Dialogue & Discourse 15(2) 36-84 doi: 10.5210/dad.2024.202

Multi-modal Anaphora and Broadcasting of Information by Gestural Post-holds

Hannes Rieser Hannes.Rieser@Uni-Bielefeld.de
Bielefeld University
Faculty of Linguistics and Literary Studies
Germany

Editor: Jonathan Ginzburg

Submitted 04/2022; Accepted 08/2024; Published online 09/2024

Abstract

This paper deals with verbal anaphora, multi-modal anaphora and the top-down broadcasting of information using gestural post-holds in multimodal dialogue. First, a new account of definite, pronominal and pro-adverbial anaphora is given. The approach is then extended to multimodal anaphora where part or all of an anaphor's meaning is contributed by iconic or deictic gestures. As tradition has it, anaphora work "bottom-up"; making use of Process Algebra, here an inverse relation "broadcasting" is defined, where information is distributed top down and input to receiving ports. Anaphora are modelled with this relation; the same is attempted for dialogue coherence relations which generally exhibit anaphora-like behaviour. As video data show, broadcasting is tied to gestural post-holds, the holding of a gesture's stroke over a stretch of discourse, independently of alignable speech. Multi-modal anaphora and broadcasting cross single contributions and turns. This leads to considering post-holds from a new perspective, stressing their speech-independent function and their relevance for indicating topic-continuity. The data come from the SaGA (the Speech and Gesture Alignment) corpus, a set of route-description dialogues generated in a Virtual Reality-setting. The calculus used to model the anaphora and broadcasting dynamics is the λΨ-calculus, a specifically developed Process Algebra. It uses the Ψ-calculus for input-output, data transport and broadcasting. Extending the Ψ-calculus, the data transported are typed λ-calculus formulae equipped with compositional links; they can be verbal only, gestural only or multi-modal. Following the Ψ-calculus, information chunks are modelled as communicating agents sending or receiving information. The λΨ-calculus is also used for the fusion component unifying gestural and verbal information; hence, the paper is also an independent contribution to multi-modal fusion.

Keywords: Anaphora, iconic gestures, gestural post-holds, broadcasting, multimodal fusion, multimodal meaning, Ψ-calculus, λΨ-calculus

1 Introduction

Multi-modal anaphora is a difficult topic to treat and so is the evaluation of gestural post-holds[1] and the modelling of their meaning. Both are modelled using information transfer, i.e., broadcasting of information. Broadcasting of information in discourse and its seemingly unrestricted distribution top-down might be unfamiliar to researchers in semantics, pragmatics and gesture semiotics. These

[1] McNeill (1992, p. 83 and pp. 288-89) also deals with post-stroke holds but tries to harmonise them with his synchrony concept.

themes are tightly bound up with multi-modal fusion, the idea, especially entertained in AI (see the literature review, Sec. 3), that information from different sources, linguistic, gestural, facial or body posture are brought together in one representation. We discuss and model the fusion of speech and aligned iconic gesture using a passage from a route-description dialogue. It is taken from the Bielefeld multi-modal SaGA corpus and provides a wealth of data on (multi-modal) anaphora, two-handed coordinated gestures, gestural post-holds and broadcasting (top down information passing). The data discussion already provides an idea of which properties an algorithm must have to simulate these. These properties are independence of gesture and speech, concurrency, temporal asynchrony of gesture and speech, and composition of gesture semantics and speech semantics. The discussion of the literature, also tailored to the datum, shows that coherence-bound linguistic anaphora theories in the Hobbs-Hume-Kehler-Rohde tradition are important as are DRT and SDRT accounts of coherence relations. The multimodal fusion mechanism developed is set up against recent multimodal fusion approaches in AI. The afore-mentioned characteristics motivate the calculus we introduce, the Ψ-calculus, a powerful extension of Milner's π-calculus (Milner, 1999, Parrow, 2001). The Ψ-calculus has agents as basic entities and works with channels, input-output facilities, information transfer and operators for concurrency, replication of information as well as operators from standard logics. The agents can be seen as monads able to do internal computation and to communicate information. They can embody expressions of higher order logics and transport these via output and input channels. The formulas transported are standard formulas of a typed λ-calculus. Therefor the two-tiered machinery is called λΨ-calculus. The λ-calculus formulas are static, the dynamics rests with the Ψ-calculus component. Both rely on top-down information processing leading to a new mechanism for anaphora resolution and confirming the independence of gesture information and its role as a topic holder. Due to the focus on anaphora and broadcasting there is not much on dialogue structure in this paper. In the Appendices it is shown how the mechanism used for anaphora resolution and broadcasting can be used to model rhetorical relations and anaphora resolution in PTT (Poesio and Traum 1997, Poesio and Rieser 2010). Due to the main topic of the paper several issues are treated with lesser care such as the characteristics of German spoken language syntax, turn-taking regularities, and dialogue update; see however Sec. 5.4 on Follower's semantics, Follower's clarification request and Follower's decision for an anaphora antecedent[2].

The paper extends research on multi-modality elaborated in Rieser (2015, 2017), Lawler, Hahn, and Rieser (2017), and, Rieser and Lawler (2020). It offers a new approach to verbal anaphora taking up a suggestion due to Chastain (1975), multi-modal anaphora using a multi-modal fusion account, treats two-handed concurrent gestures and provides a reconstruction of the semantics of gestural holds across contributions and turns. The algorithmic background for that is the λΨ-calculus.

Sec. 2 provides the annotated data, a passage from a Virtual Reality (VR) route description dialogue and an analysis of which algorithmic devices will be needed for its description, an issue taken up in more detail in Sec. 4. Sec. 3 discusses some of the literature relevant for the anaphora relations in the datum, work in the Hobbs-Hume-Kehler-Rohde research line, DRT, SDRT, and definite anaphora. A short sub-section deals with speech-gesture alignment. Multi-modal fusion approaches integrating speech semantics and information from other modalities are briefly commented upon. Against this background, coherence relations, anaphora and multi-modal anaphora in the datum receive a first scrutiny. The literature review section closes with remarks on methodology and the contributions the paper provides. Sec. 4 gives the Process Algebra set-up, i.e. the standard Ψ-calculus definitions, and introduces the λΨ-calculus. This done, we informally map the capabilities of the λΨ-calculus onto the structures of the datum outlined in Sec. 2. The discussion of the route-

[2] Given the technical possibilities of the λΨ-calculus it is quite clear how turn-taking and semantic update could be treated, the first by channel communication and the latter by "upwards" recursive processes.

description data and their formal rendering in λΨ comes in Sec. 5. Anaphora and broadcasting of information are described in some detail, identifying central paradigms of gesture-speech cooperation. Finally, Sec. 6 offers with conclusions and suggestions for further research. The Appendices contain λΨ-deductions for the central datum and suggest how the λΨ-account can be extended to capture rhetorical relations in SDRT and anaphora resolution in PTT.

2. Datum and Annotation

The section of dialogue serving as our main example comes from the SaGA corpus (see Lücking et al., 2013), dialogue V5. The corpus is a collection of 25 VR route-description dialogues in which a Route-Giver has made a ride on a "bus" through a VR site passing various landmarks. These were a sculpture, a town-hall, a market place with two churches, a park with a pond, a chapel and a fountain, all connected by streets (see Figure 1, the route through the scenery, for details). After the ride the experience is reported to a Follower expected to do the tour. The Route-Giver-Follower exchange chosen (see Figure 1 and the transcript Figure 2) captures the ride from a hedge surrounding the park, its doorways and the main entrance towards a pond, and around it up to a place where an ice-cream man offers his services. The "bus"-tour leaves the park at the ice-cream stand and goes on to a chapel (1.1-G to 11-RH below).

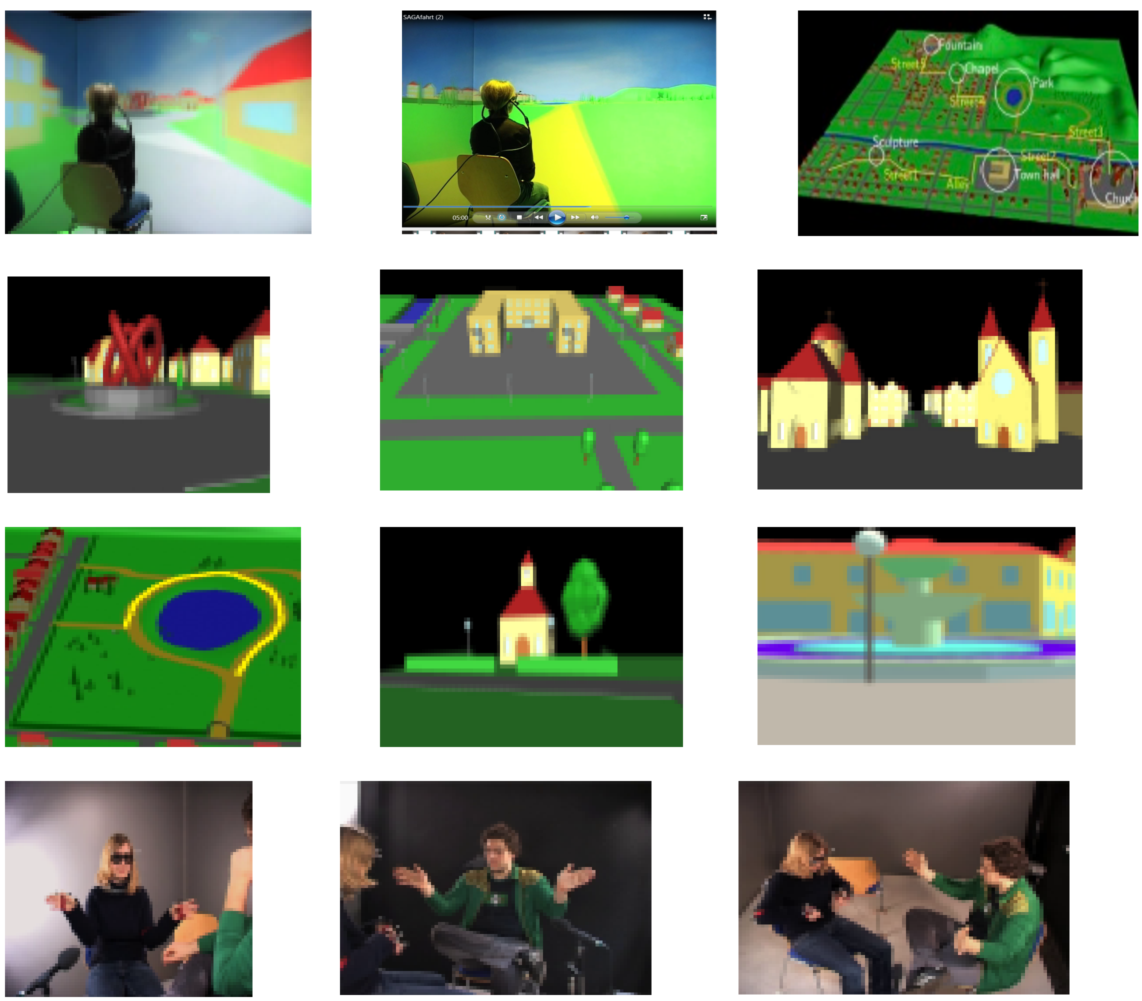


Figure 1. Route-Giver's VR ride on "bus", Route-Giver entering the park with pond in the distance, the route through the scenery, landmarks: sculpture, townhall, two churches, park entry and path around the pond, chapel, fountain; an exchange: Route-Giver signing the two churches, Follower signing left and right church, both concurrently, Route-Giver and Follower interacting.

In the annotation (see Figure 2 below) we provide the German speech transcription (-G), an English word-by-word translation (-E) and a transcription of the Right-Hand (RH) and Left-Hand (LH) gestures. In the main text, the English translation is used except in cases, where German data must be considered. For characterizing the natural hand-gestures of the datum we use approximations to ASL-handshapes. In order of appearance these are the following (see Figure 3 below): C, small c, q, G, B, loose B (not represented in Figure 3), O, and D. Next to the ASL-handshapes you find an approximation of them in the datum. Note the differences in palm- and finger positions. Complementing the ASL-handshapes we work with observational predicates like "small arc-shape", "large grey arc-shape", "drawing trajectory right angle", "palm flat", "indexing location at embankment of pond". Since we deal with two-handed gestures in the paper, the LH-RH-coordination is all important. We notate speech-gesture overlap with [] and {} in the following way: L-Handshape is marked with brackets "[….]", R-Handshape with curly brackets "{….}". In the datum we have a two-handed gesture (or two coordinated gestures) and see that the L-Handshape O is held throughout *[And at this pond. You {drive towards it and you drive right around it.}]* and the R-Handshape D underpins *{drive towards it and you drive right around it.}*. In other words, the pond gets an LH-O shape constantly held and the driving gets a trajectory with D shape extending over most part of 4.1 and 4.2. Observe that the gestures cover much more than their natural alignment expression[3]. This shows a characteristic temporal asymmetry between gesture and speech discussed in (Rieser and Lawler 2020). In this paper we do not provide the original ELAN corpus annotation and the map from corpus annotation to semantics but assume it as given.

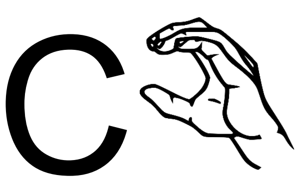
1.1-G	Route-Giver: Dann {kommt auf der rechten Seite eine Hecke}, {eine grüne Hecke}. {Zum Teil auch mit Eingängen.}
1.1-E	Then comes at the right side a hedge, a green hedge. Partly also with doorways.
1.1-LH	……
1.1-RH	{R-Handshape C, }, {R-H C, small c }. {R-Handshape q }.
1.2-G	Route-Giver: Die sind dann so bogenförmig, wie so n' Garteneingang.
1.2-E	These are then like arch-shaped, like such a garden entrance.
1.2-LH	………
1.2-RH	{R-Hands. small-arch-shape},
1.3-G	Route-Giver: 1.3.1 Bis irgendwann einmal dann der Haupteingang kommt. 1.3.2 Das ist ein großer grauer Bogen.
1.3-E	Until sometime then the main entrance comes. That is a large grey arch.

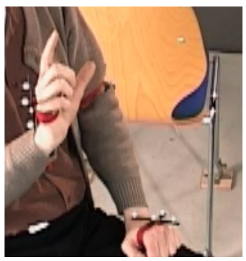
[3] The standard segmentation of gesture going back to McNeill (1992, pp. 83-85) and Kendon (2004, pp. 111-124) is in preparation, stroke and retraction. The cited datum has no observable preparation phases or retractions. The impression is that gesturing proceeds from stroke to stroke. It can well be that post-holds have a dual function; besides preserving the semantics of the stroke they could function like a quasi-retraction.

1.3-LH	………
1.3-RH	{R-Hands. large arc-shape + small c}.
1.4-G	Route-Giver: Und da mußt du sofort, scharfer rechter Winkel, rechts rein.
1.4-E	And there must you quickly, sharp right angle, to the right into.
1.4-LH	………..
1.4-RH	{G, drawing trajectory right angle, + modelling bus }.
1.5-G	Follower: In den Bogen rein?
1.5-E	In the arch into?
1.5-LH	
1.5-RG	Route-Giver: { G, drawing trajectory right angle }
2.1-G	Route-Giver: 2.1.1 Ja, in den Bogen rein. 2.1.2 Es ist keine Wand sondern eine Hecke.
2.1-E	Yeah. Into the arch. It isn't a wall, it's a hedge.
2.1-LH	[L-Handshape palm flat]
2.1-RH	{ G, drawing trajectory right angle } { R-Handshape palm flat }
3.1-G	Follower: OK.
3.1-LH	……………………….
3.1-RH	……………………….
3.2-G	Route-Giver: 3.2.1 Wenn du dort eingefahren bist, 3.2.2 fährst du geradeaus auf einen Teich zu. Einen Teich.
3.2-E	If you have there driven in, you drive [{straight} towards a pond. A pond.]
3.2-LH	C; [B parallel trajectory] [B parallel trajectory] [L-Handshape O ctnd.]
3.2-RH	{B parallel trajectory } {R-Handshape C > open O ctnd }
4-G	Route-Giver: [4.1 Und an diesem Teich. 4.2 Du {fährst drauf zu und 4.3 du fährst rechts herum. }]
4-E	[And at this pond. You {drive towards it and you drive right around it. }]
4-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.ctnd. ctnd.]
4-RH	{R-Handshape D ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.ctnd. ctnd. }
5-G	Route-Giver: [Die Hecke, {die geht noch ungefähr} so 50 m.]
5-E	[The hedge,{it runs another roughly} 50 m.]
5-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd.]
5-RH	{R-Handshape loose B ctnd. ctnd. ctnd. ctnd. }
6-G	Route-Giver: [Und dann sind dort {auch} hin und {wieder} {Sitzbänke}.]
6-E	[Then there are {also} here and { there } {benches} .]
6-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]

6-RH	{R-Handshape loose D} (first two overlaps) gesture expressing doubt: {wiggling of loose D Handshape} (third overlap)
7-G	Route-Giver: Ab[{er du fährst um den Teich herum. Rechts herum. }]
7-E	[{But you drive around the pond. Right around. }]
7-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
7-RH	{R-Handshape D ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. }
8-G	Route-Giver: 8.1 [Und manchmal ist da auch {nen Eisverkäufer }.8.2 Und an dem fährst du rechts ab.]
8-E	[And sometimes there is {'n ice-cream man there.} And at him you drive off to the right.]
8-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
8-RH	{R-Handshape G ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.}
9-G	Follower: Was heißt "manchmal"?
9-E	What does "sometimes" mean?
10-G	Route-Giver: 10.1 [Ja, könnte verändert werden. 10.2 Auf jeden Fall, {auf} meiner Tour war dort ein Eisverkäufer.]
10-E	[Well, could be changed. In any case {on} my tour there was an ice-cream man there.]
10-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
10-RH	{R-Handshape D, 3 beats/indexing location at embankment of pond}
11-G	Route-Giver: Ein Wagen mit Schirm.
11-E	[{A cart] with umbrella. }
11-LH 11-RH	[LH + RH pushing Pantomime] {RH modelling umbrella}

Figure 2. Annotation of the paper's main datum. Information relevant for anaphora is marked yellow; broadcasting cases are marked blue; "." represents larger and "," smaller speech pauses. Observe the amount of broadcasted information.





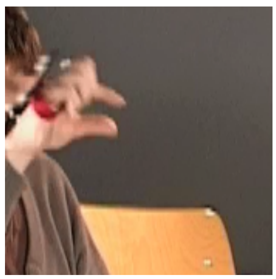


Figure 3. Used ASL shape names C, small c, q, G, B, O, D (right hand) with natural gesture occurrences for comparison. Observe that the natural D gesture in the datum uses the index finger and middle finger.

Data and Algorithmic Handling

As a preview, we consider what an algorithm modelling some features of the datum must cover: First of all, we need independent renderings of gesture and speech information since they carry independent information: From the annotation of speech and gesture, RH, LH, it can be seen that RH and LH run in parallel with speech and with each other. Therefore, we need some mechanism to express parallelism/concurrency. The roundness shape of the arch is initially only given by iconic gesture. In order to get "round arch", round'[4], the gesture semantics, and arch', the word semantics, have to communicate. Specifically, we need mechanisms of information transfer and information composition. We have the relation of 1.1-G to 1.2-G, an antecedent-anaphora relation; to capture this we need a device mapping the antecedent information onto the anaphora. The blue marks indicate persistence of information across contributions and turns, e.g., gesturing a right angle, the

[4] Information given in red indicates that it was provided by gesture.

pond or the moving-towards-the-pond trajectory. Lines 3.2 and 4 contain conditionals, hence a formal analysis of conditionals is needed.

3. Relevant Literature Background Vis-à-Vis the Datum

Anaphora accounts differ as to how they conceptualize the relation between antecedent and anaphora and use it in predictions of which antecedent will be selected. This relation can be one of coherence as in coherence theories, identity of discourse referents as in DRT and DRT-like approaches or a definiteness algorithm in corpus-based research. For a background to speech-gesture integration we look into literature on multi-modal fusion. These are broad fields: As general overviews we recommend Poesio (2016), Poesio et al. (2016) for anaphora, Geurts, Beaver, Maier (2020) for DRT and Johnston (2019) for Finite State models.

3.1 Coherence-bound Theories

The literature on linguistic anaphora mainly focusses on 3rd person pronominal resolution (Krahmer and Piwek, 2013, King and Lewis 2018, Muskens 1996, Rohde 2018). There, the works of Hobbs gave the initial impulse: Hobbs (1978) suggested a coherence based semantic system for pronoun interpretation relying on the inference patterns Contrast, Cause, Violated Expectation, Temporal Succession, Paraphrase, Parallel and Example. Kehler's (2002) coherence notion for pronoun resolution drew on Hobbs (1978 and 1979 *inter alia*). It takes up Hume's assumptions on connection among ideas (Hume 1748/1955), "*Resemblance*, *Contiguity* in time or place, and *Cause* or *Effect*". Kehler's class of Resemblance Relations includes Parallel, Contrast, Exemplification, Generalisation, Exception, and Elaboration. For pronoun resolution he distinguishes Coherence-Driven Approaches (like his own) from Attention-Driven Approaches such as Centering Theory (Brennan et al. 1987). Kehler et al. (2007) investigated the coherence-driven approach with respect to four interpretation biases, grammatical role parallelism, thematic role, implicit causality, and subjecthood. Recently, Rohde (2018), following this line of research, distinguished two types of accounts: a coherence-based one (Hobbs 1978, 1979, Kehler 2002) and a surface form based one as in, e.g., Centering Theory. She argues that an efficient model of pronoun reference must cover both, coherence- and surface-relatedness: Coherence Theory (semantics bound) and Centering Theory (surface bound) cover distinct factors of her experimental findings. A unified approach subsuming both, coherence- and surface-relatedness, is proposed, formulated in Bayesian terms.

3.2 DRT and DRT-like Approaches

Kamp and Reyle (1993) have the widest empirical coverage (singular and plural pronouns, tense and aspect) for anaphora. They treat the anaphora problem as follows: starting from surface syntax, they map syntactic structures onto representations (DRSs, Discourse Representation Structures) and then provide these with a first order interpretation[5]. Pronominal anaphora are analysed as relations between pronouns and discourse referents introduced in the discourse by an accessible antecedent. They show that numerous grammatical constraints determine accessibility. A further step in understanding the workings of anaphora in discourse was achieved with Asher and Lascarides (2003), referred to as AL below. While building on the insights of Kamp and Reyle (1993) concerning grammar-based mechanisms of anaphora resolution, they argue that one must consider the rhetorical structure of discourse and the right frontier constraint. Both influence anaphora resolution and the regularities of information packaging. In addition, they extend the anaphora notion to sentence pairs linked by cue-words like *but*, *because* or *and-then* and show that rhetorical relations are operative in dialogue. This is similar to the Hobbs-Hume-Kehler coherence relations. Von Heusinger (2007) treats *inter alia* alternations of definite NPs and pronouns in

[5] One of the reviewers noted that Muskens (1996) should also be mentioned in this context.

anaphoric chains. Against standard DRT and SDRT approaches he proposes five aspects of accessibility: activation, accessibility relation, accessibility hierarchy, accessibility structure, and salience. One of our papers which touches on many topics relevant in the anaphora context and specific issues of multi-modality is (Poesio and Rieser, 2011), formulated in PTT. It observes linguistic and psycholinguistic findings which plausibly show that interpretation of reference is incremental and interpreters' reference hypotheses run in parallel. The focus of the 2011 paper is *inter alia* on definite and pronominal anaphora. Its anaphora theory uses contexts (situations in the sense of Situation Semantics) represented as DRSs specifying the antecedents of anaphora. Building on Poesio and Rieser (2011a), this context theory uses visual situations to provide the semantics of indexicals, pointings and focus movement.

3.3 Definite Anaphora

Corpus work on definite description anaphora is rare, Poesio and Vieira (1998) being an exception. They ran two experiments asking subjects to classify the uses of definite descriptions in a corpus of articles from the Wall Street Journal. The classification schemas used were taken from Hawkins (1978), Prince (1981), and Fraurud (1990). In one experiment the classification tags were Co-referential, Bridging, Larger Situation, Unfamiliar, and Doubt; note the difference from the parameters relied on in the Hobbs-Kehler tradition. Remarkably, subjects disagreed about the antecedents of definite descriptions. This is a crucial finding, since it points to the non-determinism of the antecedent-anaphora relation.
We now turn to the alignment relation of speech and co-speech gestures.

3.4 Speech-gesture Alignment[6]

Giorgolo and Verstraaten (2008) carried out two perceptual experiments using shifts of gesture with respect to phonological information. Their experiments pointed to a correlation between gesture information and phonetic peak. They looked for the anchor at which gesture is coupled to speech. Using evidence from motion tracking data they found that gestures' peak velocity is closely aligned to the pitch (F0) of speech. However, their results also indicated that semantic information played some role. Leonard and Cummins (2010) investigate asynchrony between beats and speech. Subjects' detection of asynchrony of post-word beats was considerably better than with pre-word beats (first experiment). A second experiment revealed a relatively tight alignment of gestural apex and pitch accent. The authors stress that their findings do not extend to naturally occurring gestures. Navarretta (2011) points to Eisenstein and Davies' (2006) studies which show that gestural holds can be useful for co-reference resolution. She investigated gestures semantically related to anaphora and co-referring expressions in Danish spontaneous two-and three-party conversations. Gestures related to anaphora are deictic or iconic; they can occur with speech or on their own. If antecedents and anaphora are affiliated to gestures, the shape descriptions of both show "many common attributes and values" (p. 8). Navarretta concludes that "The shape of hand gestures might therefore contribute to the resolution of anaphora and co-referring expressions". Esposito and Esposito (2011) conducted experiments with different age groups which showed that speech pauses are highly synchronized with gestures: however, in Esposito et al. (2001) also non-synchronisation data were described. Navarretta (2021) offers data about corpus investigations and classification experiments concerning the role of silent and filled speech pauses and accents for the identification and classification of individual and abstract anaphora in Danish.
As far as we are aware, there is virtually no research on gestural post-holds. In contrast to coherence bound theories, DRT-like approaches to anaphora resolution or definiteness accounts which all work bottom up, we develop an alternative resting on the idea that antecedents and anaphora

[6] This section has been introduced following the suggestion of an anonymous reviewer for D&D. For the perspective on speech-gesture asynchrony maintained here, cf. the section "Methodology and Contributions of Paper" in 3.8.

communicate information from (in)definites coming top down and that this is taken up by anaphora ports. The topic of multi-modal anaphora is, as far as we can see, modelled here for the first time, although our intuitions resemble those of Navarretta (2011). Concerning speech-gesture alignment, we start from the assumption of speech-gesture asynchrony mediated by concurrent speech-gesture communication which is orthogonal to the Kendon-McNeill tradition.

3.5 Multi-modal Fusion

One of the main aims of this paper is to show how gesture meaning and speech meaning can be integrated. The following aspects of multi-modal integration are important for our concerns: (a) integrated accounts of at least two different modalities, (b) the special mechanism of integration used, (c) composition of information[7] , (d) incremental (step-by-step) input, (e) handling of temporal modality co-occurrence.

Aspect (a) means that modalities occurring together are mapped onto one representation format. Aspect (b) looks to the means of integrating the modalities, such as, e.g., frames, types of unification or finite automata. How information from different modalities is integrated is the topic of (c). By (d) we understand the successive modelling of bits of information such as, for example, incoming words and gestures which was the main topic of Poesio and Rieser (2011). The temporal relation between the modalities is considered in (e). As work from AI and linguistics shows, multimodal integration can start from independent representations for gesture and speech and then provide some fusion machinery; (a) is usually taken for granted. Development of fusion machineries is one of the most productive research fields in recent AI. They can be implemented as program-based multi-modal integrators combining uni-modal information as in robotic accounts (Bergmann et al. 2011, Kopp et al. 2004, Pfleger 2007), interrelated frames (Koons et al. 1993), semantic networks (Mehlmann and André 2012), as Prolog unification (Mehlmann and André 2012), typed feature structure unification (Johnston et al. 1997, Johnston 1998, Alahverdzhieva & Lascarides 2010, 2011, Lücking 2013, Lücking and Ginzburg 2021), parallel automata (Mehlmann and André 2012) or Finite State Automata (Bangalore and Johnston 2009, Johnston 2019). The fusion mechanisms mentioned share some features since they must handle information transfer. Lascarides and Stone’s (2009) gesture-speech account differs from all of these insofar as they link gesture semantics to speech semantics (both propositional) with rhetorical relations.

3.6 Coherence Relations in the Datum

Due to the fact that we deal with the description of a VR-tour which goes from turning point to turning point and from landmark to landmark (see Figure 1, route), Route-Giver and Follower are focused on end-states of events and the beginning of the next event. This is roughly what is captured by the Hobbs-Kehler coherence relation of Occasion (Kehler et al. (2007), p. 24). Occasion captures sentence pairs and the end state bias affecting the choice of pronouns in the follow-up sentence, as for example with *der Haupteingang/das, the main entrance/that* in the datum 1.3-G/E. For getting at overall coherence in a dialogue one needs more; on that one can consult the rhetorical relations of AL. Although AL deal with pronoun interpretation biases only marginally, they have a general definition for antecedents to anaphora (Definition 15, p. 149) which can be used *in tandem* with the rhetorical relations listed in their Appendix D. These can be extended to dialogue (AL, pp. 203-273, vide especially “Why Dialogue and Monologue are Similar)”) and to multi-modal dialogue (Lascarides and Stone 2009). In-turn relations of our datum can be modelled using AL’s Continuation and Narration, where Narration is roughly equivalent to Hobbs-Kehler Occasion[8]. For

[7] One of the reviewers drew our attention to Kuhn (2022) who develops a dynamical semantics for iconic gestures in sign language. Although Kuhn (2022) and λΨ are both compositional, they differ in how compositionality is implemented. This is probably due to the fact that sign language is language (one level) and we deal with the speech-co-speech gesture relation (two levels).

[8] We take Narrations up in the Appendix, 2.1 Rhetorical Relations.

an alternative, modelling coherence in terms of contexts/situations, see Poesio and Rieser (2011). One's take on the anaphora problem in dialogue must in the end be guided by some dialogue theory, which demands more than a standard version of dynamic semantics can provide. This is clear from an early paper of Eckert and Strube (1999) hooking up to the then existing dialogue research. They specify the domain of anaphora antecedents using the notion of common ground existing between discourse participants and across turns. These issues are beyond the scope of the present paper. We dealt with some of them in Poesio and Rieser (2010, 2011).
So far, we haven't considered multi-modal anaphora to which we now turn.

3.7 Anaphora and Multi-modal Anaphora in the Datum

As could be expected, our take on anaphora in the SaGA corpus also reveals a tight connection between discourse structures and anaphora occurrence mainly due to the existence of coherence relations between sections of discourse[9] but in contrast to the regimented data preferred in CL and experimental linguistics, our datum in Fig. 2 shows great variance, cf. (a) to (f):

(a) PrepPhrase/def. article: *[mit] Eingängen/die*, with doorways/these;
(b) NP/neuter demonstrative pronoun: *der Haupteingang/das*, the main entrance/that;
(c) NPpl./NPsingular (subspecies): *mit Eingängen/der Haupteingang*, with doorways/the main entrance;
(d) NPsing./local adverb: *ein großer grauer Bogen/da*, a large grey arch/there;
(e) def. NP/def. NP: *[in]den Bogen/[in]den Bogen*, the arch/the arch;
(f) event or situation/local adverb: Station der Busfahrt/*dort*, stage of bus tour/there.

To provide an impression of the full multi-modal data, we present the anaphora (a) to (f) above in their actual multi-modal occurrence as (a') to (f'). The multi-modal information is only stated in the German examples:

(a') PrepPhrase/def. article: *mit Eingängen + R-Handshape q* (see Figure 4)/*die + R-handshape small arc shape* (see Figure 5); with doorways/these;

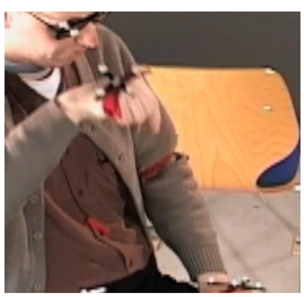


Figure 4. *R-Handshape q*

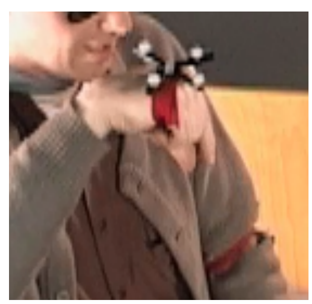


Figure 5. *R-Handshape small arc shape (bottom of arch, representing curb)*

(b') NP/demonstrative pronoun: *der Haupteingang + iconic information from above/das + R-Handshape large arc-shape + small c* (see Figures 6, 7); the main entrance/that;

[9] See Jasinskaja and Karagjosova (2020) for a general discussion of rhetorical relations.

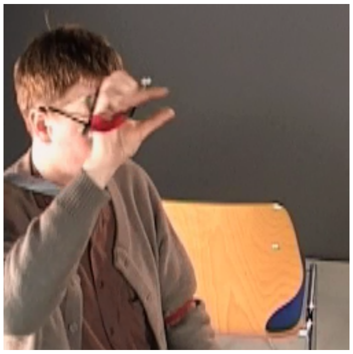


Figure 6. *R-handshape large arc-shape (by an arc drawing gesture) + small c for curb*

Figure 7. *R-handshape large arc-shape + (by a drawing gesture) base of arch*

(c') NPplural/NPsingular(subspecies): *mit Eingängen + R-Handshape small arc shape* (see Figure 5) */der Haupteingang + iconic information from above*; with doorways/the main entrance;

(d') NPsing./local adverb: *ein großer grauer Bogen + R-Handshape large arc shape /da + G drawing trajectory right angle + modelling bus plus track of bus* (see Figures 6 and 8); a large grey arch/there;

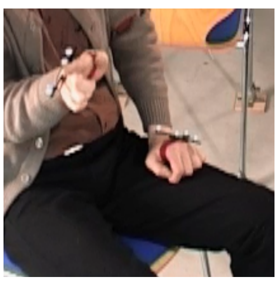


Figure 8. *G drawing trajectory right angle (route into main entrance to park)*

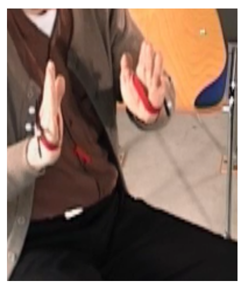


Figure 9. *Modelling (sides of) bus*

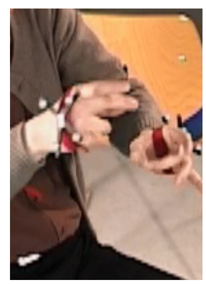


Figure 10. *Stage of bus tour + indexing; observe a pond still signed and held in the left hand*

(e') def. NP/def. NP: *den Bogen + iconic information from above + G drawing trajectory right angle + modelling bus plus track of bus /den Bogen* + iconic information from above + *G drawing trajectory right angle + modelling bus plus track of bus* (see Figures 8 and 9); the arch/the arch;

(f') event or situation/local adverb: *stage of bus tour + indexing /dort + indexing* (see Figure 10); stage of bus tour/there.

In (a') the doorways are specified by ASL q and the uptake of these by a small arc shape, hence the multi-modal meanings are "arch-shaped doorways" and "doorways with small arches", respectively. (b') The main entrance being among the doorways is given iconic information from these. (c') *With doorways* is affiliated with gesture information for their shape; *main entrance* is fused with an iconic gesture as before. (d') The large grey arch gets a large arch shape by drawing using a small C gesture. With the anaphora *there* a trajectory is gesticulated for the movement of

the bus into the arch. (e') gets information already multi-modally expressed plus trajectory information. (f') Anaphora and antecedent are differently indexed.
After these overviews, we give a brief summary of the contributions this paper makes to the topics referred to in the literature section.

3.8 Methodology and Contributions of Paper

Based on longstanding research on the SaGA corpus (see Lücking et al. 2012), we specified in Rieser and Lawler (2020) the following desiderata as a methodological guide line for gesture-speech research. Concretely, a satisfactory account should
(a) Asynchrony: accommodate cases where gesture strokes come (substantially) earlier or later than the "fitting" speech part, i.e., cases where gestures introduce new meaning or modify an existing one regardless of their precise temporal occurrence.
(b) Independence: accommodate the independence of gesture and speech, for instance, it must accommodate cases where gesture strokes are held throughout several utterances.
(c) Blocking: allow for the blocking (non-communication) or postponing of semantic information.
(d) Broadcasting: accommodate broadcasting cases, e.g., by allowing for the replication or repetition of meaning pieces across contributions or turns.
We also add the desideratum that speech-gesture meaning coordination should be determined algorithmically. There should be perspicuity regarding how the gesture meaning coordinates with speech meaning, and the coordination should not be represented in an *ad hoc* fashion but rather be the result of (finite) rule-bound procedures. This enables a systematic explanation of speech-gesture meaning coordination and a generalization to a variety of data and contexts:
(e) Algorithmic determination: A satisfactory account should algorithmically determine a gesture's speech relatum and its coordination term.
These desiderata call for a dynamic machinery, considerably different from Dynamic Semantics or the information flow idea in Barwise and Seligman (1997): Formulated in terms of processes, we need output processes which can give semantic information a "piggyback ride" and we need input processes which receive this semantic information, get it, and hand it on to the right place, the "right place" being (as a rule) already existing information.
In this paper we treat more data-related matters. We suggest solutions to the following problems using the λΨ-calculus where the focus is restricted to semantic representation:

- handling of a broad spectrum of anaphora
- attribution of meaning to iconic and deictic gestures
- fusion of speech meaning and gesture meaning
- handling concurrency and temporal asynchrony of speech and gesture
- treating (chains of) multimodal anaphora
- investigating broadcasting of post-holds across contributions and turns
- handling non-deterministic multi-modal antecedents
- contributing a new field (dialogue, anaphora, multimodal meaning) to the application domain of the Ψ-calculus
- providing examples for the interaction of λ-calculus and Ψ-calculus.

This list looks more ambitious than it is. Once one has developed a speech-gesture fusion account and a machinery relating antecedents and anaphora, solutions for the other problems follow directly.
Anaphora intuitions are quite strong as the literature shows. However, there is no general definition of anaphora. Instead researchers depend on intuitions tied up with examples. The reason for this seems clear: Sec. 3.7 shows that anaphora semantics is denotationally too varied to be subsumed under a common semantic definition. A possibility would be to resort to disjunction. We take a different route: after many dialogue data have been analysed we look into their commonalities in terms of λΨ-derivations in Sec. 5.5 Résumé and Look-back.

4. Process Algebra Set Up

Process calculi (or process algebras) belong to an intensively researched field of AI, Logics, and Philosophy. They have been invented to do away with sequentialism in programming and to establish models for the communication of processes running in parallel (see Bergstra et al. (eds.), 2001 for an impressive view of the topics relevant for the field). By now there are many variants, among them CSP (Communicating Sequential Processes), CCS (Calculus of Communicating Systems), ACP (Algebra of Communicating Processes), Ambient Calculus, π-Calculus, the more recently developed Ψ-calculus, and variants and extensions of these.[10] Most of them share the following features:

Processes (agents[11], processes, threads) which can run sequentially or concurrently.
Process interaction is via input and output facilities.
Process interaction can be regulated to achieve composition, choice, blocking and sequential behaviour.
Processes are used for data transport where the data can come with their own logics.
Processes can send information recursively.

Some of the application domains of process calculi are concurrent programming languages, cryptography, security protocols, business processes, molecular biology or social behaviour (see Fokkink 2000, Introduction, for examples). We now provide an informal description of the λΨ-calculus preparing for **Definitions 1, 2,** and **3** below.

4.1 Basic Intuitions

For matters of easier exposition we provide an intuitive description of the λΨ-calculus ontology, its objects and relations. Its basic dynamic units are so-called agents (processes, threads) interconnected by channels on which communicative processes can run. These agents can receive, internally compute, store and send information. Think of these agents as programs equipped with input and output channels and a simple internal computing mechanism. The agents communicate information coded in a formal language. The computing mechanism consists essentially in receiving a value on an agent's input channel and filling an information slot in the receiving agent. If an agent has completed the internal computations, it can send out the result to arbitrarily many other agents to receive it and work with it. There is a restriction on communication; for communication to succeed between *n* agents, their output and input channels have to correspond. In effect, there is, hence, a communication pipeline from an outputting agent to a receiving one. Each such communication track is typed to handle a special type of informational entity.[12] An agent can send information to different agents distributively and can receive information from different agents. See the remarks about broadcasting below. A special type of information propagation is the so-called replication. It happens if an agent iteratively sends information to other agents. The following Process Algebra set up for the Ψ- and the λ-Ψ-calculus is largely adapted from (Rieser and Lawler 2020) where interested readers can find more detailed information concerning the history of Process Algebra, the history of the Ψ-calculus and some of its formal AI-background.

[10] For more information see the wiki entries for "process calculus" and "π-calculus" which provide easily accessible information. Abramsky (2008), Baeten (2004), and Wing (2002) are also useful. Johansson (2010) is recommended for the Ψ-calculus, Fokkink (2000) for Process Algebra.

[11] The agent notion is widely used in the π-calculus, the Ψ-calculus and other process algebra research traditions. It should not be confused with the "intentional agent" notion used in dialogue theory, action theory and plan-based AI.

[12] How the typing is done following a Montagovian tradition is explained by Rieser and Lawler (2020). See also comment under **Definition 1**.

4.2 Process Algebra Definitions

We do not present and discuss the full Ψ-calculus here but only the parts needed to describe multi-modal integration and the transfer of linguistic and gestural information[13] (for a more elaborate version of the λΨ-calculus cf. (Rieser and Lawler 2020)). What does the Ψ-calculus provide to model the intuitions concerning anaphora and multi-modal fusion laid out under Sec. three? In principle, we have parameters, i.e., variables or names, (concurrent) operators on agents or processes, and dynamically operating agents at our disposal. The semantic representation of linguistic and gestural information is coded in data terms. These can come from any (higher order) logic. In some process algebras like the π-calculus (cf. Parrow 2001) only variables (called names) get associated with input-output channels, but in the Ψ-algebra they can be associated with arbitrary data, e.g., channels or typed λ-terms. Channels help to transport data from one increment in a multi-modal structure to another, and most importantly, across dialogue contributions and turns; we do not know of any other formalism which has this flexibility while maintaining strict control. Hence, they are an obvious means to provide *inter alia* the semantics of anaphors. The parameters indicating Ψ's data types are given in the following **Definition 1** (from Bengtson et al. 2011, pp. 4-14)[14].

Definition 1:
T the (data) terms, ranged over by M, N, N', ….
C the conditions, ranged over by φ
A the assertions, ranged over by Ψ.

In our application data terms T will be standard formulas from a typed λ-calculus such as a property M = λFZ (like'(arch-shaped')(Z) ∧ F(Z)); there λ binds higher order information, a complex property variable F and a set variable Z; the red coloured arch-shaped' indicates that the information comes from an annotated arc gesture semantically mapped onto an arch-shape. The types, not given explicitly but presupposed to be intuitively evident, are standard, echoing a Montagovian tradition. Conditions C are used in the antecedents of case constructs/if-then-elses, see φ_i below. Assertions are λ-terms without free variables. They can be combined by ∧ and are, e.g., properties of agents, embedded in conditions of the case construct of **Definition 2** or used to specify contexts of derivation.

The Ψ-calculus agents or processes, indicated by P, Q, …, are:

Definition 2:

0	Nil, the 0-agent
$\overline{M}$N.P	Output-agent
$\underline{M}(\lambda\tilde{x})$N.P	Input-agent
case φ_1: P_1 ⏸ … ⏸ φ_n: P_n	Case construct[15]
P | Q	Parallel (concurrency) operator between agents P and Q; P and Q can be executed independently or in parallel
!P	Replication; P can be repeated arbitrarily often
δ	Deadlock

The 0-agent is one being inactive. 0 is needed, e.g., if a local computation terminates.

[13] For example, we do not integrate mechanisms for hiding (encapsulating) information and for scope extension. We also do not deal with foundational matters in terms of Structural Operational Semantics.

[14] Here we keep to the wording of Bengtson et al. 2011, pp. 4-14, and to their use of "Definition".

[15] Instead of case constructs we often argue with conventional if-then-elses in the following.

Two agents implement symmetric channel communication:

"$\overline{M}$N.P" (M overbar, N dot P) puts a data structure N, more precisely, N's value, onto output channel $\overline{M}$, sends it out and continues with process P, possibly a 0 process, if no "transportable" material remains. "$\overline{M}$N." is informally also called "prefix" and similarly for "$\underline{M}(\lambda\tilde{x})$N." in the next line.

"$\underline{M}(\lambda\tilde{x})$N.P" indicates that a datum is received on the input channel $\underline{M}$ and substituted for $\tilde{x}$ in N and P.

"$\underline{M}(\lambda\tilde{x})$N.P" will be further commented upon below, where we discuss data inputs for an actual example. Also, note that the "λ" here is part of the Ψ-object language. Therefore, it differs from the λ-operator in the transported language. $\tilde{x}$ is a sequence of bound variables (bound by λ). The syntax device "." in, e.g., "$\overline{M}$N.P" separates the prefix "$\overline{M}$N" from the subsequent process "P". "." also functions as a scope marker. We frequently leave out ".", if no misunderstanding can arise.

In the case construct "case φ_1: P_1 ⏸ … ⏸ φ_n: P_n" alternatives are indicated by " ⏸ "; one alternative P_i is chosen given that φ_i is true. The case construct is also used to model the non-deterministic "either or" and the "if then else". "if then else" is captured by case φ_1: P_1 ⏸ $\neg\varphi_1$: P_i; a generalized "either or" results, if more than one φ_i holds.

The parallel operator "|" enables P and Q to expand independently or to communicate with each other *via* output and input operators, perhaps after several independent expansions. For the communication case across "|" we have, e.g.

$\underline{M}(\lambda\tilde{x})N.P \mid \overline{M}N'.Q \rightarrow N.P[N'/x] \mid .Q,$

where "→" signifies the transition relation of the operational semantics. Here, on the front side, N' goes on the output channel $\overline{M}$ (right side of "|") and enters the input channel $\underline{M}$ substituting for free x in N and P, hence, N.P[N'/x]. On the right-hand side of "|".Q remains.

Example:

In our application in Sec. 5[16] we have, for example,

Agent(1): $\underline{ch}_1$ ent-i λF$\overline{!\,ch}_{4a}$ ιx(hedge'(x) ∧ green'(x) ∧ F(x)).∃x(hedge'(x) ∧ green'(x) ∧ F(x)).0 (ent-i).

This embodies the information *hedge'(x) ∧ green'(x)* and can receive information *via* its input channel $\underline{ch}_1$. Assume, e. g., that some Agent(2) provides on his output channel $\overline{ch}_1$ the information *with-doorways'*. Since Agent(1)s input channel and Agent(2)s output channel correspond, indicated by the subindex 1, they communicate. This leads to the substitution of the Ψ-name ent-i with *with-doorways'*. Now, λβ-conversion applies and we get the information *∃x(hedge'(x) ∧ green'(x) ∧ with-doorways' (x))*. Agent(1) can now send the definite information *ιx(hedge'(x) ∧ green'(x) ∧ with-doorways'(x))*[17] in its prefix out *via* its output channel $\overline{!\,ch}_{4a}$. Indeed, it can do so arbitrarily often. On that see the explanation to Replication !P below and the Sec. 5.5 Broadcasting of Information.

Replication !P is understood as equivalent to P|!P, meaning that P can be emitted arbitrarily often, and can for example be broadcast down across subsequent turns in descriptions of multi-modal dialogue. So, !P is a technical means to model trans-turn communication, such as for example, involving multimodal anaphora resolution. We will, for example, use !$\overline{ch}_3$ λx(arch'(x)), to indicate the arches of doorways provided by gesture which are referred to several times in the datum in Figure 2.

[16] The example in Sec. 5 has a few downward and upward recursive steps and is therefore slightly more complicated.

[17] The definiteness information is based on the revised Chastain rule motivated in Sec. 5.1. However, a DRT-based argument would lead to a similar result.

The deadlock δ is taken from Fokkink (2000, pp.7, 25) and used here merely terminologically[18] instead of the Frege ⊥ in the original Ψ-calculus papers. In our application, deadlock δ could capture semantic inconsistencies, as among a property ball' and a property square', excluding square'∧ ball'.[19] δ's difference from 0 is that 0, being an idle agent, marks a non-action but does not indicate a fail, therefore it does not impede the process of semantic computation.

In addition to the agents, we also have operators. The equivariant ("equivariance" defined by α-equivalence[20]) operators are as in **Definition 3**:

Definition 3:

$\leftrightarrow: T \times T \rightarrow C$	Channel Equivalence[21]
$\otimes: A \times A \rightarrow A$	Composition
$\vdash\ A \times C$	Entailment

As remarked, for our descriptive needs we will use a reduced version of the Ψ-calculus, focussing on the tuple <0, $\overline{M}N.P$, $\underline{M}(\lambda\tilde{x})N.P$, case φ_1: P_1 ‖ … ‖ φ_n: P_n, P | Q, !P, ".", δ> in the following. Following Bengtson et al. (2011), "⊗" is implemented as "∧". The transitions of operational semantics used (see "→ " above in the formula $\underline{M}(\lambda\tilde{x})N.P \mid \overline{M}N'.Q \rightarrow N.P[N'/x] \mid .Q$) will be captured by informal descriptions dealing with the change of states as in the above example. 0-agents will often be omitted.

4.3 Data and Algorithmic Handling Matched

We now look back at the description of the datum in Sec. two, where we explained some of its features and take up the terminology coined there: Our modelling instances are now λΨ-agents. Agents can be seen as small programs which receive and send information. Independence of gesture and speech is captured by establishing different communicating agents for them. Concurrency of gesture and speech can be handled with the |-operator ranging over agents. Sequentialization is achieved by typing of information transfer. For transfer of information we have the input-output-device between agents. Compositionality of λ-data transported and the interface between Ψ-variables and λ-variables such as ent-i and F in the example for Agent(1) above, guarantee compositionality among the finally assembled multi-modal meanings. In other words, we work with a two-tiered approach captured by Ψ's transporting agents (the first tier) and the data transported with these (the second tier; see **Definitions 1** and **2** above).[22] Persistence or repetition of information across contributions is modelled with the !-operator: Information sent *via* ! can be received by a type-conforming input channel or blocked, i.e. if no fitting input channel is available; this serves as the major means for dealing with temporal gesture-speech (a-)synchrony. Finally, natural language conjunction, conditional and ⊥ have corresponding operators in Ψ (see **Definitions 2** and **3**).

---

[18] Meaning that we do not use Fokkink's process algebra version here, which we, anyhow, view as an alternative to the Ψ-calculus.

[19] To really achieve this, we would have to enlarge our λ-component investing words with a fine structure allowing agents to check for compatibility. Incompatible readings would then have to be marked with δ.

[20] α-equivalence means substitution of bound variables in formulas.

[21] Taking up a suggestion of Cooper and Larsson's (pers. comm.), we notate channel equivalence by identity sub-indexing, e.g., ch_i, instead of the axiom $\leftrightarrow: T \times T \rightarrow C$. This has repercussions for the use of the calculus in the derivations in Sec. 5. The most recent work on channel connectivity we are aware of is Pohjola (2020).

[22] Two-tieredness is, however, not an absolute necessity, as the concurrent λ-calculus of Dezani-Ciancaglini (1996) shows.

Arguments for the λΨ-calculus

The Ψ-calculus has been chosen both for theoretical and practical reasons, see the main reasons listed and the motivation for them below. It has (a) concurrency operator |, (b) a replication operator ! (c) a case construct, it entertains (d) a channel notion, (e) input-output facilities, (f) it can transport any higher order logic formulas. In more detail: (a) Some concurrency operator must be used because of gesture-speech concurrency which cannot be captured using sequential modelling, the deeper reason being that multi-modal communication works in parallel. (b) The replication operator in tandem with (e) cares for the top-down distribution of information. (c) The non-deterministic case construct is useful for expressing natural language conditionals and the context-dependence of gesture (see Fn. 24). (d) Ψs channel notion does service for modelling the visual-auditory channel. (f) Ψ-agents can transport typed λ-structures. Footing on Operational Semantics (see, however, the restrictions of this paper mentioned in fn. 28 and fn. 14), the λ-formulas care for the composition of verbal meaning on the one hand and co-speech gesture meaning and speech meaning, i.e. multi-modal fusion, on the other hand. This suggests a move towards a unified semantics for multi-modal behaviour.

Next, we move on to the description of anaphora and broadcasting (Sec. 5) and its reconstruction in the λΨ-calculus.

5. Verbal Anaphora, Multi-modal Anaphora, Broadcasting of Information

5.1 Verbal Anaphora

We start our investigation of multi-modal anaphora with a closer look at the linguistic anaphora in our datum. We then will explain in greater detail which meanings[23] co-verbal gestures contribute to the linguistic anaphora and what the resulting multi-modal meanings will be. Anaphora in natural dialogue present problems different from those discussed in the computational linguistics or the Philosophy of Language literature: they extend over stretches of discourse and non-regimented anaphoric relations are extremely varied as we will see. A remarkably early tool to approach some of these observations is the nearly forgotten Chastain (1975), which also relates the anaphora problem to the general discussion of the reference of terms at issue (including Russell, Strawson, Quine, Donnellan, and Kaplan). We accept the following of Chastain's hypotheses[24] with modifications, which we indicate in a. to d. below:

(a) Anaphora occur not only in pairs but in anaphoric chains as in *Eingänge/doorways - Die/these - der Haupteingang/the main entrance - das/it - da/there - in den Bogen/into the arch - in den Bogen/into the arch - dort/there* (1.1-G - 1.2-G - 1.3-G - 1.4-G – 1.5-G 2.1-G - 3.2-G) (p. 204).

(b) Indefinite descriptions starting an anaphoric chain are not taken as existentially quantified expressions as, for example, the later DPL and DRT accounts would essentially suggest,

[23] We agree with reviewers A, B, and C that gestures can only be interpreted in context. We dealt with the context-dependence of gesture in Lawler, Hahn, and Rieser (2017) "Gesture Meaning Needs Speech Meaning to Denote. A Case of Speech-Gesture Meaning Interaction". There, the idea is, using Ψs Case construct, to make the meaning of a gesture conditional on some VR object(s) introduced by speech. So, for example, the O-gesture as shown in Figures 11a, 12a, and 13a would depend for its meaning on *pond/Teich*, the q-shape in Figure 13 on its verbal context *arch-shaped, like such a garden entrance/bogenförmig wie so'n Garteneingang*. In this paper we assume that a context-dependent meaning of the gesture has been computed using the Case construct device. The gesture meanings supposed in this paper follow the decisions of the SaGA corpus annotators.

[24] Unfortunately, we cannot do justice to the philosophical insights of Chastain (1975) here. Page numbering is given according to Chastain (1975).

but as singular terms with a definite reference in order to account for definite uptake (pp. 206-215).
(c) The descriptive content of singular terms can be enlarged over time in the discourse (p. 232).
(d) In chains multiply linked terms can exist, i.e. different anaphoric chains can result (p. 267).

Now, as to our modifications: (b) is too strong and goes against the tradition in logic as, e.g., originally defended in Reichenbach (1947, §§ 42, 47). We will weaken it in the following way, using Ψ's agent notion for dynamic processes and information transfer.

a. An indefinite description can send a singular term for further anaphoric use.
b. Definite terms can always be sent to the outside.
c. The terms of an identity can be sent to the outside and computation can continue with either term.
d. Extending standard Ψ-calculus practice, we will allow arrays in the output prefix indicated with "< …>", the fields of which can be accessed as usual. Arrays are a source of non-determinism. They mark a sequence of possible dialogue continuations.

The sending and receiving notion is bound to the Ψ-calculus definitions (see **Definition 2**, Sec. four). It will be demonstrated in more detail in the application Sec. 5.4.
Mainly due to the route description dialogues of the SaGA corpus, land-marks, objects, definite and indefinite locations, routes, and trajectories are reflected in anaphora formation as shown in (a). (b) states a precondition for explaining the transition from the first indefinite member of an anaphoric chain to a following definite one; clearly, an existentially quantified expression cannot be equivalently substituted by, say, a definite description in general, which has been a main reason for introducing discourse referents in DRT. Observation (c) allows for dynamically extending the meaning of successive members of an anaphoric chain. As a consequence, a definite description sent in the manner of (b) can contain a slot for further accumulating information; we, however, will only indicate an extension informally and avoid slots to avoid cluttering up the formalism. Finally, (d) specifies different antecedents in anaphoric chains. We have to wait for applications of (d) until Secs. 5.4 and the Appendix 1.2 (see the function of Agent(1) there).

Remark on NL syntax: As the dialogues in SaGA show, there is an intimate connection between antecedents, anaphora, German word order and information structure management. In order to describe that we informally use the so-called field theory of the German sentence in the following manner[25]:

prefield	sentence bracket enclosing the middle field[26]	postfield
Aber du	*fährst um den Teich herum.*	*Rechts herum.*
But you	*drive around the pond.*	*Right around.*

The sentence bracket is given by the finite and the infinitival parts of the verb, *fährst/drive* and *herum/around*; the postfield is taken by the local adverb *rechts* and the factorized *herum*. The middle field is underlined. So much for the syntactic distinctions which will be addressed.
We now successively walk through the antecedent-anaphora relations in the datum. 1.1-G: *a hedge, a green hedge. Partly also with doorways.* Here we see an extended indefinite *a hedge, a green*

[25] For more information on the field theory of the German clause, see Höhle (1986) and Wöllstein (2018).

[26] The most extensive study we know of the German middle field is Haider (2006).

hedge. Partly also with doorways. We have a rhetorical anaphora *a hedge* with two elaborations in the repetition. The indefinite *a hedge, a green hedge. Partly also with doorways* is the first member of an anaphoric chain. Its descriptive content is *hedge, a green hedge. Partly also with doorways.* Remember that the second member in a chain must be marked as a singular term. Working with the Kamp and Reyle (1993 p. 305-482) plural classification, we think that the whole NP can be classified as either dependent plural or a case of distribution Δi[27] over which we can distribute in order to capture the plural *Eingänge/doorways.* This is also taken into account in our λΨ-reconstruction which specifies these two options (see Agent(4) in Appendix 1.1).
In order to shed some light on 5.1 a. above, we turn to *eine Hecke/a hedge – die Hecke/the hedge* (1.1G/E-5 G/E). A first order representation would render this pair as $\exists x\ hedge(x)$ - $\iota x\ (hedge(x))$, the first being a general and the second a singular referential term. The anaphora intuitions for cases like these are clear: the singular term depends on the general term. Hence everything hinges on the explication provided for "depends". We assume that once an object has been introduced by existential quantification, it can subsequently be taken up by a referential term. The new method used for implementing this intuition is the output channel device of the λΨ-calculus (see Sec. 4.2, **Definition 2**) working as follows: Suppose we have an agent transporting $\exists x\ hedge(x)$ …. we use λΨ's output channel $\overline{ch}_i$ and a definite value $\iota x\ hedge(x)$, in this manner $\overline{ch}_i\ \iota x\ hedge(x)$**.** $\exists x\ hedge(x)$ …. It then transports $\iota x\ hedge(x)$ in the prefix to the outside. $\iota x\ hedge(x)$ can then be received by an input channel $\underline{ch}_i$ (subscripted "i" to indicate identity) of some follow-up agent invested with $\underline{ch}_i$ and enter into compositional relations with it. The anaphorical port, that is the representation of a pronoun, definite description, local adverb or whatever, waits for information from above.

5.2 The Contribution of Gesture Meaning to Anaphora Meaning

In this section we discuss the co-verbal gestures in the datum. Remember Figure 3 for the ASL-shapes and their natural data occurrences.

The Arch Information by Gesture

Beginning with *a hedge, a green hedge*, we have a sequence of RH gestures moving from loose C over small C to a classifier for "thin". Loose C is often used to support

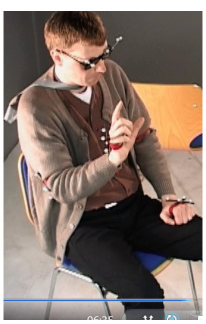


Figure 11. *loose C*

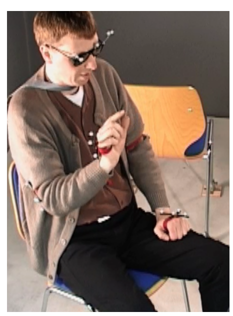


Figure 12. *Small C going to "thin"*

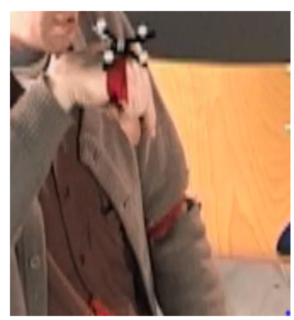


Figure 13. *q-shape*

a height or width indication for objects but this does not match well with small C and the classifier for "thin" here. It could well be that this sequence is due to the motor planning of a preparatory sequence for the R-Handshape q which is basically a rotated version of a sloppy thin classifier. Figures 11-13 represent the gesture sequence in the video V5 datum. The handshape q is aligned with *partly also with doorways*. For all we know from SaGA, an equally plausible interpretation would be to regard the loose C-shape as an interactional gesture, as an attention seeking device

[27] For distribution we use a distribution operator Δ and the Kamp-Reyle *.

more often expressed by loose G. However, the q shape can be seen as modelling two vertical trajectories connected by an arc (see Figure 13). Considered this way, the handshape provides additional information, since then it is about the round shape of the top side of the doorways as against a possibly straight shape. The gesture works in parallel with *partly also with [doorways]*. Note that in order to be semantically appropriate, it has to skip over *partly also* and combine with *doorways* achieved in the formal reconstruction *via* an assumed typing of the output-input facilities. The content of q denotes an arch property. So, we have a concurrent interaction of several agents as detailed in Appendix 1.1.

5.3 Extracting Agents from the Video Data

As already emphasised, the formal machinery is based on the agent idea of the λΨ-calculus. The identification of agents proceeds from the multimodal SaGA datum according to the following guide-lines:

(1) Gesture and speech yield different agents. Gesture information is marked in red for easier identification: λx(arch'(x)), pond'_{LH}, right-angle' etc. which represent, respectively, gesture meaning arch', pond' generated by the left hand and gesture meaning right-angle'[28]. Supra-segmental phonological marking of structures is taken as an indication for identifying agents. These can be

a. pauses indicated by "." and "," in the datum transcript: *eine grüne Hecke. Zum Teil auch mit Eingängen/ A green hedge. Sometimes with doorways.* (1-G/E) This is modelled, e.g., by interacting Agents (1), (2), and (4) in Appendix 1.1,

b. final rises or level tone indicating continuation: *ein Fluß über den auch Brücken führen/a river crossed by bridges*. The level tone is on *Fluß/river*. The final tone is on *führen/crossed*,

c. final falls indicating termination of chunk of speech in the audio data: *wie so n' Garteneingang/ like such a garden entrance.* See decision for Agent(7) in Appendix 1.1,

d. material in the prefield: *Und an diesem Teich./ And at this pond.* Observe that there is also a pause, marked by "."; represented in Agent(24) in Appendix 1.3,

e. material in the postfield: *rechts herum/right around it*. Cf. Agent(27), Appendix 1.3.

And similarly for

(2) parentheticals: *scharfer rechter Winkel/sharp right angle* (1.4G/E) with level tone; represented in Agent(13), Appendix 1.2 .

Finally, decisions due to the neo-Davidsonian verb-frames and considerations for coming as close to a first-order language as possible can also provide motivation for postulating agents: *Teilweise auch/Partly also with* and *mit Eingängen/with doorways* (1.1-G/E) yield Agent(2) and Agent(4), Appendix 1.1.

5.4 Anaphora Resolution in the λΨ-calculus

We now turn to some of the multi-modal information of the Route-Giver's and the FOLLOWER's dialogue contributions in the datum Fig. 2. We do not think that the semantic values of gesture and speech differ metaphysically. Nevertheless, we want to indicate the source of the meaning produced, in order to point out which source contributed which meaning; therefore, as mentioned earlier, we colour gestural meanings. Coloured and non-coloured meanings are fused into a multi-modal meaning as discussed in Sec. 3.5. In the end, colouring will give us a rough idea of which

[28] In this paper we do not provide a model-theoretic semantics for gesture and speech but stay at the level of semantic representation to avoid unnecessary length and complexity. If we did, red expressions would be given the same type of semantics as the black ones. This also follows from the idea of a unified semantic representation as developed in Sec. 3.5 *Multi-modal Fusion*.

objects were gesticulated, i.e. of the contribution of gesture to ontology. This is perhaps also a proper place to add a caveat as far as the meaning representation for iconic gestures is concerned. What we get through corpus-annotation are essentially constellations of trajectories in R^4 (4-dimensional reals), non-standard topological entities with their (non-standard) *sui generis* denotations as geometrical objects. In practice, there is a long way, for example, from two separated orthogonal surfaces in R^4 connected by an arc-trajectory, as in 1.2-G/E or in 1.3-G/E, to an architectural arch-concept. Here, we fix the relation on an observational basis substantiated by the VR-representations of the scenery got from the SaGA corpus (see the Route-Giver's tour and landmarks in Fig.1).

A complete representation of the Agents and their interaction is provided in Appendix 1; the numbering of Agents in main text matches the numbering in the Appendix. For detailed information it is useful to consult the Appendix 1.

The Arch Information

In this section the focus is on

- communication and exchange of information according to Ψ-calculus definitions,
- the communication of gesture agent and speech agent, and
- anaphora resolution with the revised Chastain anaphora rule
- plural readings of NPs and plural anaphora.

Anaphora resolution, following the revised Chastain rule in Sec. 5.1, proceeds in a top-down fashion, originating from a source which is information on some output channel and "looking for" an input channel where it can hook up. This process might be one-to-one, i.e. one output channel linking up with one ready input channel or it might be distributive, i.e. one source information linking up iteratively with several input channels, indicated by "!". The latter cases are mainly treated in the broadcasting section.

Examples:

Agent(8) gets information on input channel $\underline{ch}_8$ and instantiates the Ψ-name *m-entr* with it. The value is handed on to the λ-variable *p* by λβ-conversion, so the information $\underline{ch}_8$ *m-entr* is in place. Agent(8) now sends the main-entrance information *ιv(main-entrance'(v))* on $\overline{ch}_9$ to the corresponding Agent(10),

Agent(8): $\underline{ch}_8$ m-entr $\overline{ch}_9$ ιv(main-entrance'(v)). λp until-sometime-then'(p) (m-entr).0 |,

where it is integrated via the Ψ-name *m-ent.*

Agent(10): $\underline{ch}_9$ m-ent λu∃y((large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) ∧ y = u) (m-ent) .0 |.

Ψ-names are arbitrary. *m-entr* and *m-ent* were chosen to indicate that they belong to different agents. However, it would not do any harm to use either *m-entr* or *m-ent* in both places, since they are substituted by the values transported by $\underline{ch}_8$ and $\underline{ch}_9$ which are different.

Agent (22) below, after having got information from channels $\underline{ch}_{15c}$, $\underline{ch}_{14}$, $\underline{ch}_{11}$, $\underline{ch}_{18}$, $\underline{ch}_{19}$, communicates ιx(pond'(x)), i.e. the description *the pond*, or the pond gesture semantics ιx($pond_{LH}$'(x)) (see revised Chastain rule in Sec. 5.1) distributively via !$\overline{ch}_{17}$:

Agent(22): $\underline{ch}_{15c}$ pa-i $\underline{ch}_{14}$ veh-i $\underline{ch}_{11}$ the-i $\underline{ch}_{18}$ po_{LH} $\underline{ch}_{19}$ po_{RH} λpth2 λveh λthere ! $\overline{ch}_{17}$ <ιx(pond'(x)), ιx($pond_{LH}$'(x))>. (drive'(e2, FO, veh, pth2, there) ∧ straight-forward'(e2, veh, pth2) ∧ towards'(e2, FO, veh, λrs ∃x (pond'(x) ∧ x = r ∧ x = s)))(pa-i)(veh-i)(the-i)(po_{LH})(po_{RH}).0 |

Similarly, see also the pond-information of Agent(24): !$\overline{ch}_{20}$ $pond'_{LH}$. at'(ιx (pond'(x)) = $pond'_{LH}$).0 |. This provides also a good example for observing that gesture semantics and speech semantics are treated as being on a par.

Following the incremental sequence, we start the formal rendering with the sequence of agents

Agent(1) | ,……, | Agent(53),

where "|" indicates concurrency according to **Definition 2**[29]. So, *a green hedge* is the first agent and *There was an/the ice-cream man's cart which had a handle bar (horizontal grasp pantomime) and an umbrella with vertical handle (vertical grasp pantomime)* the last one. The numbering of channels is arbitrary. What matters is their identity and the role they fulfil for Agents(*n*). Due to channel communication and assumed typing regime, we can sequentialize agents, see for example the interaction of Agent(1), Agent(2) and Agent(4) in Appendix 1.1 which prescribes an execution order. Initially, the algorithm starts with an Agent in need of information to be provided by some other Agent further down the line (see Agent(1)) or by an Agent providing some information for one or several other Agents to come. The receiving agents wait for information without which they are blocked. This is similar to an algorithms entrance condition familiar from programming. So, we get in the end a communication net in which the agents depend on each other in terms of information. Apart from occasional idle 0-Agents, only two agents, 37 and 16, do not communicate. Looking at the execution order, Agent(1) needs information via the input channel $\underline{ch}_1$; the Ψ-name ent-i receives a value sent from some corresponding output channel $\overline{ch}_1$. Agent(2) will be able to send on $\overline{ch}_1$ but needs first information on $\underline{ch}_2$. Now, information seeking percolates down, Agent(2) must have some information on $\underline{ch}_2$. This is provided by Agent(4). Here we deal with a parallel agent indicated by | as introduced in Sec. 4, **Definition 2**. One side can provide an output via $\overline{ch}_2$, once it received a value on its $\underline{ch}_3$ input channel served by gestural Agent(3) modelling the information of q (see Figures 3 and 4 for the datum q). The q information for small arch provided by orthogonal sides and the connecting arc is implemented as the arch-information which can be iterated arbitrarily often due to "!" (see Sec. 4, **Definition 2**).

A closer look at the datum *a green hedge. Partly also with doorways* reveals that we have concurrently a dependent plural reading of *doorways*, intuitively, the hedge's doorways, as well as a distributive reading indicated by Δi since either could obtain. The information flow branches non-deterministically; only the right-hand side information is considered in the following.[30] $\overline{ch}_2$ $(\Delta i\ (doorway'(i) \wedge of'(i,a) \wedge \lambda x(arch'(x))(a)))$ finds Agent(2) for communication; Agent(2) ends up with an output $\overline{ch}_1$ ready to interact. Communication among Agent(2) and Agent(1) attributes to any member of a set the property of being a hedge, green and partly also with arched doorways as follows:

$$\exists x(hedge'(x) \wedge green'(x) \wedge (partly\text{-}also'(with'(x, \Delta i\ (doorway'(i) \wedge of'(i,a) \wedge arch'(a)))))) \mid .$$

We let Agent(1) send out a definite information according to the revised Chastain rule for indefinite introduction and definite resumption (see Sec. 5.1, a.). We use an array with two fields (see generalisation in Sec. 5.1). The fields read as *The green hedge, partly also with arched doorways* and *the arched doorways*, respectively. Both have been derived and can be replicated independently to fill anaphora slots to come.

The *die/these* Anaphor, the Arch Gesture and the Park Entrance

In this section we deal with

- co-operation of handshape semantics and speech semantics,
- plural descriptions incorporating gesture representations based on the Kamp-Reyle plural distribution,
- non-deterministic ("upwards branching") anaphora,
- isolated multi-modal agent information,
- contradiction of gesture meaning and speech meaning,

[29] Obviously, we have, concerning the multi-modal dialogue extract, a sequence (except for speech-gesture mappings) however, remember that sequentialization can be forced by the typing regime.

[30] We could also work with the left-hand side. It would not matter much for our example.

and on the dialogue side,

- Follower's semantics,
- dual function of Route-Giver's gesture,
- Follower's clarification requests and Route-Giver's answers,
- Follower's decision for an anaphora antecedent.

These in 1.2-G/E takes up the doorways already introduced within the antecedent. On these a bit later. We have a semantic cooperation of *arch-shaped* and the R-Handshape small-arch-shape, where "small" is due to ASL q (see Figure 4) and the little information it covers in gesture space. We let it denote the constant small arch. As we shall see below, Agent(6) embodies the denotation of the R-Handshape small-arch-shape information and cooperates with the verbal information *like such a garden entrance* provided by Agent(7). Agent(6) can then output something like "small arch like such a garden entrance". Agent(5), representing the meaning of *so like arch-shaped* fuses first with "small arch like such a garden entrance" and then with the input doorway information. Using Kamp-Reyle style plural distribution[31], the result of the computation is informally "The doorways with their arches are arch-shaped, all of them [by distribution] have a small arch, and they [by distribution] are like such garden entrances".

The doorways are taken up by the anaphora *these*. Agent(5) sets the scene for dealing with the anaphora relation, receiving the value Δi (doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a)) from Agent(1)'s output. Agent(5) needs input on channel $\underline{ch}_5$ for the comparison clause *like such a garden entrance*. Agent(6) and (7) cooperate to provide that. Agent(7) implements the comparison. [32] The information sitting on $\overline{ch}_6$ can be received by Agent(6). $\underline{ch}_{4a\ second}$ transports the set of arched doorways, Δi (doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a)), a definite set; its representation can, hence, be used like a singular term. So, the final derivation for Agent(5) is:

like'(arch-shaped')(Δi (doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a))) $\wedge$ (of'(Δi (doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a)), small-arch') $\wedge$ like'(such'$\exists$U *garden-entrance' (U))(Δi (doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a)))).0 |

This means that each i in the set operated on by Δ that is an arched doorway, i.e., is of'(i,a) $\wedge$ arch'(a), is like'(arch-shaped') [first conjunct], has a small-arch' [second conjunct] and is like such a garden entrance [third conjunct]. Some of this information will be sent out by $\overline{ch}_7$ specified below given that it finds a corresponding input channel. So, we have solved the anaphoric relation among *Eingängen/doorways* and *die/these*; observe that *die/these*, the anaphora, is represented as Δi(doorway'(i) $\wedge$ of'(i,a) $\wedge$ arch'(a)), i.e. as the set of arched doorways.

Next we consider *Until sometime then the main entrance comes. Main entrance* might be an anaphora candidate for *doorways.* So, the main entrance is in the set of doorways which are arch-shaped, have a small arch and are like such a garden entrance, as derived. The main entrance has to be tied up with this information in Agent(9). *That* then takes up the definite information "the main entrance is the arch-shaped object, having a small arch like such a garden entrance" as sent by Agent(9). Agent(10) specifies that it is large, grey and has a curb (see Figure 6 repeated).

[31] I owe the argument that the example cannot be represented using the Kamp-Reyle Σ-operator (which is an abstraction operator) but must use (an equivalent of) the Kamp-Reyle * to Andy Lücking.

[32] For the * see Kamp and Reyle (2003), p. 327, fn 13. We use Δ as bare plural operator and * for predicates which can have either singular or plural denotation.

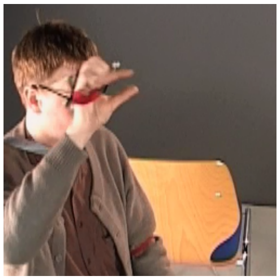


Figure 6. *R-handshape large arc-shape (by an arc drawing gesture) + small c for curb*

Figure 8. *G drawing trajectory right angle (route into main entrance to park)*

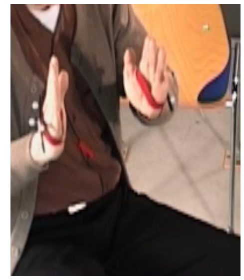


Figure 9. *Modelling (sides of) bus*

The extent information concerning the arch is provided by the gestures q *vs* an R-Handshape arch-shape and the adjective *large*. Assembling the information, we get a contradiction of small *vs* large[33]. The curb-information comes from the R-Handshape arch-shape plus the drawn small C information indicating an architectural frame structure (1.3 G/E). It is embodied in Agent(10). Agent(10) sends the information "the large grey arch with the curb which is the main entrance" to the outside. It is taken up by the anaphora *there* in *And there must you quickly, sharp right angle, to the right into.* and modelled in Agent(12), see Figure 8 repeated for the *sharp right angle*.
By theory, the anaphora *there* can hook up non-deterministically to either the main entrance or the arch. This is due to the fact that a new definite can be created out of the NP in the predication. As the clarification request below shows, the Follower takes *a large grey arch* as the relevant antecedent. As we see in the formal rendering in the Appendix, this has major repercussions for the anaphora analysis. In the datum 1.4-RH we also observe a dual function of the Route-Giver's RH gesture: it models the moving vehicle and the route to be taken into the arch (see Figure 9 repeated). Both iconic gestures continue up to the end of the Route-Giver's answer to the Follower's clarification request in 2.1-G/E. Both, clarification request (Agent(17)) and answer (Agent(18)) are highly elliptical and get their information from Agent(12), multi-modally specifying the way of moving into the arch.
Now we have to implement the anaphoric relation of *the main entrance*: The main entrance must be in the set of doorways having a small arch and being like such a garden entrance. Agent(8) provides a simplified temporal prefix for *Until sometime then* and will in the end have to send out *the main entrance* information after input on $\underline{ch}_8$.
Agent(9) submits the proposition *the main entrance comes* and expresses that the main entrance is one of the arched doorways which has a small arch and is like such a garden entrance. Agent(5) is active on channel $\underline{ch}_7$ and provides the necessary input. The information derived is sent to the outside. *The main entrance* is *one of the arched doorways, arch-shaped, has a small arch and is like such a garden entrance*. Agent(10) cares for *That is a large grey arch.* (1.3-G/E). The anaphoric indexical *That* is resolved to *ιv(main-entrance'(v))* by information transfer of either $\overline{ch}_8$ or channel $\overline{ch}_9$: We have anaphora non-determinism, an upward branching anaphora as remarked in Sec. 5.1 (d). Agent(10) embodies the {R-Handsh. large arc-shape + small c, Figures 6 and 7} information represented as curb'.
We are now ready to deal with the move into the park. Agent(12) cares for the verbal structure of the contribution (1.4-G/E) *And there must you quickly, sharp right angle, to the right into*. Its multi-modal structure is schematically set up as follows:
(must-in(current event, Follower, vehicle, path taken, destination) $\wedge$ F(current event)),

[33] From the observer's point of view, contradictions are quite frequent in the SaGA corpus. They are either overridden or (hardly ever) negotiated by the participants.

where "must-in" represents informally a modal+verb-combination, "current event" the current event of the route, "Follower" the Route-Giver's addressee, "vehicle" the bus, "path taken", the trajectory followed by the bus, and "destination" the current event's endpoint. The gesture meaning which has to be integrated is {G, drawing trajectory right angle + modelling bus}[34], as shown in Figures 8 and 9.

In more detail, Agent(12) gets information from various communicating sources: Agent(10) supplies the main entrance information. Agent(13) provides the gesture information for right angle, sent on $\overline{ch}_{15}$: ! $\overline{ch}_{15}$ right-angle'.0 |. Agent(14) contributes the vehicle gesture on $\overline{ch}_{14}$: ! $\overline{ch}_{14}$ vehicle'.0 |.

The right-angle information and the vehicle information, both frequently distributed in the datum, are replicated, "!".

Agent (15a) communicates event information and adverbial information on $\overline{ch}_{15b}$.What is represented is roughly, "you must quickly [go] directly into the large grey arch with curb [which is the main entrance]".

We now come to the Follower's clarification request *In den Bogen rein?/ In the arch into?* represented in Agent(17). The problem that arises is "What is the semantics of the Follower's notion of *arch*"? We introduce a test on the Follower's side for a specific interpretation: *ιu(arch(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))* expressing "The arch is the large grey arch with curb". The question sends the information to the outside and Agent(18), encoding the answer, takes it up. So, we have for the Follower's clarification request Agent(17) interacting with Agents(10), for providing the main entrance, (12), for specifying the event e1, (14), for pinpointing the vehicle, and (13), for giving the route direction into the arch. Agent(17)'s output is

$\overline{ch}_{36}$ (ιu (arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))).0 |.

Agent(18), encodes the answer to the Follower's clarification request[35]. The feeding channels are as in Agent(17); the output value is the filled verb frame:

$\overline{ch}_{37}$ into'(e1, FO, vehicle', right-angle, in' (ιu(arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))). 0 |.

Channel $\underline{ch}_{15}$, the path information indicated by the Route-Giver's gesture, ! $\overline{ch}_{15}$ right-angle', is active during the clarification request and the answer. Therefore, the clarification request is a cooperative undertaking between Route-Giver and Follower: In order to be sure about the direction of the bus route, i.e., the path to be taken, one needs the Route-Giver's right-angle' indication which works like an anaphora, in our account by top-down information transfer.

It is not clear why the Route-Giver remarks *It isn't a wall, it's a hedge* (2.1-G/E), since the bus must go through the arch which was just clarified. From a gestural perspective it is an interesting datum, since the front of the hedge is modelled with a vertical slanted two-handed ASL-flat B gesture, taken here as meaning front' (see Figure 16) and modelled with Agent(16a).

[34] Remember that blue marks indicate broadcasted information, the discussion of which is postponed to Sec. 5.5.

[35] For a KoS and TTR perspective on a similar datum see Ginzburg and Lücking (2021b).

Figure 16. *Left: VR-entry into park with large grey arch* *Right: two-handed gesture for hedge*

Figure 11a. *Pond first gesticulated by co-verbal LH and RH gestures with slightly different O-shapes*

So we end up with $(\exists x(\neg \text{wall'}(x) \wedge \text{hedge'}(x) \wedge \text{of'}(x, \text{front'})))$. No information is on an out-going channel showing the singularity of the Route-Giver's contribution[36].

The anaphoric chain cited in the context of Chastain's anaphora theory in Sec. 5.1 is modelled with Agent(1) (*doorways*) – Agent(5) (*these*) -Agent(8) (*the main entrance*) – Agent(10) (*it*) – Agent(10) (*there*) – Agent(17) (*into the arch*) – Agent(15a) (*there*).

5.5 Broadcasting of Information

We have so far met gestural information about objects, events, routes and directions which have been freely replicated one-to-one across dialogue contributions. The idea is that information of this sort is output, received and integrated somewhere in the dialogue on a single output-target basis. We now turn to an example where information transfer is generated by a continuous post-stroke hold. The existence of these post-holds in SaGA is one of the main observations dealt with in the present paper. It is relevant likewise for empirical gesture research (such as gesture annotation or experiments concerning the relation of gesture and pitch accent as reported in Pow and Dixon (2019) or Navarretta (2021)) and its algorithmic rendering (e.g. the obvious need for an iterative/recursive information emitting device). The formal device to capture the broadcasting of information is the "!" operator as introduced in **Definition 2**. The !-operator can emit the information in its scope, which is an output pre-fix of a formula, arbitrarily often. It is handed on, if it finds a fitting input channel. However, even after this act, "!" goes on to send the information in its scope; it further emits the information for concurrent or upcoming input channels, if any. The behaviour of "!" can be compared metaphorically to a top-down impulse.

A first case of broadcasting is found in 1.4-RH, 1.5 RH, 2.1 RH (blue marks) where the Route-Giver signs the route through the main entrance into the park. Signing continues from the Route-Giver's first directive up to the end of the Follower's first clarification request. It lasts as long as the route into the park is at issue. In the formalism this is reflected in the right-angle information of Agents(12), (13), and (17). However, the main occurrences of broadcasting in the datum are found in 3.2-G to 10-G, tied up with the pond-information, see the blue marks in the excerpt of the datum below.

First, briefly, again about the move into the park. In the conditional's antecedent in 3.2-G/E *If you have driven in there* we have *there* anaphorically related to *arch* in 2.1-G/E. Two parallel trajectories with flat LH and RH ASL shapes B indicate again two things, the vehicle's movement (distance between both hands for the vehicle breadth) and the path into the arch (distance between

[36] Nevertheless, in SDRT terms it could perhaps be taken as an elaboration. It could therefore be placed on an outgoing channel but perhaps not taken up.

both hands for the track width, see Figure 9, (Agent(21)) and towards the pond (Agent(22)). This becomes ultimately pooled in Agent(20).

3.2-G	Route-Giver: 3.2.1 Wenn du dort eingefahren bist, 3.2.2 fährst du geradeaus auf einen Teich zu. Einen Teich.
3.2-E	If you have there driven in, you drive [{straight} towards a pond. A pond.]
3.2-LH	C; [B parallel trajectory] [B parallel trajectory] [L-Handshape O ctnd.]
3.2-RH	{B parallel trajectory } {R-Handshape C > open O ctnd }
4-G	Route-Giver: [4.1 Und an diesem Teich. 4.2 Du {fährst drauf zu und 4.3 du fährst rechts herum. }]
4-E	[And at this pond. You {drive towards it and you drive right around it. }]
4-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.ctnd. ctnd.]
4-RH	{R-Handshape D ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.ctnd. ctnd. }
5-G	Route-Giver: [Die Hecke, {die geht noch ungefähr} so 50 m.]
5-E	[The hedge,{it runs another roughly} 50 m.]
5-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd.]
5-RH	{R-Handshape loose B ctnd. ctnd. ctnd. ctnd. }
6-G	Route-Giver: [Und dann sind dort {auch} hin und {wieder} {Sitzbänke}.]
6-E	[Then there are {also} here and { there } {benches} .]
6-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
6-RH	{R-Handshape loose D} (first two overlaps) gesture expressing doubt: {wiggling of loose D Handshape} (third overlap)
7-G	Route-Giver: Ab[{er du fährst um den Teich herum. Rechts herum. }]
7-E	[{But you drive around the pond. Right around. }]
7-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
7-RH	{R-Handshape D ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. }
8-G	Route-Giver: 8.1 [Und manchmal ist da auch {nen Eisverkäufer }.8.2 Und an dem fährst du rechts ab.]
8-E	[And sometimes there is {'n ice-cream man there.} And at him you drive off to the right.]
8-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
8-RH	{R-Handshape G ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.}
9-G	Follower: Was heißt "manchmal"?
9-E	What does "sometimes" mean?
10-G	Route-Giver: 10.1 [Ja, könnte verändert werden. 10.2 Auf jeden Fall, {auf} meiner Tour war dort ein Eisverkäufer.]
10-E	[Well, could be changed. In any case {on} my tour there was an ice-cream man there.]
10-LH	[L-Handshape O ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd. ctnd.]
10-RH	{R-Handshape D, 3 beats/indexing location at embankment of pond}

11-G	Route-Giver: Ein Wagen mit Schirm.
11-E	[{A cart] with umbrella. }
11-LH 11-RH	[LH + RH pushing Pantomime] {RH modelling umbrella}

Figure 3: Broadcasting fragment of Figure 2.

As the datum 3.2-G/E shows, the pond (Fig. 12a) is first gesticulated by co-verbal LH and RH gestures with slightly different O shapes (Figure 11a). Since there is only one pond in the VR-setting acting as our model, their denotations must be identical and identical to the pond denotation expressed in speech. For antecedent and consequent of the conditional we have "If you have driven into the arch, you (next) drive towards a pond". "Drive" is underpinned with gestural information for vehicle, path (RH) and the pond goal (LH). This completes the information of Agent(22). It distributes the relevant pond information in speech and gesture by "!" replication; ! broadcasts the pond information up to 10-LH and beyond. 4-G/E is concerned with the pond and the path to reach it. Instantly, the function of the RH and the LH changes: The LH still broadcasts the pond information as before while the RH now signs the path towards the pond and then around it. We deal with a complicated conditional expressing "As for the pond, if you have driven towards it, you (next) drive right around it". The Route-Giver, after again focussing the pond, moves one step back in the route description and repeats *you drive towards it* and then adds *and then right around it*. *It* takes up *this pond* anaphorically by LH O and continues throughout the contribution 4-G/E. This way it also contributes the pond information for the elliptified *towards it* and *right around it*. The L-Handshape O is statically held on pond, while the R-Handshape D is drawing the route. Looking at the VR-rendering of the pond and its environs (see Figures 12a and 14), we see that we get the correctly gesticulated information related to the pond and the route around it, in other words, gesture information seems truth functional[37].

Figure 12a. *Left: The path into the park through the main entrance (arched main entrance from the street into the park at the bottom of slide), around the pond and out of it. Right: The RH draws the path towards the pond and then around it. The L-Handshape O is statically held while the R-Handshape D is drawing the route towards the pond and around it.*

Agents(26) and (27) continue the gestural information for vehicle and path. In 5-G/E, "The hedge, it runs another roughly 50 m", the Route-Giver's left hand still continues to hold the pond

[37] See, however, fns 33 and 39 which call truth functionality into question.

representation while the RH gesticulates, using loose B, the distance covered by the vehicle; the speech-gesture overlap is *it runs another roughly*. The Route-Giver continues to hold the pond information with LH in 6-G/E, *Then, there are also here and there benches*.

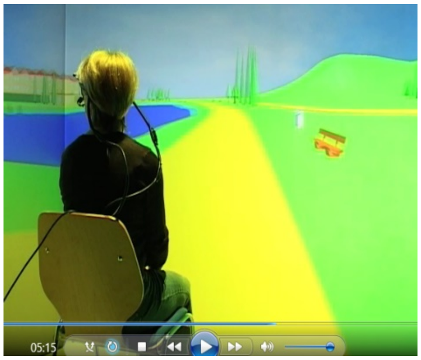


Figure 14. *The VR-ride towards and around the pond; one of several benches*

Figure 13 a. *Indexing location of ice-cream-man at bank of pond. Observe that LH is still holding the pond.*

The pond gesture continues to be held across the Follower's new clarification request *What does "sometimes" mean?* (see 9-E). It only stops when the cart of the ice-cream-man is gesticulated by a pantomime using both hands (see Fig. 15 below).

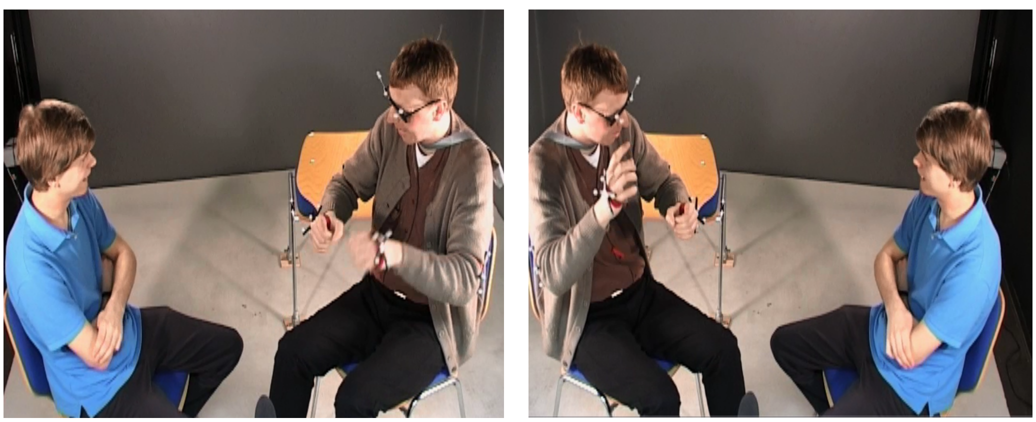


Figure 15. *The cart pantomime of the Route-Giver (left) and the umbrella pantomime of the Route-Giver with right hand using lax C (right). Observe that the left hand is still holding the pond*[38].

The stop makes perfect sense, since the Follower has to turn off to the right at the location of the ice-cream man. The Route-giver's gestural message is now that we are done with the route around the pond and the pond-gesture can be dropped for the time being. It is again taken up in a subsequent extensive repeat not discussed in this paper.

Résumé and Look-back: We started with mapping information chunks of the datum onto the formal agents using, e.g., SaGA pauses and intonation contours. We then began the discussion of the multi-modal anaphora in the datum and explained how the anaphora resolution is working in λΨ. Multi-modal antecedents consisting of speech and gesture are generated by a fusion device combining speech semantics and gesture semantics in one representation as suggested in the multi-

[38] Observe that the umbrella stick is gesturally larger than the pond and also separated from it. So, if we take that at face value, we have a gestural contradiction: truth functionality of gesture gets lost.

modal fusion literature in Sec. 3.6. The fused information is sent on Ψ's output channel to the outside and can be received by some anaphoric representation for pronouns, adverbs or definites showing a corresponding input channel. The linking of the antecedent information to the anaphoric site is achieved through Ψ's channel concept and the interface of Ψ-names and λ-variables. The outgoing multimodal information can be distributed to several receiving ports broadcasting across stretches of non-alignable speech, thus showing the independence of gesture and speech. Due to the communication concept of λΨ, the anaphora intuitions laid out are different from what we reported in the literature overview (Sec.s 3.1, 3.2, 3.3). They are an inverse to those proposals, which, starting from syntactic anaphora information, look bottom up for exactly one antecedent. Essentially, the λΨ-calculus' output-input device captures the "Familienähnlichkeit" (family resemblance) of multi-modal anaphora of different sorts using top-down communication.

At the end we can look back at the Sec.s 3.8 "Methodology and Contributions of Paper" and 4.3 "Data and Algorithmic Handling Matched" and review what has been cashed in in which ways. As far as we know this is the first paper dealing with multimodal anaphora and the distribution of information in dialogue (broadcasting). These rest on the λΨ-techniques. For handling independent gesture information co-operating with speech information as in the arch' and small arch case we need a communication vehicle, served by λΨ's output-input agents. For speech-gesture concurrency as with vehicle or right angle one needs a concurrency concept like the |-operator. Top-down anaphora resolution also relies on the communication mechanism of λΨ, more specifically, the distribution of information as in the case of pond'_{LH}, on the !-operator. No existing static grammar-bound speech-gesture account could handle the problems listed in Sec. 3.8. Therefore, at present, there is no alternative to λΨ- techniques or similar ones which might exist given the broad spectrum of process algebras.

6. Conclusions and Further Research

In this paper we dealt with verbal anaphora, multimodal anaphora involving one-handed and two-handed gestures, and the broadcasting of multimodal information across speech contributions and turns. Anaphora resolution is based on the λΨ-calculus especially on the output-input information transfer by a channel mechanism and the same holds for broadcasting. In contrast to DRT-like approaches which treat anaphora by discourse referents (DRs) and accessibility relations linking the anaphor DR to the antecedent DR we transfer full terms. Here we project the whole antecedent information on an output channel down to its receiving channel changing the type of expression from (existential) quantification to singular term (the "Chastainian shift", if you like), if necessary. Note that outputs often get fairly complicated. In effect, if a (multi-modal) singular term has been coined, it can be sent further down. However, it may happen that no corresponding input channel exists. Then the information on the firing output channel suggests a possible discourse extension not exploited. Simple cases of speech-gesture integration demand that one starts with extracting information from the corpus annotation, map it onto a logical formula (a λ-term) and shift it around using Ψ-techniques, always guided by observing speech-gesture asymmetry, i.e. the fact that a gesture can come too early, too late or overlap only partly with the speech semantics it is to be aligned with. In such cases, speech and gesture can be asynchronous but bounded within a construction or at least within one turn. Note that broadcasting cases are different: There, gestural information is sent out without meeting a semantically fitting speech or gesture input channel in same contribution. The information held tries to communicate across constructions and turns but is kept from fusing; it is in a wait loop, so to speak. In the cases of the datum discussed broadcasting communication finally succeeds but it could also fail and peter out after some time. From a dialogue perspective, the following problems were treated in the main text: sequential and parallel dialogue contributions using λΨ's Agent concept, in-turn anaphora and multi-modal anaphora using Ψ's channel device and its input-output facilities, multi-modal anaphora in clarification requests, in-

turn and trans-turn communication by broadcasting gesture meaning, composition of meanings across (long) distances using λΨ. In the Appendix we provide a trace of the axioms used and show how coherence relations and anaphora in PTT can be modelled using λΨ-techniques.[39] Now, questions worthwhile to further investigate are: Which other information processes in dialogue could be described using the same Ψ-techniques and which cannot, in other words, what are the limits of the λΨ-techniques as developed in this paper? We first turn to the cases which concern the dialogue structure and which can possibly be handled: The route descriptions in SaGA are plan-based insofar as the landmarks are clear from the outset and the Route-Giver instructs the Follower how to get from one landmark to the next. Selection of objects of orientation is first of all due to the Route-Giver. In order to proceed on the route to the next landmark/object of orientation certain paths have to be taken. On the whole, the dialogue's "and next" structure suggests a reconstruction in terms of Hobb's Occasion relation, SDRT's Narration relation or PTTs context resource situations (see the literature discussion in Sec. three). So, as a first starting point, we can resort to a case structure (see **Definition 2** in Sec. four above): If the sculpture has been passed, next comes the round-about and the straight passage out of it etc.
However, there seem to be other structures which are decidedly bottom-up, so for example dialogue moves hooking up to the information already given by the Route-Giver such as acknowledgements, clarification requests, corrections, repeats of part or whole of the bus-tour description by Route-Giver or Follower. Repeats are often *ad-hoc*, following Route-Giver's or Follower's sporadic intentions, so the divide seems finally to be between systematic, hence expectable, dialogue moves and merely contingent ones. The systematic ones like acknowledgements could perhaps be rendered as answering to Route-Giver's expectations, hence in a top-down way, but we are not sure whether this will work. There is no easy way to deal with the divide and we leave that as our next big question for a follow-up paper.

Acknowledgements
I want to thank Andy Lücking who read a preliminary version of this paper. Thanks go also to the anonymous D&D reviewers and to the editor for their suggestions to improve the paper.

Appendices

1. Trace of λΨ-rendering of the Datum in Figure 2

1.1 The Arch Information

Agent(1): → $\underline{ch}_1$ ent-i λF∃x(hedge'(x) ∧ green'(x) ∧ F(x)) (ent-i) |
← !$\overline{ch}_{4a}$ < ιx(hedge'(x) ∧ green'(x) ∧ (partly-also'(with'(x, Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)))))), Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))>.0 |

Agent(2): → $\underline{ch}_2$ ent-i $\overline{ch}_1$ λYλy(y, partly-also'(with'(Y))) (ent-i) |
← $\overline{ch}_1$ λy (y, partly-also'(with'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))))).0 |

Agent(3): ←$\overline{!\ ch}_3$ λFZ. F(Z) ∧ of'(Z, s1) ∧ of'(Z, s2) ∧ side'(s1) ∧ side'(s2) ∧ connect'(a, s1, s2) ∧ arch'(a).0 = Def. $\overline{!\ ch}_3$ λx(arch'(x))(a).0 |

Agent(4): → $\underline{ch}_3$ arc-g $\overline{ch}_4$ λFZ(doorway'(Z) ∧ of'(Z, a) ∧ F(a)) (arc-g) | $\underline{ch}_3$ arc-g $\overline{ch}_2$ λFΔi (doorway'(i) ∧ of'(i,a) ∧ F(a)) (arc-g). 0 |
← $\overline{ch}_4$ λZ(doorway'(Z) ∧ of'(Z, a) ∧ λx(arch'(x))(a)).0 | $\overline{ch}_2$ (Δi (doorway'(i) ∧ of'(i,a) ∧ λx(arch'(x))(a))).0 |

---

[39] These issues are still quite low-level in character, however, theories of stepwise turn-production, -reception, in-turn communication, acknowledgement, clarification and mutual belief could build on them mainly due to the dynamics of λΨ..

Table 1. Agents for deriving the arch information tagged for input → and output ←.[40]

Agent(1) gets information on $\underline{ch}_1$ from Agent(2) sending on $\overline{ch}_1$. Agent(2) needs first information on input channel $\underline{ch}_2$ provided by Agent(4). Agent(4) is a parallel agent indicated by |, see Sec. 4, **Definition 2**. However, Agent(4) can only output via $\overline{ch}_2$ on the right hand side of | once it received a value on its $\underline{ch}_3$ input channel served by gestural Agent(3) modelling the information of gesture q (see Figures 3 and 4 in main text for the datum). The q information is implemented as the arch-information which can be iterated arbitrarily often due to replication "!" (see Sec. 4, **Definition 2**). Communication among Agent(2) and Agent(1) attributes to the member of a set the property of being a hedge, green and partly also with arched doorways. So, Agent(1) can send out either ιx(hedge'(x) ∧ green'(x) ∧ (partly-also'(with'(x, Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)))))) or Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)), see Sec. 5.1 (d). The doorways are taken up by the anaphora *these*. Agent(5) (see below) deals with the anaphora relation, receiving the value Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)) from Agent(1) via its output-channel.

A note on the multi-modal fusion perspective provided by λΨ

As can be seen from the transcript in Figure 1, 1.1-G and 1.1 RH, we have speech-gesture overlap of two uni-modal information processes, *Zum Teil auch mit Eingängen/Partly also with doorways* and the gesture {R-Handshape C, q}; C indicates width and q the shape built by thumb, index finger and the space these enclose, oriented downwards. The {R-Handshape C, q} gesture conveys independent information conceptualized in Agent(3) as replication $!\overline{ch}_3$ λx(arch'(x))(a).0, see Fn. 24 on the context-problem. Since *Partly also with* is also in the scope of the gesture stroke, we have speech-gesture asymmetry; this means that the semantics of the q-gesture cannot be directly combined with the semantics of *Partly also with* but has to be shifted until it finds the doorway-semantics. Observe that the definition of $!\overline{ch}_3$ λx(arch'(x))(a).0 mimics the hand- and finger-position of RH using side'(s1), side'(s2), and connect'. Agent(3) distributes its information via !. By channel equivalence, Agent(3) can communicate with Agent(4) which is a concurrent agent, itself composed of two concurrent agents. λx(arch'(x))(a) fuses with the verbal doorway information by λβ-conversion in both agents. The result is a multi-modal expression percolating up to communicate with Agent(2), which in turn contributes to Agent(1).

1.2 The *die/these* Anaphor, the Arch Gesture and the Park Entrance

Agent(5): → $\underline{ch}_5$ comp $\underline{ch}_{4a\ \text{second}}$ entr-i λF Z(like'(arch-shaped')(Z) ∧ F(Z)) (comp) (entr-i).0 |
← Δi (doorway'(i) ∧ like'(arch-shaped')(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(i, small-arch') ∧ like'(such'∃U *garden-entrance' (U))(i)) . (like'(arch-shaped')(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))) ∧ of'((Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))), small-arch') ∧ like'(such'∃U *garden-entrance' (U))(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)))).0 |

Agent(6): → $\underline{ch}_6$ comp $\overline{ch}_5$ λFV(of'(V, small-arch') ∧ F(V))(comp).0 |
← $\overline{ch}_5$ λV(of(V, small-arch') ∧ like'(such'∃U *garden-entrance' (U))(V))).0 |

Agent(7): ← $\overline{ch}_6$ λS like'(such'∃U *garden-entrance' (U))(S).0 |

Agent(8): → $\underline{ch}_8$ m-entr $\overline{ch}_9$ ιv(main-entrance'(v)). λp until-sometime-then'(p) (m-entr).0 |
← $\overline{ch}_9$ ιv(main-entrance'(v)). until-sometime-then'((ιv(main-entrance'(v)) = j ∈(Δi

[40] Since channels, except for 0-agents, are always marked for input or output by underscore as, e.g., $\underline{N}$ or overbar as, e.g., $\overline{N}$ (see **Definition 2**) tags are strictly speaking not needed. However, as discussions with anonymous D&D reviewers B and C on the communication behaviour of agents showed, the communication threads among agents are easier to follow with tags. Hence, they serve a useful didactic purpose.

(doorway'(i) ∧ arch-shaped'(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(i, small-arch') ∧ like'(such'∃U*garden-entrance' (U)))(i) ∧ comes'(j))).0 |

Agent(9): → $\underline{ch}_7$ p-ent $\overline{ch}_8$ ιv(main-entrance'(v)). λZ(ιv(main-entrance'(v)) = j ∈ Z ∧ comes'(j)) (p-ent).0

← $\overline{ch}_8$ (ιv(main-entrance'(v)) = j ∈(Δi (doorway'(i) ∧ arch-shaped'(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(i, small-arch') ∧ like'(such'∃U *garden-entrance' (U)))(i) ∧ comes'(j)). 0 |.

Agent(10): → $\underline{ch}_9$ m-ent λu∃y((large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) ∧ y = u) (m-ent) .0

← $\overline{ch}_{11}$ <∃y(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)). 0 |

Agent(11): .0 |

Agent(12): → $\underline{ch}_{11}$ garc $\underline{ch}_{14}$ veh $\underline{ch}_{15}$ quiri $\underline{ch}_{15b}$ ev λpath λvehi λarc λF (must-in(e1, FO, vehi path, arc) ∧ F(e1))(garc)(veh)(quiri)(ev).0 |

← ! $\overline{ch}_{16}$. (must-in(e1, FO, vehicle', right-angle', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ quickly(e1) ∧ right-into(e1, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)))). 0 |

Agent(13): ← ! $\overline{ch}_{15}$ right-angle'.0 |

Agent(14): ← ! $\overline{ch}_{14}$ vehicle'.0 |

Agent(15): → $\underline{ch}_{11}$ arc λarc-i $\overline{ch}_{15b}$ λe(quickly(e) ∧ right-into(e, arc-i)) (arc)

← $\overline{ch}_{15b}$ λe(quickly(e) ∧ right-into(e, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')))).0

Agent(16): → $\underline{ch}_{38}$ fronth(λfro(∃x(¬ wall'(x) ∧ hedge'(x) ∧ fro(x))))(fronth).0 |

Agent(16a): ← $\overline{ch}_{38}$ λz of'(z, front')) .0 |

Agent(17): → $\underline{ch}_{11}$ arc-i $\underline{ch}_{16\ second}$ ev $\underline{ch}_{14}$ ve $\underline{ch}_{15}$ pa λarc λe λvehi λpath ?into'(e, FO, vehi, path, in'(ιu (arc'(u)) = arc))(arc-i)(ev)(ve)(pa).0|

← $\overline{ch}_{36}$ (ιu (arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))).0 |

Agent(18): → $\underline{ch}_{16\ second}$ e $\underline{ch}_{14}$ veh $\underline{ch}_{15}$ sra $\underline{ch}_{11}$ arc λev λvehi λpath λarc(into'(ev, FO, vehi, path, in' (ιu (arc(u)) = arc)))(e)(veh)(sra)(arc). 0 |

← $\overline{ch}_{37}$ into'(e1, FO, vehicle', right-angle', in' (ιu(arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))). 0 |

Table 2: Agents needed to derive the *die/these* anaphor, the arch gesture and the park entrance gestures with tags for agents' input → and agents' output ←.

Agent(5) gets input on channel $\underline{ch}_5$ for the comparison clause *like such a garden entrance*. Agent(6) and (7) provide that. Agent(7) implements the comparison.[41] The information on $\overline{ch}_6$ is received by Agent(6), providing the input for Agent(5). The final derivation for Agent(5) is:

like'(arch-shaped')(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))) ∧ (of'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)), small-arch') ∧ like'(such'∃U *garden-entrance' (U))(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)))).0 |

This means that each i in the set operated on by Δ that is an arched doorway, i.e., is of'(i,a) ∧ arch'(a), is like'(arch-shaped') [first conjunct], has a small-arch' [second conjunct] and is like such a garden entrance [third conjunct]. Some of this information will be sent out by $\overline{ch}_7$ given that it finds a corresponding input channel:

[41] For the * used here, see Kamp and Reyle (2003), p. 327, fn 13.

$\overline{ch}_7$ (Δi (doorway'(i) ∧ like'(arch-shaped')(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(i, small-arch') ∧ like'(such'∃U *garden-entrance' (U))(i)) . (like'(arch-shaped')(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))) ∧ of'((Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))), small-arch') ∧ like'(such'∃U *garden-entrance' (U))(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)))).0 |.

This solves the anaphoric relation among *doorways* and *these*; observe that *these*, the anaphora, is represented as Δi(doorway'(i) ∧ of'(i,a) ∧ arch'(a)), i.e. as the set of arched doorways.
The value in the prefix says that each arched doorway i is arch-shaped and has a small arch and is like such a garden entrance. Technically, we assemble properties originally attributed to the Δ-set into the Δ-set itself following the Chastainian observation of referential term extension in Sec. 5.1 (b). After action on $\underline{ch}_7$ Agent(9) looks as follows:

$\overline{ch}_8$ (ιv(main-entrance'(v)) = j ∈(Δi (doorway'(i) ∧ arch-shaped'(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(i, small-arch') ∧ like'(such'∃U *garden-entrance' (U)))(i) ∧ comes'(j))**.** 0 |.

On the convention that an existentially quantified expression can send a singular term, we get for Agent(10):

= $\overline{ch}_{11}$ <ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))>**.** ∃y(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)). 0 |

Remember that an identity can proceed with either term according to 5.1 (b) c. The output derivation for Agent(12) is then:

!$\overline{ch}_{16}$ <ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)), e1>. (must-in(e1, FO, vehicle', right-angle', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ quickly(e1) ∧ right-into(e1, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)))). 0 |
Agent(13) and Agent(14) specify the right-angle and the vehicle information, both given by gesture, respectively. Agent(15a) specifies the moving into the arch. Agent(17) embodies the Follower's clarification request.

1.3 Post-holds and Broadcasting

Agent(12): → $\underline{ch}_{11}$ garc $\underline{ch}_{14}$ veh $\underline{ch}_{15}$ quiri $\underline{ch}_{15b}$ ev λpath λvehi λarc λF (must-in(e1, FO, vehi, path, arc) ∧ F(e1))(garc)(veh)(quiri)(ev).0 |
← !$\overline{ch}_{16}$. (must-in(e1, FO, vehicle', right-angle', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ quickly(e1) ∧ right-into(e1, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)))). 0 |
Agent(13): ← !$\overline{ch}_{15}$ right-angle'.0 |
Agent(14): ← !$\overline{ch}_{14}$ vehicle'. 0 |
Agent(15): → $\underline{ch}_{11}$ arc λarc-i $\overline{ch}_{15b}$ λe(quickly(e) ∧ right-into(e, arc-i)) (a rc)
← $\overline{ch}_{15b}$ λe(quickly(e) ∧ right-into(e, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')))).0
Agent(16): → $\overline{ch}_{38}$ fronth(λfro(∃x(¬ wall'(x) ∧ hedge'(x) ∧ fro(x))))(fronth).0 |
Agent(16a): ← $\overline{ch}_{38}$ λz(of'(z, front')) .0 |
Agent(17): → $\underline{ch}_{11}$ arc-i $\underline{ch}_{16\ second}$ ev $\underline{ch}_{14}$ ve $\underline{ch}_{15}$ pa λarc λe λvehi λpath ?into'(e, FO, vehi, path, in'(ιu (arc'(u)) = arc))(arc-i)(ev)(ve)(pa).0|
← $\overline{ch}_{36}$ (ιu (arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb'))).0 |

Agent(18): → $\underline{\text{ch}}_{16\ \text{second}}$ e $\underline{\text{ch}}_{14}$ veh $\underline{\text{ch}}_{15}$ sra $\underline{\text{ch}}_{11}$ arc λev λvehi λpath λarc(into'(ev, FO, vehi, path, in'(ιu (arc(u)) = arc)))(e)(veh)(sra)(arc). 0 |
← $\overline{\text{ch}}_{37}$ into'(e1, FO, vehicle', right-angle', in' (ιu(arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ (y) ∧ of'(y, curb'))). 0 |

Agent(12): → $\underline{\text{ch}}_{11}$ garc $\underline{\text{ch}}_{14}$ veh $\underline{\text{ch}}_{15}$ quiri $\underline{\text{ch}}_{15b}$ ev λpath λvehi λarc λF (must-in(e1, FO, vehi, path, arc) ∧ F(e1))(garc)(veh)(quiri)(ev).0 |
← !$\overline{\text{ch}}_{16}$. (must-in'(e1, FO, vehicle', right-angle', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ quickly'(e1) ∧ right-into'(e1, ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v)))).0 |

Agent(13): ← ! $\overline{\text{ch}}_{15}$ right-angle' .0 |

Agent(14), (15), (16): see Table 2.

Agent(17): → $\underline{\text{ch}}_{11}$ arc-i $\underline{\text{ch}}_{16\ \text{second}}$ ev $\underline{\text{ch}}_{14}$ ve $\underline{\text{ch}}_{15}$ pa λarc λe λvehi λpath ?into'(e, FO, vehi, path, in'(ιu (arc'(u)) = arc))(arc-i)(ev)(ve)(pa).0 |
← $\overline{\text{ch}}_{36}$? into'(e1, FO, vehicle', right-angle', in'(ιu (arc(u)) = ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')))).0 |

Agent(20) = case Agent(21): Agent(22).0 |

Agent(21): → $\underline{\text{ch}}_{15a}$ pa-i $\underline{\text{ch}}_{14}$ veh-i $\underline{\text{ch}}_{11}$ ther-i λpth1 λveh λthere (have-driven-in'(e1, FO, veh, pth1,there)) (pa-i)(veh-i) (there-i) .0 |
← $\overline{\text{ch}}_{16}$ path1'. (have-driven-in'(e1, FO, vehicle', path1', ιx (large'(x) ∧ grey'(x) ∧ arch'(x) ∧ of'(x, curb')))).0 |. $\overline{\text{ch}}_{17}$

Agent(22): → $\underline{\text{ch}}_{15c}$ pa-i $\underline{\text{ch}}_{14}$ veh-i $\underline{\text{ch}}_{11}$ the-i $\underline{\text{ch}}_{18}$ po$_{LH}$ $\underline{\text{ch}}_{19}$ po$_{RH}$ λpth2 λveh λthere <ιx(pond'(x)), ιx(pond$_{LH}$'(x))>. (drive'(e2, FO, veh, pth2, there) ∧ straight-forward'(e2, veh, pth2) ∧ towards'(e2, FO, veh, λrs ∃x (pond'(x) ∧ x = r ∧ x = s)))(pa-i)(veh-i)(the-i)(po$_{LH}$)(po$_{RH}$).0|
← ! $\overline{\text{ch}}_{17}$<ιx(pond'(x)), ιx(pond$_{LH}$'(x))>. (drive'(e2, FO, vehicle', path2', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ straight-forward'(e2, vehicle', path2') ∧ towards'(e2, FO, vehicle', ∃x(pond'(x) = pond'$_{LH}$ = pond'$_{RH}$))).0|

Agent(23) = case Agent(24): Agent(25).0 |

Agent(24): ← $\overline{\text{ch}}_{20}$ pond'$_{LH}$. at'(ιx (pond'(x)) = pond'$_{LH}$).0 |

Agent(25) = case Agent(26): Agent (27).0 |

Agent(26): → $\underline{\text{ch}}_{17\ \text{first}}$ po-i $\underline{\text{ch}}_{14}$ veh-i $\underline{\text{ch}}_{21}$ pa-i λpo λvh λpth3 λpo$_{LH}$ (drive'(e3, FO, vh, pth3) ∧ towards'(e3, FO, vh, pth3, po) ∧ po = pond'$_{LH}$)(po-i)(veh-i)(pa-i).0 |
← (drive'(e3, FO, vehicle', path'$_{RH}$) ∧ towards'(e3, FO, vehicle', path'$_{RH}$, pond'$_{LH}$) ∧ ιx(pond'(x)) = pond'$_{LH}$).0 |.

Agent(27): → $\underline{\text{ch}}_{14}$ veh-i $\underline{\text{ch}}_{21}$ pa-i $\underline{\text{ch}}_{22}$ pa-ic $\underline{\text{ch}}_{20}$ po-i λvh λpth3 λpth4 λpo$_{LH}$ (drive'(e3, FO, vh, pth3) ∧ right-around'(e3, FO, vh, pth4, po$_{LH}$))(veh-i)(pa-i)(pa-ic)(po-i) .0 |
← $\overline{\text{ch}}_{22}$ path-c'$_{RH}$.(drive'(e3, FO, vehicle', path'$_{RH}$) ∧ right-around'(e3, FO, vehicle', path-c'$_{RH}$, pond'$_{LH}$)).0 |

Agent(31): ← $\overline{\text{ch}}_{25}$ (runs'(e4, (ιx(hedge'(x) ∧ green'(x) ∧ (partly-also'(with'(x,Y) ∧ doorway'(Y) ∧ of'(Y, a) ∧ arch'(a)))), Z) ∧ Z = another'(roughly'|50|meter'))). 0 |

Agent(33): .0 |

Agent(34): ← $\overline{\text{ch}}_{26}$ environs'.0 |

Agent(37): → $\underline{\text{ch}}_{26}$ area $\underline{\text{ch}}_{30}$ ben1 $\underline{\text{ch}}_{31}$ ben2 $\underline{\text{ch}}_{28}$ poss-ben λloc λloc-ben1 λloc-ben2 λben(at'(loc, of(ben, here', loc-ben1) ∧ of'(ben, there', loc-ben2))) (area)(ben1)(ben2)(poss-ben) .0| (at'(environs', of'(perhaps'(Δi (bench' (i))), here', ι loc1(of (Δi (bench'(i)),

loc1))) ∧ of(perhaps'(Δi (bench' (i))), there', ι loc2 (of (Δi (bench'(i)), loc2))))) .0 |
Agent(41): Accommodated.
Agent(42): → ch14 veh-i ch22 pth-i ch18 po$_{LH-i}$ λvh λpth λpo$_{LH}$ (drive'(e6, FO, vh, pth) ∧ round'(ιx (pond'(x)))(e6) ∧ ιx (pond'(x)) = po$_{LH}$ ∧ right'((around')(po$_{LH}$))(e6))(veh-i)(pth-i)(po$_{LH-i}$) .0 |
Agent(43): ← !ch34 ιx(ice-cream-man(x)). ∃x(ice-cream-man(x)).0 |
Agent(44): ← ch35 ιz(loc(z)).0 |
Agent(45): ← ch36 ∃y sometimes'(at(ιx(ice-cream-man'(x)), ιz(loc(z))) ∧ boundary'(y) ∧ of'(y, pond'$_{LH}$) ∧ part-of(ιz(loc(z)), y)).0
Agent(46): → ch14 veh-i ch35 loc-icm λveh λloc (drive-off(e7, FO, veh-i, at(loc)) ∧ to-right(e7))(veh-i)(loc-icm).0 | drive-off(e7, FO, vehicle', at(ιz(loc(z))) ∧ to-right(e7)).0 |
Agent(47): ← ch37 r. ask(FO, p) ∧ p = ∃q imply(sometimes'(q), r).0 |
Agent(48): → ch37 chang ch34 icr-m ch35 loc-icm λchang λice-crm λloc could-be(chang = ∃y(changed'(e8, y, q) ∧ q = be(ice-crm, at(ice-crm-loc))) (chang)(icr-m)(loc-icm) .0| could-be(r = ∃y(changed'(e8, y, q) ∧ q = be'(ιx(ice-cream-man'(x)), at(ιz(loc'(z)))))
Agent(49a) = case Agent(49): Agent(50) |
Agent(51): .0 |
Agent(49): → ch40 tour-icm ch41c-umb λp q∀e(case e = e9: p ∧ q)(tour-icm)(c-umb) .0 |
← ∀e(case e = e9: (was' (e9, ιx(ice-cream-man'(x)), on-Route-Giver's-tour'at(ιz(loc'(z))))) ∧ ∃x was'(cart'(x) ∧ of'(x, ιx(ice-cream-man'(x))) ∧ of'(x, y) ∧ handle-bar'(y) ∧ horizontally-grasped'(y) ∧ of'(x, u) ∧ umbrella'(u) ∧ of'(u, v) ∧ vertical-handle'(v))). 0 |
Agent(50): → ch34 icr-m ch35 loc-icm λicr-m λloc-icm ch40 on-Route-Giver's-tour'(was' (e9, icr-m, at(loc-icm)))(icr-m)(loc-icm) .0 |
← ch40 on-Route-Giver's-tour'(was'(e9, ιx(ice-cream-man'(x)), at(ιz(loc'(z))))) .0 |
Agent(51): → ch34 icr-m ch38 cart-h ch39 umb-h λicr-m λcart-h λumb-h ch41 (∃x was' (cart'(x) ∧ of'(x,icr-m) ∧ of'(x, y) ∧ cart-h'(y) ∧ of'(x, z) ∧ umbrella'(z) ∧ of'(z, v) ∧ umb-h'(v)))(icr-m)(cart-h)(umb-h) .0 |
← ch41 (∃x was' (cart'(x) ∧ of'(x, ιx(ice-cream-man'(x))) ∧ of'(x, y) ∧ handle-bar'(y) ∧ horizontally-grasped'(y) ∧ of'(x, u) ∧ umbrella'(u) ∧ of'(u, v) ∧ vertical-handle'(v))) .0 |
Agent(52): ← ch38 λu (handle-bar'(u) ∧ horizontally-grasped'(u)).0 |
Agent(53): ← ch39 λz (vertical-handle'(z)).0 |

Table 3: Agents needed to model post-holds and broadcasting tagged for input → and output ←.

For the Route-Giver's 3.2-G/E we use the case construct of Ψ: *if you have driven in there* works as antecedent φ and *you drive straight towards a pond. A pond* as consequent (see Sec. 4, **Definition 3** for the general format). The anaphora *there* is resolved to *a large grey arch* becoming definite. Agent(21) amounts in the end to

ch16path1'.(have-driven-in'(e1, FO, vehicle', path1', ιx (large'(x) ∧ grey'(x) ∧ arch'(x) ∧ of'(x, curb'))))).0 |.

ch15a contributes path1', ch14 vehicle', ch11 the large grey arch with curb'. The initial section of the path, path1', is returned on ch16.The consequent *you drive straight towards a pond. A pond. is* modelled as Agent(22). Assuming $pond_{LH}$ as value on $\underline{ch}_{18}$ and path2' as value on $\underline{ch}_{15c}$ [42] Agent(22) turns into

$! \overline{ch}_{17}$ <ιx(pond'(x)), ιx($pond_{LH}$'(x))>. (drive'(e2, FO, vehicle', path2', ιy(large'(y) ∧ grey'(y) ∧ arch'(y) ∧ of'(y, curb')) = ιv(main-entrance'(v))) ∧ straight-forward'(e2, vehicle', path2') ∧ towards'(e2, FO, vehicle', ∃x(pond'(x) = $pond'_{LH}$ = $pond'_{RH}$).0|.

We need the path-into, pa-i, its value coming from $\underline{ch}_{15c,}$ the vehicle, the main entrance, where the bus went in by $\underline{ch}_{11}$ and a left hand and a right-hand O-shape gesture, both denoting the pond. Given these, we end up with the derivation of the broadcasting Agent(22). The route around the pond (see VR Figure 12a) is represented as follows: We have an embedded conditional structure: Agent(23) = case Agent(24): Agent(25) and Agent(25) = case Agent(26): Agent(27) with the noticeable occurrence of the linguistic pond information syntactically in the prefield (see Sec. 5.1 on this notion). The whole conditional structure is, Agent(24) and Agent(26) covering respectively the role of φ_i in the general definition (see Sec. four, **Definition 2**): Agent(24) represents the antecedent *And at this pond*, Agent(26) the antecedent *You drive towards it*. In specifying Agent(26) we use the linguistic pond information on $\underline{ch}_{17first}$ (distributively communicated by Agent(22)), the vehicle-gesture information we already have and an assumed path drawn by the RH gesture on $\underline{ch}_{21}$, $path'_{RH}$, again accommodated, shown in Fig. 12a. The LH pond gesture $pond'_{LH}$ is held all the time as indicated by iterative "!". The central movement information is conveyed by the RH gesture $path'_{RH}$. We get as final derivation for Agent(26), observe the gestural representation of vehicle', $path'_{RH}$, and $pond'_{LH}$:
(drive'(e3, FO, vehicle', $path'_{RH}$) ∧ towards'(e3, FO, vehicle', $path'_{RH}$, $pond'_{LH}$) ∧ ιx(pond'(x)) = $pond'_{LH}$).0 |.
Agent(27) represents *and you drive right around it*. Its final derivation is
$\overline{ch}_{22}$ $path\text{-}c'_{RH}$.(drive'(e3, FO, vehicle', $path'_{RH}$) ∧ right-around'(e3, FO, vehicle', $path\text{-}c'_{RH}$, $pond'_{LH}$)).0 |,
where, looking at the λ-variables, the parameters for vehicle, vh, the path towards the pond, pth3, the RH-path around the pond, pth4, and the LH-gestured po_{LH} are filled by the values provided by the input channels. Note that substantial information is expressed by the LH and RH gestures. The path-information $path\text{-}c'_{RH}$ is sent to the outside and will be taken up in Agent(42).
Next, the Route-Giver refers to the hedge (5-G/E), *The hedge*, syntactically again situated in the prefield and resumed in the middle field. The long-distance anaphora for *The hedge* is resolved with the broadcasting hedge-information from above, i.e. by Agent(1) via channel $!\overline{ch}_{4afirst}$. This is embodied in Agent(31) which is derived as

$\overline{ch}_{25}$(runs'(e4, (ιx(hedge'(x) ∧ green'(x) ∧ (partly-also'(with'(x,Y) ∧ doorway'(Y) ∧ of'(Y, a) ∧ arch'(a)))), Z) ∧ Z = another'(roughly'|50|meter'))). 0 |.

While the Route-Giver produces the description *The hedge, {it runs another roughly} 50 m.*, we might expect that the hedge is at issue, nevertheless it is the L-Handshape O with denotation pond' which is broadcasted independently from speech. The Route-Giver's next remark is concerned with the environs (6G/E): *Then there are also here and there benches.* Again, the L-Handshape O is continued. The indexing D-shape shows the temporal asymmetry of gesture and speech characteristic for iconic gesture use. The first indexing is too early, ideally it should be aligned with *hin [und wieder]/here [and there].*

[42] Again, observe that $pond_{LH}$ as value on $\underline{ch}_{19}$ and path2' on $\underline{ch}_{15c}$ are both accommodated. A more elaborate reconstruction would have to express that path2' continues path1' in the direction towards the pond. We forego the more detailed reconstruction here.

Assume an Agent (34) which provides a value environs' for an indexical gesture and sends it out on $\overline{ch}_{26}$. Then Agent(37) does the integration work. It gets input from Agent(34) which provides the value environs' on channel $\underline{ch}_{26}$. Assume Agent(41) contributes the "doubtful benches", poss-ben, on $\underline{ch}_{28}$. Then the final derivation for Agent(37) is

(at'(environs', of'(perhaps'(Δi (bench' (i))), here', ι loc1(of (Δi (bench'(i)), loc1)) ∧ of(perhaps'(Δi (bench' (i))), there', ι loc2 (of (Δi (bench'(i)), loc2))))) .0 |.

There is no outgoing transport channel for this information stressing the singularity of this contribution as in the case of the hedge information.
7-G/E essentially repeats *But you drive around the pond. Right around*, modelled in Agent(42). So, almost the same Agents for input are active as in Agent(22): $\underline{ch}_{14}$ veh-i contributes the vehicle information, $\underline{ch}_{22}$ pth-i the gestured path around the pond, path-c'$_{RH}$, and $\underline{ch}_{18}$ po$_{LH}$ the pond assigned by the left hand. The pond is a non-official landmark, important for turning to the right and continuing the tour to the chapel and on to the fountain. The turn-off to the right is at an ice-cream cart, referred to in 8-G/E: *And sometimes there is an ice-cream man there. And at him you drive off to the right.* We need new entities for the ice-cream man and his location at the embankment of the pond. The "pond-gesture", small O, is held as before (see Fig. 12a). Therefore, we have in Agent(45):

$\overline{ch}_{36}$ ∃y sometimes'(at(ιx(ice-cream-man'(x)), ιz(loc(z))) ∧ boundary'(y) ∧ of'(y, pond'$_{LH}$) ∧ part-of(ιz(loc(z)), y)).0 |,

where Agent(43) and Agent(44) have contributed, respectively, the ice-cream man and his location. We hand on the location of the ice-cream man ιz(loc(z)) which we need to resolve the anaphora *at him*. We have the pond gesture as before and a G-bound gesture indexing the ice-cream man's location. The datum shows that the R-Handshape G being on {nen Eisverkäufer}/{'n ice-cream man} comes too early.
Ideally, it should align with *da/there*. We represented "manchmal/sometimes" as propositional operator in Agent(45) which is sufficient for our purposes[43].
Using Agent(14) and Agent(44), Agent(46) expands to

drive-off(e7, FO, vehicle', at(ιz(loc(z))) ∧ to-right(e7)).

At this stage of the dialogue we get an interesting clarification request of the FOLLOWER's (9-G/E): "Was heißt 'manchmal'?/"What does 'sometimes' mean"?. Due to the polysemy of German "heißen"/"mean" it can be read as "What does 'sometimes' imply/come to"? So we render it as Agent(47)[44].

2. Further Applications

2.1 Rhetorical Relations

In the literature discussion Sec. 3.1 we pointed out the importance of semantic/pragmatic relations among sentence pairs for anaphora resolution as observed in the Hobbs-Hume-Kehler-Rohde research line. Why does one, in addition, need rhetorical relations? Essentially, they cover the intuition that some parts of a discourse are more tightly bound together than others as exemplified

[43] We do not want to go into quantifying across time intervals in this paper which an existential quantification for *sometimes'* would need.

[44] Andy Lücking attributes the Follower's clarification request to Ginzburg's "intended content clarification", cf. Ginzburg (2012), p. 153.

in SDRT (AL, p. 147). The primary task of rhetorical relations and the right border constraint is to enable anaphora resolution which, as will be clear by now, we do differently, by direct message passing. But we do admit that there is a marked substructure in the route-giver's directives in the datum: first we have the introduction of the hedge with its doorways and the shape of the doorways, then we get the main-entrance with its properties and so on. We assume that *dann* + *sind*/then + are acts as a cue phrase for continuation and *dann* + finite verb as a cue phrase for narration. We let rhetorical relations supervene on anaphora resolution below. In order to reduce discourse information we use the following simplified version of the given datum:

1-E	There are doorways.
1.1-RH	{R-Handshape q }
1.2-E	These are then arch-shaped.
1.2-RH	{R-Handshape small-arch-shape}
2-E	[Until sometime] then the main entrance comes.
3-E	That is a large grey arch.
3.1-LH	{R-Handsh. large arc-shape + small c}

Table2: Simplified version of datum.

We have by increments A1 < A2 < ,….,< A7 and A1 | A2 | , …., |A7. Agent(1) captures the doorways plus an accompanying R-Handshape q gesture. Agent(2) introduces the Continuation relation between this information and the doorways being arch-shaped. The co-verbal gesture signs small arches for the doorways. Agent3 shows that *these* is translated into an input channel plus Ψ-name dw1 receiving the doorway-set Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)) from Agent(1). Agent(3) hands on the information that each i in the Δ-set is arch-shaped and has a small arch. Agent(4) encodes the Narration relation mapping the generated Continuation relation onto the main entrance information, the main entrance being one of the doorways with a small arch. Agent(5) outputs the main entrance information and the fact that the main entrance comes in sight on the tour. Agent(6) continues the information we have so far with a description of the main entrance being a large grey arch. The gesture depicts a large arc plus the curb-information using small C. By message passing, the main-entrance is identified with the entity being a large grey arch. Rhetorical relations are treated like anaphora, agents sending multi-modal propositional information interact with ports receiving this information and accumulating it. This makes rhetorical relations multi-modal anaphoric entities. In sum, our reconstruction of the simplified datum captures the following issues: incrementality, multi-modal fusion of gesture and speech, anaphora resolution, and rhetorical relations Continuation and Narration[45].

Agent(1) $\overline{ch}_1$ (Δi (doorway'(i) ∧ of'(i,a) ∧ λx(arch'(x))(a))).0 |[46]

Agent(2) $\underline{ch}_3$ dw $\underline{ch}_1$ dw2 $\overline{ch}_4$ $\lambda p_1 p_2$ Continuation(p_1, p_2)(dw)(dw2)
= $\overline{ch}_4$ Continuation((arch-shaped'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)), small-arch'))), Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a))).0 |

Agent(3) $\underline{ch}_1$ dw1 $\overline{ch}_3$ λ Z (arch-shaped'(Z) ∧ of'(Z, small-arch'))(dw1)
= $\overline{ch}_3$ (arch-shaped'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)) ∧ of'(Δi (doorway'(i) ∧ of'(i,a) ∧ arch'(a)), small-arch'))).0 |

[45] We do not use rhetorical relations to tie speech semantics and gesture semantics together as proposed in Lascarides and Stone (2009). See 3.6 in the literature section on their proposal in the context of multi-modal fusion.

[46] Agent1 is simplified inasmuch as the whole process of generating the arch-information and attributing it to the doorways is not shown. Since the focus here is on rhetorical relations this will not matter.

Agent(4) $\underline{ch}_4$ prop1 $\underline{ch}_{5\ second}$ prop2 $\overline{ch}_6$ $\lambda p_1 p_2$ Narration(p_1, p_2)(prop1)(prop2)
= $\overline{ch}_6$ <Narration(Continuation((arch-shaped’(Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)) $\wedge$ of’(Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)), small-arch’))) , Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a))), (ιv(main-entrance’(v)) = j $\in$(Δi (doorway’(i) $\wedge$ arch-shaped’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a) $\wedge$ of’(i, small-arch’))) $\wedge$ comes’(j))>.0 |

Agent(5) ! $\overline{ch}_5$ < ιv(main-entrance’(v)) = j $\in$(Δi (doorway’(i) $\wedge$ arch-shaped’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)) $\wedge$ of’(i, small-arch’)), (ιv(main-entrance’(v)) = j $\in$(Δi (doorway’(i) $\wedge$ arch-shaped’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)) $\wedge$ of’(i, small-arch’)) $\wedge$ comes’(j)).0> |

Agent(6) $\underline{ch}_{6\ first}$ prop2 $\underline{ch}_7$ prop1 $\lambda p_1 p_2$ Continuation(p_1, p_2)(prop2)(prop1)
= Continuation(Narration(Continuation((arch-shaped’(Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)) $\wedge$ of’(Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a)), small-arch’))) , Δi (doorway’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a))) , $\exists$y((large’(y) $\wedge$ grey’(y) $\wedge$ arch’(y) $\wedge$ of’(y, curb’)) $\wedge$ y = ιv(main-entrance’(v)) = j $\in$(Δi (doorway’(i) $\wedge$ arch-shaped’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a) $\wedge$ of’(i, small-arch’)))))

Agent(7) $\underline{ch}_{5\ first}$ me $\overline{ch}_7$ λu$\exists$y((large’(y) $\wedge$ grey’(y) $\wedge$ arch’(y) $\wedge$ of’(y, curb’)) $\wedge$ y = u)(me)
= $\overline{ch}_7$ $\exists$y((large’(y) $\wedge$ grey’(y) $\wedge$ arch’(y) $\wedge$ of’(y, curb’)) $\wedge$ y = ιv(main-entrance’(v)) = j $\in$(Δi (doorway’(i) $\wedge$ arch-shaped’(i) $\wedge$ of’(i,a) $\wedge$ arch’(a) $\wedge$ of’(i, small-arch’)))).

1.2 Anaphora in PTT

Poesio and Rieser (2011) focussed on the resolution of definite anaphora, NPs and pronouns from the perspective of incrementality in PTT. The main insights were[47]: Generalizing Löbner’s account of definites (Loebner 1987) to instantiate discourse referents. Use the resource situations of Situation Semantics to capture pointings, anaphoric reference and the function of rhetorical relations. Capture experimental results of the visual world paradigm. Use the LTAG formalism for encoding incrementality using prioritized defaults for all levels of the parsing tool developed. Developing the concept of micro-conversational events (MCEs).
For illustration purposes we give a full PTT-version of 1-E below[48]:
[K1.1, up1.1, ce 1.1|
K1.1 is [X| X is Δi (doorway’(i))],
up1.1: utter(Route-Giver, FOLLOWER, “There are doorways”),
sem(up1.1) is K1.1,
ce1.1: assert(Route-Giver, FOLLOWER, K1.1),
generate(up1.1, ce1.1).
We want to show how $\lambda\Psi$-ideas can help to account for two types of dynamics, the speech-gesture interface and the resolution of multi-modal plural anaphora. As in PTT in general, we also rely on Muskens’ Compositional DRT (CDRT) as a background formalism. The division of labour between CDRT and the $\lambda\Psi$-calculus is as follows: box-internal regularities follow the CDRT-regime, communication among box-agents is organized in the $\lambda\Psi$-manner. We restrict our focus to semantic representations; *up1.1, ce1.1*, and *generate* are not dealt with (see however the remark about further simulations below). Agents are now taken as a dynamic version of the MCEs in Poesio (1995). Information in red is again marked as contributed by gesture.

[47] We do not give detailed explanations and literature references here. See the Poesio and Rieser (2011) paper on that.

[48] K is a discourse referent in CDRSs, *up1.1* stands for “utterance phrase”, *sem* for “semantics”, *ce1.1* for conversational event, i.e. a discourse act, utterance phrases generate conversational events.

Agent(1) $\overline{ch}_1$ X. [X|X is Δi (doorway'(i))];
Agent(2) $\underline{ch}_1$ dw λY $\overline{ch}_2$ Z. [| of(Y,a), arch(a), Z is Δi (doorway'(i), of'(i,a), λx(arch'(x))(a))] (dw);
Agent(3) $\underline{ch}_2$ dwa λU $\overline{ch}_3$ V. [| arch-shaped(U), U is V](dwa);
Agent(4) $\underline{ch}_3$ dwas λW $\overline{ch}_4$ Y. [| of(W,a), small-arch(a),Y is Δi (doorway'(i), of'(i,a), λx(arch'(x))(a), small-arch(a))].

The idea is as follows: Agent(1) sends the discourse referent for the plural term Δi (doorway'(i)) out which Agent(2) receives[49]. Agent(2) sends out the multi-modal information tied to doorway. Agent(3) attributes arch-shaped to the multi-modal term and sends the input out. Agent(4) adds multi-modal information and sends the multi-modal Δi (doorway'(i), of'(i,a), λx(arch'(x))(a), small-arch(a)) out. The set variables X, Z, V, Y do essentially the duty of the resource situation variable in the Poesio and Rieser (2011) paper. In a similar manner *up*, *ce*, and *generate* can be treated.

References

Samson Abramsky (2008). Information, Processes and Games. In P. Adriaans and J. van Benthem, editors. *Philosophy of Information*. Vol. 8, pages 483 – 549 Elsevier, Amsterdam

Katya Alahverdzhieva & Alex Lascarides (2010). Analysing Language and Coverbal Gesture in Constraint-based Grammars. In *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*, pages 5-25. http://makino.linguist.univ-paris-diderot.fr/files/hpsg2010/file/abstracts/HPSG/alahverdzhieva-lascarides-hpsg2010.pdf

Katya Alahverdzhieva & Alex Lascarides (2011). An HPSG Approach to Synchronous

Speech and Deixis. In *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, pages 6–24, University of Washington, Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2011/alahverdzhieva-lascarides.pdf

Nicholas Asher (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, Dordrecht. ISBN 0-7923-2242-8

Nicholas Asher and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press, Cambridge. ISBN 0521650585

J. C. M. Baeten (2004). A brief history of process algebra (PDF). Rapport CSR 04-02. VakgroupInformatica, Technische Universiteit Eindhoven

Jon Barwise and Jerry Seligman (1997). *Information Flow*. The Logic of Distributed Systems. Cambridge University Press, Cambridge ISBN 0 521 58386 1

Srinivas Bangalore and Michael Johnston (2009). Robust Understanding in Multimodal Interfaces. *Computational Linguistics*, 35(3): 345-397. http://dx.doi.org/10.1162/coli.08-022-R2-06-26

Jesper Bengtson, Magnus Johansson, Joachim Parrow, and Victor Björn (2011). PSI-CALCULI: A FRAMEWORK FOR MOBILE PROCESSES WITH NOMINAL DATA AND LOGIC. *Logical Methods in Computer Science* 7 (1,11): 1-44. arXiv:1101.3262 [cs.LO]

[49] According to our identity condition 5.1 (b).c we could also send the term to the right of the "is".

Kirsten Bergmann, Hannes Rieser and Stefan Kopp (2011). Regulating Dialogue with Gesture. Towards an Empirically Grounded Simulation with Conversational Agents. In *Proceedings of SIGdial 2011*, pages 88-97

J. A. Bergstra, A. Ponse, S.A. Smolka, editors (2001). *Handbook of process algebra*. Online resource https://eu04.alma.exlibrisgroup.com/view/uresolver/49HBZ_BIE/openurl?u.ignore_date_coverage=true&portfolio_pid=53309823570006442&Force_direct=true

Johannes Borgström, Shuqin Huang, Magnus Johansson, Palle Raabjerg, Björn Victor, Johannes ˚Aman Pohjola, and Joachim Parrow. Broadcast psi-calculi with an application to wireless protocols. *Software and System Modeling*, 14(1):201–216, 2015.

Susan, E. Brennan, Marilyn W. Friedman, Carl Pollard (1987). A centering approach to pronouns. In *Proceedings of the 25th ACL*, pages 155-162 Stanford. http://dx.doi.org/10.3115/981175.981197

Charles Chastain, (1975). *Reference and Context*. University of Minnesota Press, Minneapolis. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/185224

Herbert H. Clark, (1977). Bridging. In P. N. Johnson-Laird and P.C. Wason, editors. *Thinking: readings in Cognitive Science*, pages 411-420 Cambridge University Press, Cambridge. https://www.aclweb.org/anthology/T75-2034.pdf

Thomas Leonard and Fred Cummins, (2011). The Temporal Relation between Beat Gestures and Speech. *Language and Cognitive Processes* 26: 1457-1471. DOI: 10.1080/0169010965.2010.500218

Mariangiola Dezani-Ciancaglini, (1996). *Logical Semantics for Concurrent Lambda Calculus*. Proefschrift, Nijmegen Universiteit, Nijmegen,The Netherlands.

Miriam Eckert and Michael Strube, (1999). Resolving Discourse Deictic Anaphora in Dialogues.In *Proceedings of the EACL'99*, pages 37-44. http://dx.doi.org/10.3115/977035.977042

Jacob Eisenstein, Randall Davis, (2006). Gesture Features for Coreference Resolution. In S. Renals, S. Bengio, J. Fiscus, editors. *MLMI 2006*, pp. 154-165. DOI: 10.1007/11965152_14

Jay Eisenstein and Randall Davis, (2006). Gesture Improves Coreference Resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, pages 37-40, New York, June. https://doi.org/10.3115/1614049.1614059

Anna Esposito, Marc-Eric McCullough, Francis Quek (2001). Disfluencies in gesture: Gestural correlates to filled and unfilled speech pauses. Researchgate

Anna Esposito, Antonietta M. Esposito (2011). On Speech and Gestures Synchrony. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., Nijholt, A., editors. *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*. Lecture Notes in Computer Science, vol 6800. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-25775-9_25

Wan Fokkink (2000). *Introduction to Process Algebra.* Springer-Verlag Berlin, Heidelberg ISBN:978-3-540-66579-3

Kari Fraurud (1990). Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7 (Issue 4):395–433. https://doi.org/10.1093/jos/7.4.395

Bart Geurts, David I. Beaver, and Emar Maier, "Discourse Representation Theory", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta, editor, URL = <https://plato.stanford.edu/archives/spr2020/entries/discourse-representation-theory/>.

Jonathan Ginzburg (2012). *The Interactive Stance*. Oxford University Press, Oxford

Jonathan Ginzburg and Andy Lücking (2021a). I thought pointing is rude: a dialogue-semantic analysis of pointing at the addressee. In P. G. Grosz, L. Marti, H. Pearson, Y. Sudo, and S. Zobel, editors, *Proceedings of Sinn and Bedeutung 25*, pages 276-291. University College London and Queen Mary University of London, https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/937

Jonathan Ginzburg and Andy Lücking (2021b). Requesting clarifications with speech and gesture. *Conference Beyond Language: Multimodal Semantic Representation*. Virtual at the University of Groningen, held in conjunction with IWCS 2021.

Jonathan Ginzburg and Andy Lücking (2022). Leading voices. Dialogue semantics, cognitive science, and its polyphonic structure of multimodal interaction. Preprint in *Language and Cognition.* November 2022.

Gianluca Giorgolo and Frans Verstraten (2008). Perception of speech-and-gesture integration. In *Proceedings of the International Conference on Auditory-Visual Speech Processing 2008*, pages 31–36

Ramūnas Gutkovas (2014). *Languages, Logics, Types and Tools for Concurrent System Modelling*. Dissertation from the Faculty of Science and Technology 1392. Upsala Universitet, Upsala

Florian Hahn, Hannes Rieser (2011). Gestures Supporting Dialogue Structure and Interaction in the Bielefeld Speech and Gesture Alignment Corpus. In *Semdial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, pages 182-183, Los Angeles, California, 21-23 Sept. 2011

Hubert Haider (2006). Mittelfeld Phenomena (Scrambling in Germanic). In Martin Everaert and Henk van Riemsdijk, editors, *The Blackwell Companion to Syntax,Vol. III*, pages 204-275, Blackwell Publishing. http://dx.doi.org/10.1002/9780470996591.ch43

John A. Hawkins (1978). *Definiteness and Indefiniteness*. Lonndon, Croom Helm.

Klaus von Heusinger (2007) Accessibility and definite noun phrases. In M. Schwarz-Friesel, M. Consten, and M. Knees, editors. *Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference*, pages 123-145, John Benjamins, Amsterdam. https://doi.org/10.1075/slcs.86.12heu.

Jerry, R. Hobbs, 1978, Resolving Pronoun References. *Lingua 44*, (1978), 4:311-33 https://doi.org/10.1016/0024-3841(78)90006-28.

Jerry, R. Hobbs, 1979. Coherence and Coreference. *Cognitive Science 3*: 67-90. https://doi.org/10.1207/s15516709cog0301_4

Tilman, N. Höhle (1986). Der Begriff "Mittelfeld". Anmerkungen über die Theorie topologischer Felder. In: Schöne, Albrecht (Hrsg.) 1986. *Kontroversen alte und neue. Akten des VII. Internationalen Germanisten-Kongresses*, Göttingen 1985. Bd. 3: Weiss, Walter E., Wiegand, Herbert, E. und Reis, Marga (Hrsgg.),*Textlinguistik kontra Stilistik? –Wortschatz und Wörterbuch – Grammatische oder pragmatische Organisation von Rede?*, pages 329-340. https://doi.org/10.5281/zenodo.1169673

David Hume (1748). *An Enquiry Concerning Human Understanding*. The Liberal Arts Press, New York, 1955 edition.

Katja Jasinskaja, Elena Karagjosova (2020). Rhetorical Relations. In *The Wiley and Blackwell Companion to Semantics,* pages 1-29. DOI: 10.1002/9781118788516.sem061

Magnus Johansson (2010). *Psi-calculi: a framework for mobile process calculi: Cook your own correct process calculus - just add data and logic.* PhD thesis, Uppsala University, Division of Computer Systems, 2010.

Michael Johnston (2019). Multimodal Integration for Interactive Conversational Systems. In Sharon Oviatt et al., editors. *The Handbook of Multimodal-Multisensor Interfaces*, Vol. 3, pages 21-77, ACM Books. DOI:10.1145/3233795.3233798 Corpus ID: 20028979

Michael Johnston (1998). Unification-based multimodal parsing. In *Proceedings of the 36th Annual Meeting on Association for Computational Linguistics – Volume I*. Annual Meeting of the ACL, pages 624–630, Montreal, Quebec, Canada: Association for Computational Linguistics

Michael Johnston, David McGee, Sharon L. Oviatt, James A. Pittman, and Ira Smith (1997). Unification-based Multimodal Integration. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics. European Chapter Meeting of the ACL*, pages 281–288, Madrid, Spain: Association for Computational Linguistics

Hans Kamp and Uwe Reyle (1993). *From Discourse to Logic.* Kluwer, Dordrecht.

Andrew A. Kehler (2002). *Coherence, Reference, and the Theory of Grammar*. CSLI Publications. http://www.loc.gov/catdir/description/cam0210/99036170.html

Andrew A. Kehler, Laura Kertz, Hannah Rohde, Jeffrey, L. Elman (2007). Coherence and Coreference revisited. *Journal of Semantics* 25: 1–44. http://dx.doi.org/10.1093/jos/ffm018

Andrew A. Kehler & Hannah Rohde (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39: 1-37. https://doi.org/10.1515/tl-*2013*-0001

Ruth Kempson, Eleni Gregoromichelaki, Ronni Cann, Stergios Chatzikyriakidis (2016). Language Mechanism for Interaction. *Theoretical Linguistics*, Bd. 42, Heft 3-4 (October 2016), pages 203-275. DOI 10.1515/tl-*2016*-0011.

Adam Kendon (2004). *Gesture. Visible Action as Utterance*. Cambridge University Press, Cambridge

Jeffrey C. King and Karen S. Lewis. "Anaphora", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta, editor. URL=<https://plato.stanford.edu/archives/fall2018/entries/anaphora/>.

David B. Koons, Kristinn, R.Thórisson, Carlton C. Sparrell (1993), Integrating simultaneous input from speech, gaze & hand gestures. In M. T. Maybury, editor. *Intelligent Multimedia Interfaces*, pages 257-276, AAAI Press/MIT Press, Cambridge, MA.

Stefan Kopp, Paul Tepper & Justine Cassell (2004). Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 97–104. http://dx.doi.org/10.1145/1027933.1027952

Emil Krahmer and Paul Piwek (2013). *ITRI-00-13 Introduction: Varieties of Anaphora*. ResearchGate. https://www.researchgate.net/publication/228712984_ITRI-00-3_Introduction_Varieties_of_Anaphora

Jeremy Kuhn (2022). A dynamic semantics for multimodal communication. In *13th International Conference DHM 2022,* LNCS 13319, pages 231-243

Alex Lascarides & Matthew Stone (2009). A Formal Semantic Analysis of Gesture. *Journal of Semantics* 26(4): 1–57. http://dx.doi.org/10.1093/jos/ffp004

Peter Lasersohn (1999). Pragmatic Halos. *Language*, Vol. 75, No. 3 (Sept. 1999) pages 522-551

http://links.jstor.org/sici?sici = 0097-507%28199909%2975%3A%3C522%3APH%3E2.0.

Insa Lawler, Florian Hahn & Hannes Rieser (2017). Gesture Meaning Needs Speech Meaning to Denote. A Case of Speech-Gesture Meaning Interaction. In Christine Howes & Hannes Rieser, editors, *Proceedings of FADLI 2017*, pages 43-47. https://www.researchgate.net/publication/317339368_Gesture_meaning_needs_speech_meaning_to_denote_-_A_case_of_speech-gesture_meaning_interaction

Sebastian Loebner (1987). Definites. *Journal of Semantics 4*: 279-326. https://www.researchgate.net/publication/31273905_Definites

Andy Lücking, (2013). *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. De Gruyter, Berlin/Boston. https://doi.org/10.1515/9783110301489.

Andy Lücking and Jonathan Ginzburg (2020). Towards the score of communication. In Proceedings of SemDial 2020/WatchDial: *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue.* At: Brandeis University, Waltham (Watch City), MA.

Andy Lücking and Jonathan Ginzburg (2021). Leading voices: the polyphonic structure of multimodal interaction. In *Language and Cognition* 15(1), pp. 1-28

Andy Lücking, Kirsten Bergmann, Florian Hahn, Stefan Kopp and Hannes Rieser (2012). Data-based Analysis of Speech and Gesture: The Bielefeld Speech and Gesture Alignment Corpus

(SaGA) and its Applications. *Journal on Multimodal User Interfaces*, pages 92-98 Springer, Berlin/Heidelberg. http://dx.doi.org/10.1007/s12193-012-0106-8

David McNeill (1992). *Hand and Mind. What Gestures reveal about Thought.* Chicago and London, The University of Chicago Press

David McNeill (2005). *Gesture and Thought*. University of Chicago Press, Chicago

Gregor Mehlmann and Elisabeth André (2012). Modelling multimodal integration with event logic charts. In *Proceedings of the International Conference on Multimodal Interfaces (ICMI)*, pages125-132. Santa Monica, CA. https://www.researchgate.net/publication/262410953_Modeling_multimodal_integration_with_event_logic_charts

Robin Milner (1999). *Communicating and mobile systems: the π-calculus.* Cambridge University Press, Cambridge. http://dl.acm.org/citation.cfm?id=329902

Reinhard Muskens (1996). Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy 19*, pages 143-186

Costanza Navarretta (2011). Anaphora and Gestures in Multimodal Communication. In *Proceedings of the 8th Discourse Anaphora and Anaphora resolution Colloquium (DAARC 2011),* pages 2-10. https://www.researchgate.net/publication/236649685_Anaphora_and_Gestures_in_Multimodal_Communication

Costanza Navarretta (2021). Speech Pauses and Pronominal Anaphors. *Frontiers of Computer Science* 3. https://doi.org/10.3389/fcomp.2021.659539

Joachim Parrow (2001). An introduction to the π-calculus. In J. A. Bergstra, A. Ponse and S. A. Smolka, editors, *Handbook of Process Algebra,* pages 479–545. Elsevier, Amsterdam.

Pamela, Perniss & Asli Özyürek. 2015. "Visible Cohesion: A Comparison of Reference Tracking in Sign, Speech, and Co-Speech Gesture". In: *Topics in Cognitive Science* 7(1), pages 36–60.

Pfleger, Norbert (2007). *Context-based Multimodal Interpretation: An Integrated Approach to Multimodal Fusion and Discourse Processing.* Dissertation, Universität des Saarlandes.

Massimo Poesio and David Traum, 1997. Conversational Actions and Discourse Situations, *Computational Intelligence*,v. 13, n.3.

Massimo Poesio (1995). A model of conversational processing based on micro conversational events. In *Proceedings of the 17th Annual Conference of the Cognitive Society*, pages 678-703. Pittsburgh, July 1995. https://www.researchgate.net/publication/2419229_A_Model_of_Conversation_Processing_Based_on_Micro_Conversational_Events

Massimo Poesio, (2016). Linguistic and Cognitive Evidence about Anaphora. In M. Poesio, R. Stuckardt and Y. Versley, editors. *Anaphora Resolution. Algorithms, Resources, and Applications,* pages 23-55, Springer-Verlag Berlin, Heidelberg.

https://www.researchgate.net/publication/305873349_Linguistic_and_Cognitive_Evidence_About_Anaphora.

Massimo Poesio and Hannes Rieser (2010). Completions, Coordination, and Alignment in Dialogue. *Dialogue and Discourse* 1: 1–89.

Massimo Poesio and Hannes Rieser (2011). An Incremental model of Anaphora and Reference Resolution Based on Resource Situations. *Dialogue and Discourse* 2(1*)*: 235-277. doi: 10.5087/dad.2011.110

Massimo Poesio and Hannes Rieser (2011a). Anaphora and direct reference: Empirical evidence from pointing. In *Proceedings of DiaHolmia, the 13th Workshop on the Semantics and Pragmatics of Dialogue,* pages 35-43, Stockholm, June 2009.

Massimo Poesio, Roland Stuckardt, Yannik Versley and Renata Vieira (2016). Early Approaches to Anaphora Resolution: Theoretically Inspired and Heuristic-Based. In M. Poesio, R. Stuckardt, R. and Y. Versley, editors. *Anaphora Resolution. Algorithms, Resources, and Applications*, pages 55-97. Springer-Verlag Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-662-47909-4_3

Massimo Poesio and Renata Vieira (1998). A corpus-based investigation of definite description use. *Computational Linguistics* 24(2*)*: 183-216. https://www.researchgate.net/publication/2289214_A_Corpusbased_Investigation_of_Definite_Description_Use

Johannes Åman Pohjola (2020). PSI-CALCULI REVISITED : CONNECTIVITY AND COMPOSITIONALITY. In : *Logical Methods in Computer Science.* Vol. 16, issue 4, 16 :1 – 16 :28 https : //lmcs.episciences.org/

Ellen F. Prince (1981). Toward a taxonomy of given-new information. In P. Cole (editor) *Radical Pragmatics*, pages 223-255, Academic Press, New York.https://www.bibsonomy.org/bibtex/258c8a5f9a8006674c7d0c66a74c86e5f/cbrewster

W. Pouw, & J. A. Dixon (2019). Quantifying gesture-speech synchrony. In: Grimminger, A., editor. Proceedings of the 6th Gesture and Speech in Interaction – GESPIN 6, pages 75-80. Paderborn: Universitätsbibliothek Paderborn. doi:10.17619/UNIPB/1-815

Matthew Purver and Jonathan Ginzburg, Patrick Healey (2001). On the Means for Clarification in Dialogue. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*, pages 235-255. DOI: 10.1007/978-94-010-0019-2_11

Hans Reichenbach, (1966). *Elements of Symbolic Logic*. First Free Press Paperback Edition. (Feb. 2010).

Hannes Rieser (2010). Factoring Out a Gesture Typology from the Bielefeld Speech-Gesture-Alignment Corpus«. In: Stefan Kopp und Ipke Wachsmuth, editors. *Proceedings of GW 2009: Gesture in Embodied Communication and Human-Computer Interaction.* pages 47–60, Berlin, Heidelberg: Springer,

Hannes Rieser (2011). Gestures Indicating Dialogue Structure. In *Semdial 2011 Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, pages 9-18, Los Angeles, California, 21-23 Sept. 2011

Hannes Rieser (2015). When Hands Talk to Mouth. Gesture and Speech as Autonomous Communicating Processes. In *Proceedings of Semdial 2015*, Gothenburg. http://semdial.org/anthology/Z15-Rieser_semdial_0017.pdf

Hannes Rieser (2017). A Process Algebra Account of Speech-gesture Interaction. Revised and Updated Version. In: Chr. Howes and H. Rieser, editors. *FADLI (ESSLLI 2017 workshop) Proceedings* 2017, pages 67-71. https://www.researchgate.net/publication/317637747_A_Process_Algebra_Account_of_Speech-gesture_Interaction

Hannes Rieser (2024). Small-scale Sub-turn Communication. A Study on the Interaction of Gaze, Speech, Gesture, Back-channels, Interactional Anaphora, Role Changes, and Repair within a Turn, pages 1-25. Under review for *Multi-modal Communication*.

Hannes Rieser and Insa Lawler (2020). Multi-modal meaning: An empirically founded process algebra approach. *Semantics and Pragmatics 13* (8). Early Access Version. http://dx.doi.org/10.3765/sp.13.8

Hannes Rieser and Massimo Poesio (2009). Interactive gesture in Dialogue. A PTT Model. *Proceedings of SIGDial 2009. The 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*, pages 87-96, Queen Mary University of London. https://aclanthology.org/W09-3912.pdf

Hannah Rohde (2018). Pronoun interpretation and production. In C. Cummins & N. Katsos, N., editors. *Oxford Handbook of Experimental Semantics and Pragmatics,* pages 452-473, Oxford University Press, Oxford. DOI:10.1093/oxfordhb/9780198791768.013.21

Hannah Rohde & Andrew Kehler (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience* 29:8, pages 912-927. https://doi.org/10.1080/01690965.2013.

Wikipedia entry *process calculus*. Page last edited on 16th July 2022

Wikipedia entry *π-calculus*. Page last edited on 16th June 2022

Jeanette M. Wing (2002). FAQ on π-Calculus. Online Resource. Columbia University.

Angelika Wöllstein, 2018. Topologisches Satzmodell. In: Jörg Hagemann, Sven Staffeldt, (Hrsgg.), Syntaxtheorien im Vergleich, pages 145-166, Stauffenberg 2018, Tübingen.