Pointing in Dialogue

Hannes Rieser
SFB 360 “Situated Artificial Communicators”
Bielefeld University
Postfach 10 01 31,
D-33501 Bielefeld, Germany
Hannes.Rieser@Uni-Bielefeld.de

Abstract

A new approach based on experiments aiming at the integration of content originating from pointing plus definite descriptions (objects called “CDs”) in dialogue is presented. We develop it against the background of the early semiotic positions of Wittgenstein, Peirce, and Quine, the intentionalism of Kaplan, *Neo*-Peirce-Wittgenstein-Quine approaches and “mixed” points of view. Our experimental data show that pointing gestures are polysemous and polymorphic entities. Polysemy of CDs is due to their different functions, pointing to objects (“object demonstration”) and pointing to regions (“restrictor demonstration”), polymorphism originates from different positions wrt the utterance. Gesture information and expression meaning are integrated into a syntax-semantics interface using constraint-based syntax and type-logical semantics. Finally, it is shown that an underspecification account for the syntax-semantic interface can be set up along the lines of Logical Description Grammars.

1 Overview and Introduction

Ch. (1) deals with constraints of pointing gestures and introduces the “↘”-notation for gesture strokes. (2) overviews Peircian to *post*-Kaplan approaches on demonstration and reference. (3) describes gesture experiments. (4) is on “object demonstration” and “restrictor demonstration”. The set-up of the interface combining constraint-based grammar and type-logics for the integration of multimodal content is specified. (5) deals with the logical form of CDs. (6) shows that an underspecification account for the syntax-semantic interface can be set up along the lines of Logical Description Grammars. Discussion and future research come in (7).

Demonstration is bound up with reference (see e.g. Levinson 1995). Demonstrations (characteristically pointings) can accompany simple or complex referring expressions. We represent the stroke of hand gestures (see Mc Neill 1992) and similar devices by “↘”. Up to section (4) the nature of the↘ sign will be left to intuition. It occurs at the position indicated in the string and marks gesture stroke occurrence.

Examples of CD-expressions:

(1) Grasp↘this/that.

(2) *grasp.

(3) *↘

(4) Grasp ↘this/that yellow bolt.

(5) * Grasp this/that yellow bolt.

(6) Grasp the yellow bolt.

(7) ↘This yellow bolt, grasp it.

(8) All the bars get fixed by ↘this yellow bolt.

(9) ↘This yellow bolt doesn’t fix all the bars.

(10) ↘This yellow bolt must fix all the bars left of it.

(7) shows a CD-expression taken up by an anaphora; *it* comprises the content provided by *this yellow bolt* and the "↘" together. (8), (9) and (10) show scope interactions of CDs and either quantifier phrases ((8) and (9)), negation or modals ((9) and (10)).

(11) $\searrow_1$This/that is different from $\searrow_2$this/that.

(12) $\searrow_1$This/that, $\searrow_2$this/that, and $\searrow_3$this/that goes into the box.

(11) and (12) have different occurrences of ↘. Anaphora, scope-like effects and multiple occurrences of ↘s are among the most convincing cases for an integrated treatment of demonstratives and demonstrations. Three things have to be considered if we want to get a fuller understanding of CDs: (a) demonstrations and their timing wrt to speech, (b) the structure of verbal expressions going into CDs, and (c) the interaction of demonstrations and expressions, i.e. what they individually contribute to the semantic or pragmatic information provided by CDs *in toto*.

2 Related Research: From Peirce to Kaplan and Beyond

A unified account of CDs will opt for a compositional semantics to capture the information coming from the verbal and the visual channel. Peirce (1932, p. 166) and Wittgenstein (1958, p. 109) consider pointing as part of the symbol. Quine (1960) is committed to a similar point of view.

At present one can distinguish three mainstream philosophical attitudes towards CDs: The line of thought farthest off the Peirce-Wittgenstein-Quine line is the intentionalism associated with Kaplan's late work (1989b). There demonstration is taken as a mere externalisation of intention. It is intention that determines reference. Later on Reimer (1992), Dever (2001), King (2001) and Borg (2000) have supported this line.

Neo-Peirce-Wittgenstein-Quinians (*neo*-PWQians) exist as well. A case in point is McGinn (1981). He holds that in establishing reference the gesture functions as part of the language. Larson and Segal (1995), Hintikka (1998) and ter Meulen (1994) also sympathize with this view. However, there is no *neo*-PWQian approach explicitly representing pointing gestures and providing a semantics for them.

Still, there is a group of "in-betweeners", stressing the contribution of intention and demonstration in fixing demonstrative reference: Among these are the early Kaplan (1989a), D. Braun (1994, 1996) and Lepore and Ludwig (2000). Of these only D. Braun explicitly represents demonstrations in his 1996 approach.

The literature referred to rests almost exclusively on intuitions concerning pointings to single, visible objects. However, pointing is a more varied phenomenon as experiments show.

3 Gesture Experiments Using Simple Reference Games

Reliable intuitions for demonstration are hard to come by. Therefore we use experimental data called "simple reference games" (see Kühnlein and Stegmann (2004)). These are set up in the following way: We have two subjects, description-giver and object-taker. The description-giver must give sufficient information to the object-taker to make him identify one of the objects on a table between them.

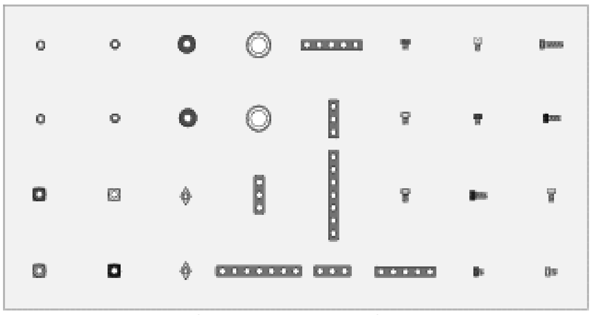


Fig.1: One type of clustering used for gesture experiments in simple reference games (Kühnlein and Stegmann (2004)).

The description-giver selects objects providing the identification information by verbal or gestural means. The object-taker may grasp the object singled out and lifts it from the table. Every game was video-taped from two perspectives, the description giver's as well as a neutral one (fig. 2). There were two types of clusterings, sameness of colour *vs* sameness of form. The games in two video-films have been annotated using the TasX-annotator (Milde and Thies (2002)) (fig. 3); location of gesture stroke, syntax and semantics of NL-expressions and the structure of discourse have been considered. Lack of space prevents us from

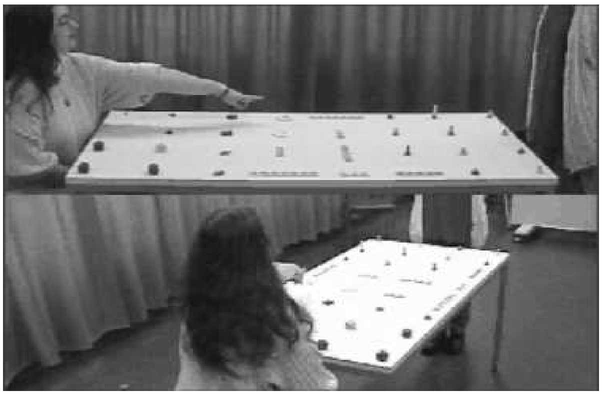
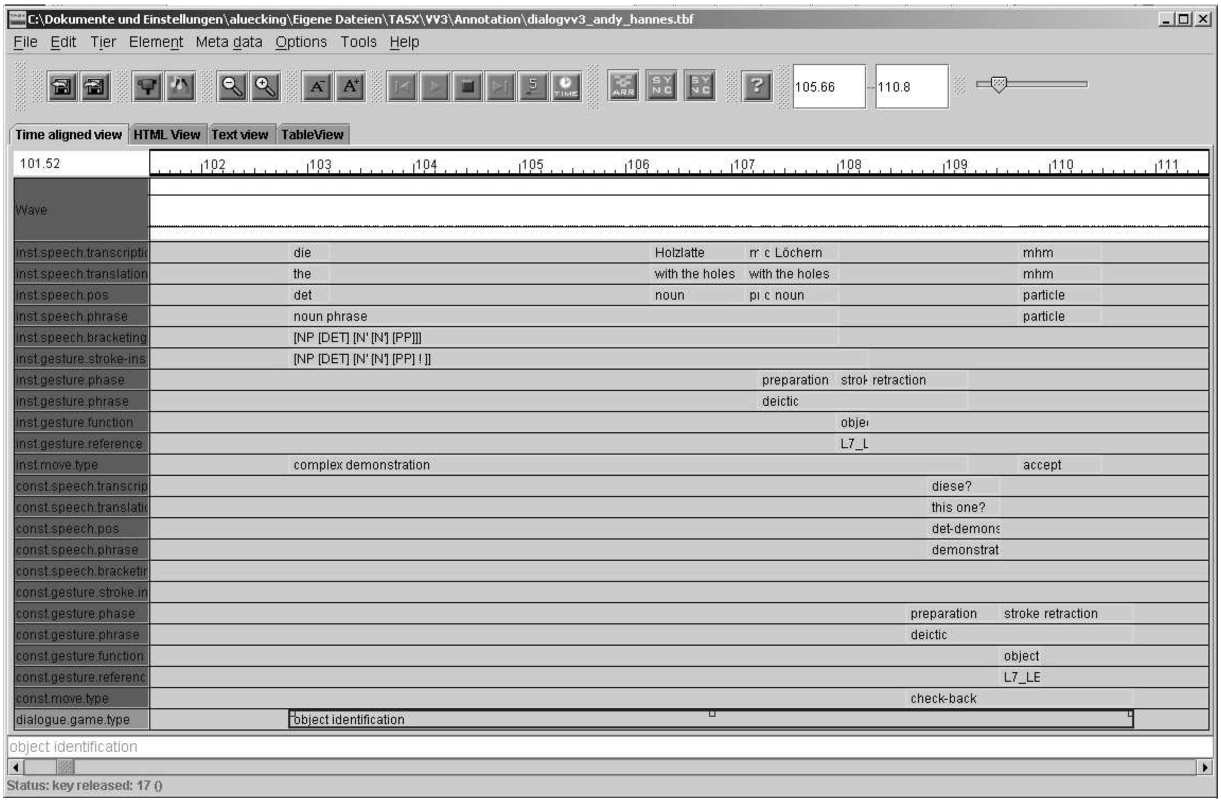


Fig. 2: Video-taped pointing gesture from two perspectives

going into descriptions of graspings and of dialogue structure here. Also, discussion of interesting statistical details must be left out (but see Lücking, Rieser, Stegmann (2004)).

4 CDs and Definite Descriptions: Object Demonstration and Restrictor Demonstration

There is a debate on whether definite NPs plus demonstrations can be regarded as definite descriptions (Kaplan 1989 a,b; Rieber 1998). A plea for taking CDs as definite descriptions comes from Quine (1960, ch. III). For us, definite NPs are definite descriptions to which demonstrations add content, either by specifying an object independently of the definite description or by narrowing down the description's restrictor. We call the first technique "object demonstration" and the second one "restrictor demonstration". Graspings are the clearest cases of object demonstration.



C:\Dokumente und Einstellungen\aluecking\Eigene Dateien\TASX\VV3\Annotation\dialogvv3_andy_hannes.tbf
File Edit Tier Element Meta data Options Tools Help
5
TIME
ARR
SY NC
SY NC
?
105.66 – 110.8
Time aligned view
HTML View
Text view
TableView
101.52
102 103 104 105 106 107 108 109 110 111

Wave								
inst.speech.transcripti	die	Holzlatte	rr c Löchern					mhm
inst.speech.translation	the	with the holes	with the holes					mhm
inst.speech.pos	det	noun	pı c noun					particle
inst.speech.phrase	noun phrase							particle
inst.speech.bracketing	[NP [DET] [N' [N'] [PP]]]							
inst.gesture.stroke-ins	[NP [DET] [N' [N'] [PP] !]]							
inst.gesture.phase			preparation	strok retraction				
inst.gesture.phrase			deictic					
inst.gesture.function				obje				
inst.gesture.reference				L7_L				
inst.move.type	complex demonstration							accept
const.speech.transcrip						diese?		
const.speech.translati						this one?		
const.speech.pos						det-demons		
const.speech.phrase						demonstrat		
const.speech.bracketir								
const.gesture.stroke.in								
const.gesture.phase					preparation		stroke retraction	
const.gesture.phrase					deictic			
const.gesture.function							object	
const.gesture.referenc							L7_LE	
const.move.type					check-back			
dialogue.game.type	object identification							

object identification

Status: key released: 17 ()

Fig.3: TasX-annotated dialogue game *object identification* comprising instructor's complex demonstration, constructor's check-back and instructor's acceptance

The Syntax-semantics Interface Used

Figs. 4 and 5 show the components of the interface used. Fig. 4 sketches the interpreted grammar, Fig. 5 its empirical coverage.

Following Sag and Wasow (1999), the interface uses constraint based grammar. It combines syntactic and semantic information in one AVM-format. Because of technical reasons we use type-logics for semantic representation, i.e. in the values of SEM-attributes, and interpret Sag and Wasow's ⊗-operator (Sag and Wasow (1999), p. 116) as functional application. Due to limits of space, we only represent logical forms here and neglect linking proper.

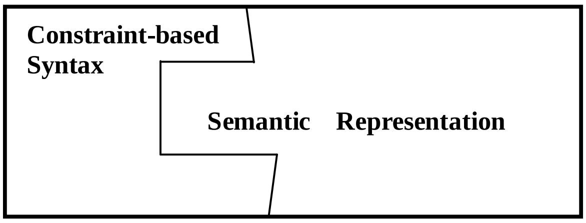


Fig. 4: Components of the constraint-based syntax-semantics interface

Following Searle and Vanderveken (1989), directives consist of an illocutionary role indicator and the proposition. They do not have truth conditions in the classical sense. The interpretation specifies satisfaction conditions for them, i.e. it singles out successful directives wrt models. The generalised notion of satisfaction used here is Recanati's (1993). In this context the relation of the definite NP to "its" demonstration is of decisive importance: If both serve their referential tasks, one condition for the satisfaction of the directive is met. The model provides conditions both for the illocutionary role and the associated proposition. Translation of the type-logical format into dynamic semantics is possible in principle (see Eijk and Kamp (1997)).

The model handles canonical uses of CDs, even cases where the demonstration follows the definite NP. Its coverage subsumes real world experimental data as well as *VR*-data.The working of the semantic component will be illustrated here discussing the toy example *Grasp the yellow bolt!*. The meaning of directives is identified with their illocutionary forces.

The imperative is represented by the illocutionary role marker F_{dir} operating on the open formula *grasp(u, v)*, the difference between imperatives and other finite forms being expressed by "F_{dir}". $\lambda P.P$ (you) ⊗ $\lambda u(F_{dir}$ (grasp(u, ιz(yb(z))))) gives us the representation of the whole directive F_{dir}(grasp(you, ιz(yb(z)))) to paraphrase as "There is exactly one yellow bolt, grasp it, addressee!"

So far we have not integrated demonstrations.

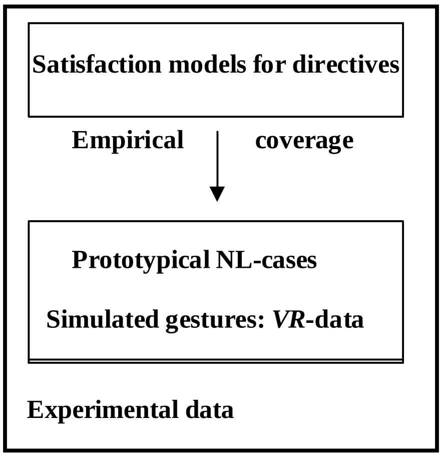

Satisfaction models for directives
Empirical coverage
Prototypical NL-cases
Simulated gestures: VR-data
Experimental data

Fig. 5: Models for the constraint-based interface and their empirical coverage

5 Logical Forms for Multi-modal Content

5.1 Integrating Demonstrations into Descriptions

Before we show how to represent demonstrations together with descriptions, we specify our main hypotheses concerning their integration. These are related to content, compositionality, *i.e.* role in building up truth-functional content for the embedded proposition, and scope of gesture. Hypothetically then, demonstrations (a) act like verbal elements in providing content, (b) interact with verbal elements in a compositional way, (c) may exhibit forward or backward dynamics, (d) involve a continuous movement over a time interval, comparable to suprasegmentals, and (e) can be described using discrete entities like the "↘".

Demonstrations introduce objects independently of the definite description ("object demonstration") or act as restrictors of descriptions (“restrictor demonstration”). Intuitively, this will invest demonstrations with two functions. However, this does not yet entail that they are ambiguous between two readings, regardless of the position of the stroke. There still could emerge arguments for a division of labour concerning semantics and pragmatics. Before we enter modelling gesture stroke, we report on the findings concerning stroke position from the empirical data. All findings are corroborated by statistical material (see Lücking, Rieser, Stegmann (2004)):

(1.) Stroke positions can be *pre*-N’, *on*-N’ or *post*-N’. Here data exhibit greater variation than commonly assumed: Demonstration does not occur before referring expressions unexceptionally. The proto-typical stroke position is on-N’. (2.) Demonstrations can fail. Descriptions they are associated with can denote nevertheless. In particular: satisfiable object demonstrations and corresponding non-satisfiable descriptions yield false propositional content. (3.) Object demonstration and restrictor demonstration are clearly separable and seem to cover together the classifiable data. (4.) Stroke positions do not indicate object demonstration or restrictor demonstration preferences. (5.) A failing description can be completed by a restrictor demonstration. We can have elliptical descriptions in CDs. (6.) In case the description is satisfied on its own, a successful restrictor demonstration is redundant. (7.) Non-classifiability can arise with respect to stroke, direction or role of demonstration, description or completed description.

The central problem is of course how to interpret demonstrations. This question is different from the one concerning the ↘’s function tied to its position in the string. We base the discussion on the following examples showing different empirically found ↘ positions and turn first to “object demonstration”:

(13) Grasp ↘this/that yellow bolt.

(13a) Grasp this/that ↘yellow bolt.

(13b) Grasp this/that yellow ↘bolt.

(13c) Grasp this/that yellow bolt↘.

5.2 Object Demonstration

Our initial representation for the propositional frame of the demonstration-free expression is

(14) $\lambda P\ \lambda u(P\ \ \lambda v\ F_{dir}\ (grasp(u, v)))$.

The ↘ provides new information. If the ↘ is independent from the reference of the definite description the only way we can express that is by extending (14) with $v = y$:

(15) $\lambda P\ \lambda u\ \lambda y(P\ \ \lambda v\ \ F_{dir}\ (grasp(u, v)\ \wedge\ (v = y)))$.

The idea tied to (15) is that the reference of *v* and the reference of *y* must be identical, regardless of the way in which it is given. Intuitively, the reference of *v* will be given by the definite description $\iota z(yb(z))$ and the reference of *y* by the ↘. We could also work with a free variable, which, however, would have a different effect (see below).

The Compositionality Problem Concerning Strokes

(15) or the free variable solution may be interesting options for type-logical expressions integrating referential expressions and demonstrations. However, an intuition frequently put forth is that demonstrations to objects act like constants in standard logical notation. Whichever route we want to follow, one thing is common to the three solutions: demonstrations are taken as referring terms, that is, we can represent them as either

(16) (a) $\lambda P \lambda x.P(x)$ (bound variable)

(b) $\lambda P.P(x)$ (free variable)

(c) $\lambda P.P(a)$ (constant)

(a), (b) and (c) do different things: (a) and (b) contribute content *via* an assignment, whereas (c) does so *via* the model’s interpretation function.

In order to get a logical form for the whole directive, we must decide on the position of the ↘ in the string. We opt for (13), *Grasp* ↘*this/that yellow bolt.*, which intuitively indicates that the reference of the ↘ is independent of the reference of the definite description *this/that yellow bolt*.

The bracketing assumed for the string is roughly

(17) [grasp [↘ this/that yellow bolt]].

This implies we have to find a representation for *grasp* which combines with $\lambda P.P(a)$ first, followed by the definite description. A workable solution for this problem is (18), as the derivation based upon it shows:

(18) $\lambda Q \lambda P\, \lambda u\, (P\ (Q\ (\lambda y\ \lambda v F_{dir}\, (grasp(u, v) \wedge (v = y)))))\ \lambda P.P(a)$ /*[grasp ↘]

(19) $F_{dir}\, (grasp(you,\ \iota z(yb(z)))\ \wedge\ \iota z(yb(z)) = a)$.

What can one say about (18)? There the reference is coded twice, once via the pointing gesture $\lambda P.P(a)$ and once via the description $\iota z(yb(z))$. The information exchange, so to speak, maintains a security principle. In most empirical data, however, demonstrations and verbal information show a sort of "division of labour". We now turn to these cases.

5.3 Restrictor Demonstration

(13a) and (13b) above are the prototypical cases where demonstration is embedded into the description, hence the only thing that matters there is the set up of the description. Object demonstration case and restrictor demonstration case are similar insofar as information is added. In the object demonstration case, this is captured by a conjunct with identity statement; in the restrictor demonstration case the ↘ contributes a new property narrowing down the verbally expressed one. The bracketing we assume for the string is roughly

(20) [[grasp] [this/that [↘yellow bolt]]].

As a consequence, the format of the description has to change. This job can be done by

(21) $\lambda D \lambda F \lambda P\, .\, P(\iota z(F(z)\ \wedge\ D(z)))$.

The demonstration "↘" in (13a) will then be represented simply by

(22) $\lambda y(y \in D)$,

where *D* intuitively indicates the demonstrated region in the domain. We use the ∈-notation here in order to point to the information from the other channel. Under "⊗" this winds up to

(23) $\iota z(yb(z) \wedge z \in D)$.

Intuitively, (23), the completed description, indicates "the demonstrated yellow bolt".

Different stroke positions come with different compositionality problems.

6 Polymorphism of ↘ Captured in an Underspecification Account

To see what the real problems are if we want to get a stab at multimodal semantics, consider the possible stroke positions marked in the labelled bracketing of example (13):

(24) $[_S\ [_S\ [_{VP}\ [_{Vinf}\ \text{grasp}]\ \searrow_{pre\text{-}NP}[_{NP}\ [_{Dem}\ \text{this/that}]\ \searrow_{pre\text{-}N'}\ [_{N'}\ \searrow_{pre\text{-}Adj}[_{Adj}\ \text{yellow}]\ \searrow_{pre\text{-}N}[_N\ \text{bolt}]\ \searrow_{post\text{-}N}]\ \searrow_{post\text{-}N'}]\ \searrow_{post\text{-}NP}]\ \searrow_{post\text{-}VP}]]\ {}^*\!\searrow_{post\text{-}S}]$

We have pre-occurrences and post-occurrences of ↘. The pre-occurrences are $\searrow_{pre\text{-}NP}$, $\searrow_{pre\text{-}N'}$, $\searrow_{pre\text{-}Adj}$, $\searrow_{pre\text{-}N}$; these are the post-occurrences: $\searrow_{post\text{-}N}$, $\searrow_{post\text{-}N'}$, $\searrow_{post\text{-}NP}$, $\searrow_{post\text{-}VP}$. $\searrow_{post\text{-}S}$ we consider as not well-formed. At the same time, every occurrence can be paired with at least two readings, that is where the polysemy comes from. Seen from the point of view of our type-logical formulas for "object demonstration" and "restrictor demonstration",

(25) $\lambda P.P(a)$ and (26) $\lambda y\ (y\ \varepsilon\ D)$

we get the problem that there emerges a clash between the "natural" context-free syntactic category of ↘ and ist semantic function. We won't solve that entirely here. Clearly, all the "post"-occurrences of ↘ are problematic in a way, nevertheless they do occur. By way of solution, we can take up a suggestion of Sag and Wasow's (1999) concerning underspecification and distinguish between descriptions, feature structures and models as follows: Descriptions can be underspecified, feature structures are complete in relevant respects and serve as models for linguistic entities. Underspecified descriptions are satisfied by sets of structures. Seen from this perspective, our discussion so far dealt entirely with the semantic side of structures. Now, we move on to descriptions.

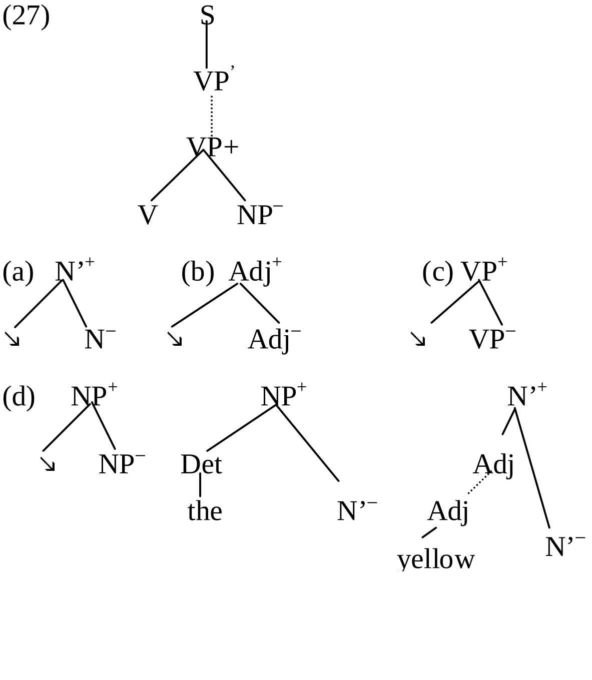
The underspecification model nearest the formalisms used here is the *Logical Description Grammars* (LDGs) account of Muskens (2001), which has evolved *inter alia* from Lexicalised Tree Adjoining Grammars and D-Tree Grammars. The structures derived within LDGs are compatible with those we get in constraint based formalisms using AVMs, hence there is no big methodological difference between the assumptions made about the theory of grammar here and LDGs. The intuitive idea behind LDGs is that, based on general axioms capturing the structure of trees, we work with a *logical description of the input*, capturing linear precedence phenomena, *lexical descriptions for words* and *elementary trees*. Then a *parsing-as-deduction* method is applied yielding semantically interpreted structures.

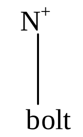
We provide the main steps of an LDG-reconstruction of the readings of (24) below.

A graphical representation of the input is given in (27). ‘+’ respectively ‘-‘ indicate components which can substitute (‘+’) as against nodes to be substituted (‘-‘). Dotted lines represent dominance and solid ones direct dominance. Models for the description in (27) are in the parsing-as-deduction approach derived by pairing off + and – nodes in a one-to-one fashion and by identifying the nodes thus paired. “I.e., each +node must be identified with a –node and *vice versa*, but not two +s and no two –s can be identified”. (Muskens (2001), p. 424). Words can come with several lexicalisations. The ↘-positions in (27) (a) to (d) have to be regarded as alternatives.

The *logical description of the input* has to provide the linear precedence regularities for our example *Grasp the yellow bolt!* Observe that these will be different from (27), which contains alternatives (a) to (d) for ↘-positions. (28) shows some precedence possibilities; the subindices on ↘ are provided to facilitate understanding.

(27)





(28)

(a) grasp < ↘$_{\text{pre-NP}}$ < this/that < yellow < bolt.

(b) grasp < this/that < ↘$_{\text{pre-N'}}$ < yellow < bolt.

(c) grasp < this/that < ↘$_{\text{pre-Adj}}$ < yellow < bolt.

(d) grasp < this/that < yellow < ↘$_{\text{pre-N}}$ < bolt.

(e) grasp < this/that < yellow < bolt < ↘$_{\text{post-N}}$.

(f) grasp < this/that < yellow < bolt < ↘$_{\text{post-N'}}$.

(g) grasp < this/that < yellow < bolt < ↘$_{\text{post-NP}}$.

(h) grasp < this/that < yellow < bolt < ↘$_{\text{post-VP}}$.

The description of the input must fix the underspecification range of the ↘. It has to come after the imperative verb, but that is all we need to state; in other words, that covers all the models depicted in (28).

The *lexical descriptions for words* will have to contain the type-logical formulas for compositional semantics. From the descriptions of the *elementary trees* we will get the basics for the “pairing-off” mechanism. It is easy to see that we can establish a proof for the NP with ↘$_{\text{pre-NP}}$ yielding (28)(a). (27) (a), (b) allow us to extend the NP with ↘$_{\text{pre-N'}}$ and ↘$_{\text{pre-Adj}}$, respectively. The “post”-versions could be generated by lexical anchors roughly similar to (27)(a) to (d). Lack of space prevents us from explaining here what has to be done at the type-logical level to ensure compositionality and well-formedness.

7 Discussion and Future Research

One of the central questions is of course whether there is an alternative to the *neo*-PWQian point of view and the ensuing methodology. A PWQian approach leads quite naturally to an integrated theory. A viable alternative might be to try an approach stressing the difference (!) between NL-expression and demonstration and to capture the role of demonstration in a different way, perhaps solely *via* the semantic model for the formal description chosen. Seen from this perspective, demonstration is an object with semantic impact but it is not part of the language. By and large this would be a Kaplan point of departure.

Keeping within *neo*-PWQian assumptions, the following points concerning the approach described here seem worthy of mentioning and need further detailed study: How can polymorphism/polysemy of demonstration be handled best? Will Logical Description Grammars do all there needs to be done? And, which division of labour between semantics and pragmatics is the correct one for setting up a theory of CDs? In addition, describing the simple reference games familiar from the data in a real discourse games approach is a worthwhile target but the other problems have to be sorted out first.

References

Almog, J., Perry, J., Wettstein, H. (eds.): 1989, *Themes from Kaplan*. New York, Oxford: OUP

Borg, E.: 2000, Complex Demonstratives. In: *Philosophical Studies 97*: 229 – 249

Braun, D.: 1996, Demonstratives and Their Linguistic Meanings. In: *Nous* Vol. 30, Nr. 2, pp. 145- XX

Braun, D.: 1994, Structural characters and complex demonstratives. In: *Philosophical Studi*es, Vol. 74, Nr. 2, pp. 193-221

Dever, J.: 2001, Complex Demonstratives. In: *Linguistics and Philosophy*, Vol. 24, Nr. 3. pp. 271-330

Hintikka, J.: 1998, Perspectival Identification, Demonstratives and “Small Worlds”. In: *Synthese* 114; pp. 203 – 232

Kaplan, D.: 1989a, Demonstratives. In Almog et al. eds., pp. 481-563

Kaplan, D.: 1989b, Afterthoughts. In Almog et al. eds., pp.565-614

King, J.C.: 2001, *Complex Demonstratives. A Quantificational Account.* Cambridge, Mass.: MIT Press

Kühnlein, P. and Stegmann J.: 2004, *Referring to Objects in Simple Identification Tasks.* Technical report of the SFB 360

Larson R. and Segal, G.: 1995, Pronouns and Demonstratives. Ch. 6 of *Knowledge of Meaning*. The MIT Press: Cambridge, Mass, pp. 197 – 227

Lepore, E.: 2000, Semantics and Pragmatics of Complex Demonstratives. In: *Mind*, Oxford, Vol. 109, Nr. 434, pp. 199-240

Levinson, St. C.: 1995, *Pragmatics*. Cambridge: CUP

Lücking, A, Rieser, H. and Stegmann, J.: 2004, Statistical Support for the Study of Structures in Multi-Modal Dialogue. *Catalog04 Proceedings*, pp.

Mc Neill, St.: 1992, *Hand and Mind*. The University of Chicago Press: Chicago and London

Milde, J.-T. and Thies, A.: 2002, *The TasX environment: Owner’s Manual*. Ms, Bielefeld, Univ.

Muskens, R.: 2001, Talking about Trees and Truth-Conditions. In: *Journal of Logic, Language and Information*, pp. 417 - 455

Quine, W. van O.: 1960, *Word and Object*. MIT Press, ch. III, The Ontogenesis of Reference

Recanati, F.: 1993, *Direct Reference. From Language to Thought*. Oxford, UK: Blackwell

Rieber, St.: 1998, Could demonstratives be descriptions? In: *Philosophia*, Vol. 26, Nr. 1-2, pp. 65-79

Sag, I. A. and and Wasow, Th.: 1999, *Syntactic Theory: a formal introduction.* Stanford, Calif.: CSLI Public.

Searle, J. and Vanderveken, D.: 1989, *Foundations of Illocutionary Logic*. Cambridge: CUP

Stanley, J. and Gendler, S.: 2000, On Quantifier Domain Restriction. In: *Mind & Language*, Vol 15, pp. 219-261

Ter Meulen, A. G. B.: 1994, Demonstratives, indications and experiments. In: *The Monist*, Vol. 77, Nr. 2, pp. 239-256

Van Eijk, J. and Kamp, H: 1997, Representing Discourse in Context. In: Van Benthem *et alii* (eds.), *Logic and Language*. North-Holland, pp. 179 – 239.

Wexelblat, A.: 1998, Research Challenges in Gesture: Open Issues and Unsolved Problems. In: Wachsmuth, I. Fröhlich, M. (eds.): *Gesture and Sign Language in Human-Computer Interaction*. International Gesture Workshop Bielefeld, Germany, September 1997 Proceedings, pp. 1-13