & val_1 \\ l_2 & = & val_2 \\ \dots & & \\ l_n & = & val_n \end{bmatrix}$$

$$\text{b.} \quad \begin{bmatrix} x & = & -28 \\ \text{e-time} & = & 2\text{AM, Feb 17, 2019} \\ \text{e-loc} & = & \text{Nome} \\ c_{temp-at-in} & = & \text{sit1} \end{bmatrix}$$

**Record Types:** a record type is a record where each field represents a judgement rather than an assignment, as in (8a). The basic relationship between records and record types 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, as in (10b):

$$(10) \quad \text{a.} \quad \begin{bmatrix} l_1 & : & T_1 \\ l_2 & : & T_2 \\ \dots & & \\ l_n & : & T_n \end{bmatrix}$$

$$\text{b.} \quad \text{The record} \begin{bmatrix} l_1 & = & a_1 \\ l_2 & = & a_2 \\ \dots & & \\ l_n & = & a_n \end{bmatrix} \text{ is of type: } \begin{bmatrix} l_1 & : & T_1 \\ l_2 & : & T_2 \\ \dots & & \\ l_n & : & T_n \end{bmatrix}$$

iff  $a_1 : T_1, a_2 : T_2, \dots, a_n : T_n$

To exemplify this, (11a) is a possible type for (9b), assuming the conditions in (11b) hold. Record types are used to model utterance types (often referred to in formal grammar as *signs*) and to express rules of conversational interaction.

$$(11) \quad \text{a.} \quad \begin{bmatrix} x & : & \text{Ind} \\ \text{e-time} & : & \text{Time} \\ \text{e-loc} & : & \text{Loc} \\ c_{temp-at-in} & : & \text{temp\_at\_in}(\text{e-time}, \text{e-location}, x) \end{bmatrix}$$

$$\text{b.} \quad -28 : \text{Ind}; 2:00\text{AM, Feb 17, 2019} : \text{Time}; \text{Nome} : \text{Loc}; \text{sit1} : \text{temp\_at\_in}(2:00\text{AM, Feb 17, 2019, Nome, } -28)$$

Contextual reasoning will be important here in several ways. First, we characterize dialogue regularities (e.g., A's assertion  $p$  gives rise to the possibility that B accepts  $p$  or alternatively that B initiates discussion of the question  $p$ ?) in terms of *conversational rules*, mappings between two cognitive states the *precondition*s and the *effects*. Conversational rules can come in two flavours, rules that each interlocutor applies in the same way to their cognitive state (*participant neutral*) 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, 2021. Most of the conversational rules we will specify will be participant neutral, as exemplified in the rules given in (12):

- (12) a. Ask QUD-incrementation: given a question  $q$  and  $\text{ASK}(A,B,q)$  being the Latest-Move, one can update QUD with  $q$  as MaxQUD.

$$\left[ \begin{array}{l} \text{pre} \\ \text{effects} \end{array} : \left[ \begin{array}{l} q : \text{Question} \\ \text{LatestMove} = \text{Ask}(\text{spkr}, \text{addr}, q) : \text{IllocProp} \\ \text{QUD} = \langle q, \text{pre.QUD} \rangle : \text{poset}(\text{Question}) \end{array} \right] \right]$$

- b. QSPEC: this rule characterizes the contextual background of reactive queries and assertions – if  $q$  is MaxQUD, then subsequent to this either conversational participant may make a move constrained to be  $q$ -specific (i.e., either About or Influencing  $q$ ).

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

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 QUD and FACTS 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 (13a), 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. (13b) provides the specification for visual situation update rule:

- (13) 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:

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

The final logical notion we introduce is the situation semantics notion of an Austinian proposition (Barwise and Etchemendy, 1987). Deriving from Austin's (1950) theory of truth (a true assertion involving a situation token matching a situation type), they were originally proposed

to explicate assertions and relatedly beliefs. In TTR they are identified with records of the form (14a) whose truth conditions are defined in (14b):

- (14) a.  $\left[ \begin{array}{l} \text{sit} \quad = \quad s \\ \text{sit-type} \quad = \quad T \end{array} \right]$
- b. A proposition  $p = \left[ \begin{array}{l} \text{sit} \quad = \quad s_0 \\ \text{sit-type} \quad = \quad ST_0 \end{array} \right]$  is true iff  $s_0 : ST_0$

Subsequently, such propositions have been used in modelling utterance processing (Ginzburg, 2012). Ginzburg (2012) proposes that dialogue interaction is, to a large extent, structured by a series of branching points where an utterance is either grounded (Clark, 1996) or gives rise to clarification interaction or repair. Ginzburg (2012) shows that the specific conditions for grounding and possibilities for repair of an utterance  $u$  can be read off the *locutionary proposition* defined by  $u$  and a grammatical type  $T_u$ , intuitively the *sign* (in the Saussurean sense) associated with  $u$ . That the locutionary proposition involves the entire sign and not merely its semantic components is motivated, in part, by the fact that this enables the locutionary proposition to characterize the forms that are possible means to ground or request clarification about  $u$  and these exhibit significant syntactic and phonological parallelism with  $u$  (Ginzburg and Cooper, 2004). (15b) exemplifies lexical entries we will posit below for laughter and its ilk. Here it is a somewhat simplified lexical entry for the particle *mmh* used to by B to acknowledge understanding of a prior utterance by A. It has fields for phonological and syntactic types, as well as for the *contextual parameters* of the utterance (DGB-PARAMS) needed to resolve the content of an utterance of *mmh* on a given use. In this case the contextual parameters are an utterance token and the conversational participants:

- (15) a. A locutionary proposition  $\left[ \begin{array}{l} \text{sit} \quad = \quad u_0 \\ \text{sit-type} \quad = \quad T_{u_0} \end{array} \right]$  is true iff  $u_0 : T_{u_0}$ , in other words iff the sign fully classifies the utterance; otherwise, repair interaction ensues.
- b.  $\left[ \begin{array}{l} \text{phon} : \text{mmh} \\ \text{syncat} : \text{interjection} \\ \text{dgb-params} : \left[ \begin{array}{l} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{u} : \text{sign} \\ \text{c1} : \text{address(addr,spkr,u)} \end{array} \right] \\ \text{cont} = \text{Acknowledge}(u,\text{spkr}) : \text{IllocProp} \end{array} \right]$

## 4. Lexicalizing addressee pointing gestures

### 4.1. CG

We already encountered a CG example in (6) in Sec. 2. We left the discussion by observing that the CG gesture indicates for some constituent of the current utterance that it is the constituent of a common ground proposition. Having the tools from Sec. 3 at disposal we can make this interpretation precise since both the constituents of an utterance and FACTS (i.e. CG) are contextual parameters, among others. The working of CG pointing can be captured in terms of the lexical entry in (16). Concretely, it indicates of some sub-utterance that is a constituent of

(the maximally) pending utterance – an utterance still in the process of being grounded – that it fills an argument role of an already grounded proposition  $p$  ( $p$  is part of FACTS, see c2).

$$(16) \quad CG \mapsto \left[ \begin{array}{l} \text{Shape : PointType} \\ \text{dgb-params : } \left[ \begin{array}{l} \text{MaxPending : LocProp} \\ \text{u : sign} \\ \text{c1 : In(u,MaxPending.constits)} \\ \text{R : Rel} \\ \text{a : IND} \\ \text{p = R(a) : Prop} \\ \text{c2 : In(FACTS,p)} \end{array} \right] \\ \text{cont = } \left[ \text{c3 : =(u.cont,a)} \right] : \text{RecType} \end{array} \right]$$

Note that the lexical entries we provide here are simplified in that they abstract over different tiers or channels. They can be embedded into a tier-based framework of dialogue gameboards, however (Lücking and Ginzburg, 2020). Of course, the pointing gesture alone is not able to discern the constituent which is indicated to be already familiar. This is achieved by the accompanying speech that in case of example (6) involves a repetition of the constituent's PHON type.

However, identification of the grounded constituent does not necessarily involve segmental repetition. In example (17) (SaGA V2, 9:16) the constituent in question is identified by means of its order of appearance in the route direction:

(17)

- 1 F: ok\_nochmal beim anfang dieses <<pointing at R> mit den säulen scheint ja
- 2 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?

How does this work? The FACTS field is populated *inter alia* by (descriptive contents of) grounded moves. MOVES are stored within a list. The addressee of *CG pointing* from (17) just has to identify the first move from the route direction list and retrieve the constituent(s) it introduces.

A related example is given in (18) (SaGA V4, 9:43):

(18)

- 1 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?

The difference to (17) is that the constituent-relevant move is indicated in a relative manner (namely after some other route direction component) by *then*, rather than according to its order of appearance.

#### 4.2. Utterance anaphora

While CG pointing indicates that a given constituent is already known, Utt pointing (utterance anaphora) emphasizes a DR of the actual utterance. Accordingly, Utt pointing often occurs with topic (DR) introduction, an affirmation of the utterance of the other interlocutor, clarification request, or corrections – see the frequencies collected in (5) in Sec. 2. It should be noted that Utt pointing formally is not only realized by pointing at R or F, but also by index finger raising, which is not a proper pointing at the addressee. We cannot explore here further such form/function variations, however, though it is a potentially important consideration for future work.

A real-world example is shown in (19) (SaGA V2, 7:30):

(19)

1 R: °hh und dann kommen halt äh (-) die ((pointing at F)) BÄUme



/ and then there will just eh be the (*UTT pointing*) TREEs

Albeit a kinematically modest Utt pointing, R, while prosodically stressing *trees*, points at F (the addressee). Once more, F is not the index of the pointing gesture. Rather, the gesture put emphasis on a DR of the accompanying utterance – in case of (19) this is the constituent associated with the plural noun *trees*. In the context of the route direction this Utt pointing highlights a new landmark. It is therefore bound up with topic introduction or topic switch and contributes to the structure of the ongoing conversation.

The lexical entry we provide in (20) assigns as the content of Utt uses the speech event associated with a sub-utterance of the (maximally) Pending utterance.

(20) Utt  $\mapsto$   $\left[ \begin{array}{l} \text{Shape : PointType} \\ \text{dgb-params : } \left[ \begin{array}{l} \text{MaxPending : LocProp} \\ \text{u : sign} \\ \text{c1: member(u1, MaxPending.sit.constits)} \end{array} \right] \\ \text{cont = u.sit : Rec} \end{array} \right]$

#### 4.3. SCTM

SCTM pointing (*something's coming to mind*) indicates just that: the speaker suddenly recalls something different from what he or she is talking about. SCTM is best illustrated by means of an example (SaGA V4, 5:23):

(21)

- 1 R: da gehst du rein (-) °h da kommt n SEE: / there you enter, and there is a  
 2 LAKE  
 3 R: ah gut ((pointing at F)) (.) ich glaub / well (*SCTM pointing*) I guess



- 4 es kam doch erst der park / there was the park first

In (21) the direction giver R continues her route description by introducing what she believes to be the next landmark/topic (namely the lake). She then recognizes that she was confused: the park was before the lake. The point of recall is indicated by particles in speech (*ah gut*) and by addressee pointing. Intuitively SCTM pointing signals something like *wait a moment* and *there will be a modification/repair*.

So at the bottom line SCTM involves topic change. More precisely, SCTM pointing is akin to a *forward looking disfluency* (Ginzburg et al., 2014), a discourse particle that provides indications about a looming utterance, in this case that the issue it concerns is distinct from the current one. In this case we capture the effect in terms of a lexical entry that expresses the move effected by the pointing and a conversational rule that indicates a subsequent contextual update such a move underwrites.

$$(22) \quad \text{SCTM} \mapsto \left[ \begin{array}{l} \text{Shape : PointType} \\ \text{dgb-params : } \left[ \begin{array}{l} \text{spkr : Ind} \\ \text{addr : Ind} \\ \text{utt-time : Time} \\ \text{c1 : Address(spkr,addr,utt-time)} \\ \text{Pending.cont : IllocProp} \\ \text{q : Question} \\ \text{c1 : About(Pending.cont,q)} \end{array} \right] \\ \text{cont = ChangeTopic(spkr,q) : IllocProp} \end{array} \right]$$

$$(23) \quad \text{SCTM conversational rule} \mapsto \left[ \begin{array}{l} \text{preconds : } \left[ \begin{array}{l} \text{q : Question} \\ \text{LatestMove = ChangeTopic(spkr,q) : IllocProp} \end{array} \right] \\ \left[ \begin{array}{l} \text{spkr = preconds.spkr : Ind} \\ \text{Pending.cont : IllocProp} \\ \text{c2: } \neg \text{About(Pending.cont,preconds.q)} \end{array} \right] \end{array} \right]$$

We found SCTM in two variants: as addressee pointing and as index finger raising.

#### 4.4. GrabTurn

Probably the most straightforward kind of discourse pointing is GrabTurn: it effectuates turn change. Accordingly, it is affiliated to turn-taking expressions – in (25), for instance, a request to pose a clarification question. With just two instances, GrabTurn was the least frequent kind of

discourse pointing in our corpus, and both occurrences were produced by index finger raising. So it remains to be seen whether there is also an addressee pointing variant, as we suspect. The first occurrence of GrabTurn in our sample is the following (SaGA V4, 4:28):

(24)

- 1 R: du bleibst auf jeden fall auf der straÙe wo du bist und gehst geradeaus °h /
- 2 in any case you stay on the street where you are and go straight ahead
- 3 F: <<index raised, repeated>ich frage nochmal kurz was nach> (.) also ähm / I
- 4 have a quick clarification request ehm



F interrupts R with a raised index finger. F tries to catch R's attention both visually as acoustically. The reason for the interruption is then explained. As with STCM, we analyze GrabTurn by means of a lexical entry and a conversational rule that gives rise to turn change:

$$(25) \quad \text{GrabTurn} \mapsto \left[ \begin{array}{l} \text{Shape : PointType} \\ \text{dgb-params : } \left[ \begin{array}{l} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time : Time} \\ \text{c1 : Address(spkr,addr,utt-time)} \end{array} \right] \\ \text{cont = GrabTurn(addr,utt-time) : IllocProp} \end{array} \right]$$

$$(26) \quad \text{GrabTurn conversational rule} \mapsto \left[ \begin{array}{l} \text{Preconds : } \left[ \begin{array}{l} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time : Time} \\ \text{LatestMove = GrabTurn(addr,utt-time) : IllocProp} \end{array} \right] \\ \text{Effects : } \left[ \begin{array}{l} \text{spkr = pre.addr : Ind} \\ \text{addr = pre.spkr : Ind} \end{array} \right] \end{array} \right]$$

## 5. Conclusions

In sum, 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. Accordingly, a dialogical notion of grammar is required, in terms of which discourse pointing can be analyzed.

It is tempting to think about a coherent framework for the various uses of pointing gestures: identification, abstract, deferred, and discourse. A common theme seems to be that in any of these uses the pointing gesture acts as an instruction for the addressee to find the referent (which in turn is further described in speech or by contextual salience). Just the search domain differs:

- visual domain in concrete deictic pointing
- knowledge domain for indirect classification in deferred reference
- geometric projection in abstract deixis

- utterance context and dialogue management in discourse pointing

A coherent framework will emerge in future work. This work includes extended corpus work in order to get a better quantitative picture of the distribution of discourse pointing, and to identify potentially further uses of addressee pointing. We have only very briefly covered gaze and intonation. Following a truly multimodal analysis, however, the functional interaction of discourse pointing with other verbal and non-verbal signals will be examined. Accordingly, multi-tier extensions of dialogue gameboards provide a starting point (Lücking and Ginzburg, 2020).

## References

- Alahverdzhieva, K., A. Lascarides, and D. Flickinger (2017). Aligning speech and co-speech gesture in a constraint-based grammar. *Journal of Language Modelling* 5(3), 421–464.
- Austin, J. L. (1950). Truth. In *Proceedings of the Aristotelian Society. Supplementary*, Volume xxiv, pp. 111–128. Reprinted in John L. Austin: *Philosophical Papers*. 2. ed. Oxford: Clarendon Press, 1970.
- Barwise, J. and J. Etchemendy (1987). *The Liar*. New York: Oxford University Press.
- Barwise, J. and J. Perry (1983). *Situations and Attitudes*. Bradford Books. Cambridge: MIT Press.
- Bavelas, J. B., N. Chovil, D. A. Lawrie, and A. Wade (1992). Interactive gestures. *Discourse Processes* 15(4), 469–489.
- Bergmann, K., O. Damm, F. Freigang, C. Fröhlich, F. Hahn, S. Klett, A. Löcking, J. Kopp, Stefan Letetzki, H. Rieser, N. Thomas, and N. Wittwer (2014). *Documentation – Sagaland*. Bielefeld University: SFB 673, Project B1. <https://www.phonetik.uni-muenchen.de/Bas/BasSaGAeng.html>.
- Betarte, G. and A. Tasistro (1998). Martin-Löf’s type theory with record types and subtyping. In G. Sambin and J. M. Smith (Eds.), *25 Years of Constructive Type Theory. Proceedings of a Congree held in Venice, October 1995*. New York: Oxford University Press.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer Verlag.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Cooper, R. (2005). Austinian truth, attitudes and type theory. *Research on Language and Computation* 3(4), 333–362.
- Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, and T. Fernando (Eds.), *Handbook of the Philosophy of Science*, Volume 14: Philosophy of Linguistics, pp. 271–323. Amsterdam: Elsevier.
- Cooper, R. (2021). From perception to communication: An analysis of meaning and action using a theory of types with records (TTR). <https://github.com/robincooper/ttr>. Book Draft.
- Cooper, R. and J. Ginzburg (2015). Type theory with records for natural language semantics. In C. Fox and S. Lappin (Eds.), *Handbook of Contemporary Semantic Theory, second edition*, Oxford. Blackwell.
- Cooper, R. and M. Poesio (1994). Situation theory. In *Fracas Deliverable D8*. Centre for Cognitive Science, Edinburgh: The Fracas Consortium.
- Ebert, C. and C. Ebert (2016). The semantic behaviour of co-speech gestures and their role in demonstrative reference. Invited talk given at Institut Jean-Nicod, Département d’Études

- Cognitives Ecole Normale Supérieure, Paris.
- Fernández, R. (2006). *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph. D. thesis, King's College, London.
- Frege, G. (1918). Der Gedanke. *Beiträge zur Philosophie des deutschen Idealismus* 1(2), 58–77.
- Fricke, E. (2007). *Origo, Geste und Raum*. Number 24 in *Linguistik – Impulse & Tendenzen*. Berlin, New York: Walter de Gruyter.
- Fricke, E. (2012). *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken*. Number 40 in *Linguistik – Impulse und Tendenzen*. Berlin, Boston: De Gruyter.
- Ginzburg, J. (1994). An update semantics for dialogue. In H. Bunt (Ed.), *Proceedings of the 1st International Workshop on Computational Semantics*. Tilburg: ITK, Tilburg University.
- Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.
- Ginzburg, J. and R. Cooper (2004). Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy* 27(3), 297–366.
- Ginzburg, J. and R. Fernández (2010). Computational models of dialogue. In A. Clark, C. Fox, and S. Lappin (Eds.), *Handbook of Computational Linguistics and Natural Language*, Oxford: Blackwell.
- Ginzburg, J., R. Fernández, and D. Schlangen (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics* 7(9), 1–64.
- Ginzburg, J., C. Mazzocconi, and Y. Tian (2020). Laughter as language. *Glossa: a journal of general linguistics* 5(1).
- Ginzburg, J. and M. Poesio (2016). Grammar is a system that characterizes talk in interaction. *Frontiers in Psychology* 7, 1938.
- Ginzburg, J. and I. A. Sag (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in *CSLI Lecture Notes*. Stanford: California: CSLI Publications.
- Grosz, P. G. (2019). Pronominal typology and reference to the external world. In *Proceedings of the Amsterdam Colloquium 2019, AC'19*, pp. 563–573.
- Higgleton, E. and A. Seaton (Eds.) (1996). *Harper's Essential English Dictionary*. New Delhi, India: Allied Chambers.
- Holler, J. (2010). Speakers' use of interactive gestures as markers of common ground. In S. Kopp and I. Wachsmuth (Eds.), *Proceedings of Gesture Workshop 2009*, Number 5934 in *Lecture Notes in Artificial Intelligence*, pp. 11–22. Springer.
- Kaplan, D. (1978). Dthat. In P. Cole (Ed.), *Pragmatics*, Number 9 in *Syntax and Semantics*, pp. 221–243. New York, San Francisco, London: Academic Press.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes from Kaplan*, pp. 481–563. New York, Oxford: Oxford University Press.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication*, Number 25 in *Contributions to the Sociology of Language*, pp. 207–227. The Hague: Mouton.
- Kühnlein, P. (1999). Dynamics of complex information. In E. André, M. Poesio, and H. Rieser (Eds.), *Proceedings of the Workshop on Deixis, Demonstration and Deictic Belief at ESSLLI XI*. Paper 11.
- Larsson, S. (2002). *Issue based Dialogue Management*. Ph. D. thesis, Gothenburg University.

- Lascarides, A. and M. Stone (2009). A formal semantic analysis of gesture. *Journal of Semantics* 26(4), 393–449.
- Lücking, A. (2013). *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. Berlin, Boston: De Gruyter. Zugl. Diss. Univ. Bielefeld (2011).
- Lücking, A. (2018). Witness-loaded and witness-free demonstratives. In M. Coniglio, A. Murphy, E. Schlachter, and T. Veenstra (Eds.), *Atypical Demonstratives. Syntax, Semantics and Pragmatics*, Number 568 in Linguistische Arbeiten. De Gruyter.
- Lücking, A. and J. Ginzburg (2020). Towards the score of communication. In *Proceedings of The 24th Workshop on the Semantics and Pragmatics of Dialogue, SemDial/WatchDial*.
- Lücking, A., T. Pfeiffer, and H. Rieser (2015). Pointing and reference reconsidered. *Journal of Pragmatics* 77, 56–79.
- Lücking, A., K. Bergmann, F. Hahn, S. Kopp, and H. Rieser (2010). The Bielefeld speech and gesture alignment corpus (SaGA). In *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, LREC 2010, pp. 92–98. 7th International Conference for Language Resources and Evaluation.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review* 92(3), 350–371.
- McNeill, D., J. Cassell, and E. T. Levy (1993). Abstract deixis. *Semiotica* 95(1-2), 5–19.
- Nunberg, G. (1993). Indexicality and deixis. *Linguistics and Philosophy* 16(1), 1–43.
- Purver, M. (2006). Clarie: Handling clarification requests in a dialogue system. *Research on Language & Computation* 4(2), 259–288.
- Ranta, A. (1994). *Type Theoretical Grammar*. Oxford: Oxford University Press.
- Rieser, H. (2004). Pointing in dialogue. In *Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*, Catalog '04, pp. 93–100.
- Seligman, J. and L. Moss (1997). Situation Theory. In J. van Benthem and A. ter Meulen (Eds.), *Handbook of Logic and Linguistics*. Amsterdam: North Holland.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann, K. Birkner, E. Couper-Kuhlen, A. Deppermann, P. Gilles, S. Günthner, M. Hartung, F. Kern, C. Mertzluft, C. Meyer, M. Morek, F. Oberzaucher, J. Peters, U. Quasthoff, W. Schütte, A. Stukenbrock, and S. Uhmann (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402.