

Master Thesis

Gesture Resolution and Definiteness

Jonathan Weinrich
Goethe University Frankfurt

handed in 24.08.2020
first uploaded 31.12.2021

Notes about this version

The contents of this document have not been edited since it was handed in as a master thesis in 2020. The only changes involve the omission of metadata required by university standards, such as a German language abstract or personal data, and the addition of this page.

Contents

1	Introduction	1
2	Discourse referents, frameworks and gestures	3
2.1	The nature of discourse referents	3
2.2	DRT	4
2.3	SDRT	8
2.4	The application of SDRT by Lascarides & Stone (2006, 2009)	13
3	Gestural discourse referents as pronouns	21
3.1	Speech-only antecedents	24
3.2	Gesture-only antecedents	30
4	Gestures with definite description semantics	39
4.1	Formal literature on definiteness	40
4.2	Definite descriptions, pronouns and the definite/indefinite distinction . . .	42
4.3	Existence and unique maximality	44
4.3.1	Partitives and general number	44
4.3.2	Felicity by existence condition or bridging alone	46
4.3.3	Unique maximality and relevance	50
4.3.4	Maximal by form or by subtype	55
4.3.5	Approaching formal relevance	57
4.4	The definite approach in a dynamic framework	64
4.5	Applying Donnellan's (1966) referential-attributive distinction	68
5	Gestural behavior unexplained by a definiteness account	71
5.1	Representative gestures	71
5.2	Overgeneration of antecedent members	75
6	Conclusion	78

1 Introduction

According to a proposal by Lascarides & Stone (2006, 2009) co-speech gestures, gestures co-occurring with simultaneous speech, can be analyzed by a semantic formalism in the framework of SDRT (Segmented Discourse Representation Theory; Asher & Lascarides, 2003). Gestures are analyzed by formulas which relate the depicted content to the spoken one, but several gestures may also be connected with each other. One example they provide for gestures building on other gestures is the discourse in (1):

- (1) a. and [took his] [HATCHET and
First: Speaker's right hand grasps left hand, with wrists bent.
Second: Speaker lifts poised hands above right shoulder.
- b. with] [a mighty SWEEP]
First: Hands, held above right shoulder, move back then forward slightly.
- c. [(pause 0.4 sec)] [SLICED the wolf 's stomach open]
First: Speaker turns head
Second: Arms swing in time with sliced; then are held horizontally at the left
(cited as in Lascarides & Stone, 2009, p. 400; originally from Kendon, 2004, p. 136)

As Lascarides & Stone (2009, pp. 425-427) analyze the discourse, the gestures co-occurring with the bracketed speech segments in (1) described by *First:* and *Second:* are connected by *rhetorical relations*. The crucial observation is that certain characteristics of one gesture are picked up by the following gesture: The imaginary object the hands are taking in the first gesture in (1a) is the handle of the hatchet, and it is still the same handle of the same hatchet in all subsequent gestures which is depicted to be held by the two hands of the speaker.

One of the concerns of Lascarides & Stone (2009) is therefore the proper way to formalize when properties of similar looking gestures (e.g. holding an imaginary object with both hands) can be interpreted in a coherent or even uniform way. One other is how to formalize constraints on the coherence of gestures with spoken content. Both points are addressed by the usage of *discourse referents*, semantic representatives which license the usage of pronouns. The main point of the analysis is therefore formalizing gestures in a way that allows for anaphoric dependencies, the relationship found with pronouns and many instances of definite descriptions. Definite descriptions are taken here to be phrases of the form *the NP* or NPs preceded by a possessive determiner. The properties of pronouns and definite descriptions are collapsed by the term *definiteness* in some accounts (e.g. Kadmon, 1990). The term *definiteness* as used in the title of this thesis is intended in this sense. In a less formal way they can be said to have in common that they

refer to given entities salient from context or previous discourse.

With the framework of Lascarides & Stone (2006, 2009) as a starting point this thesis investigates two possibilities of capturing instances of gestures involving definiteness: An approach where gestures behave like pronouns, and one where they behave like definite descriptions. In order to provide the necessary background information, Section 2 discusses the nature of discourse referents, how the discourse theory of DRT (Kamp & Reyle, 1993a, 1993b) incorporates them, how the framework SDRT (Asher & Lascarides, 2003) is built on top of DRT and how Lascarides & Stone (2006, 2009) use this background to account for gestures. This background is then put to use in Section 3, which examines the possibility of analyzing the discourse given parts of a gesture's semantics as pronouns. By employing large monologues defying the pronoun resolution constraints of SDRT and by minimal pairs with speech pronouns and definite descriptions it is established that an analysis as pronouns does not capture the data in question. Likewise rhetorical relations are shown not to capture all anaphoric data.

In Section 4 an alternative account based on definite descriptions is developed. After some background on previous literature (Section 4.1) and how to distinguish gestures as definite descriptions from other analyses (Section 4.2) empirical data is used to establish which conception of definite descriptions is plausible (Section 4.3). First it is established that in case of a plural antecedent the gestural data indicates that often only some entities of the antecedent are involved with an event and that every account involving definite descriptions must do so by a partitive analysis. The next part presents evidence that some gestures can only be used felicitously at all if the context provides entities suitable as an antecedent, while others only have to obey a bridging constraint of Lascarides & Stone (2006, 2009). Since partitives do not allow for testing the definiteness of gestures by simply testing if they contribute the unique maximal antecedent as such, a maximal antecedent will be justified by examples where a notion of relevance leaves open important implicit questions with speech partitives and gestural partitives alike. Having established that some notion of a maximal antecedent is involved with the gestures of interest, subsequent discussion allows to decide that the denotation of a gesture should be a contextually unresolved predicate rather than an interpreted version. In order to complete the fundamental work for a definite description approach first steps towards a more formal account of the notion of relevance involved with testing maximal antecedents are undertaken. During the course of Section 4 an algorithm for (mainly definite) gestural resolution is developed.

Section 4.4 applies the findings to Lascarides & Stone's (2006, 2009) version of SDRT, while Section 4.5 makes concluding remarks about the two kinds of definite

description uses, attributive and referential (Donnellan, 1966), in gestures.

Towards the end Section 5 discusses problems encountered with the account of Section 4, and while possible solutions to some aspects are pointed out, still some points will have to remain unexplained or at least slightly *ad hoc*. The phenomena involve non-exhaustive inferences with definite gestures and antecedents the algorithm of Section 4 predicts to be available but which are not. Section 6 concludes the thesis with an overall evaluation.

2 Discourse referents, frameworks and gestures

2.1 The nature of discourse referents

Karttunen (1971) employs the term *discourse referent* in order to discuss constraints on coreference, especially across sentence or clause boundaries. He uses examples like the following to illustrate the existence of such constraints:

- (2) a. Bill has [a car]_i. It_i / [the car]_i / [Bill's car]_i is black.
b. Bill doesn't have [a car]_i. *It_i / [the car]_i / [Bill's car]_i is black.
(Karttunen, 1971, p. 4; formatting modified)
- (3) I don't believe that Mary had [a baby]_i and named her_i Sue. *[The baby]_i has mumps.
(Karttunen, 1971, p. 18; formatting modified)

In (2b) the denial of the existence of a car of Bill's seems to prevent coreferential phrases or pronouns, although coreference is readily established in the positive case (2a). In (3) the pronoun *her* can successfully establish coreference with *a baby*, but *the baby* cannot do so anymore. Karttunen (1971) considers the constraints in question a linguistic phenomenon since they do not seem to be affected by world knowledge:

- (4) a. Bill saw [a unicorn]_i. [The unicorn]_i had a gold mane.
b. Bill didn't see [a unicorn]_i. *[The unicorn]_i had a gold mane.
(Karttunen, 1971, p. 5; formatting modified)

Heim (1982) provides examples which further illustrates the linguistic behavior of such discourses:

- (5) a. (i) I dropped ten marbles and found all of them, except for one.
(ii) It is probably under the sofa.
b. (i) I dropped ten marbles and found only nine of them.

(ii) ?It is probably under the sofa.

(Heim, 1982, p. 21; attributed to Barbara Partee)

(6) a. John has a spouse. She is nice.

b. ?John is married. She is nice.

(Heim, 1982, p. 24)

Heim (1982, pp. 21-24) observes that although in the two-sentence sequences (5a) and (5b) the truth conditions of the first sentences (5a-i) and (5b-i) are identical, the pronoun *it* can hardly be used in (5b-ii) as coreferential with the missing marble. The same pattern can be seen in (6): Having a spouse and being married are two interchangeable predicates, but only *has a spouse* licenses the coreferential pronoun *she*.

Karttunen (1971, p. 5) chooses to define discourse referents as something introduced by indefinite noun phrases “just in case it justifies the occurrence of a coreferential pronoun or a definite noun phrase later in the text”. Heim (1982, p. 250) points out that the term *discourse referent* is not a term for individuals (like the referent of *John* may be assumed to be the name-bearing individual John himself), but a *representative* of individuals. Drawing on Karttunen’s (1971) work, she formalizes the phenomena presented above by dynamic *file cards*, discourse units collecting information by criteria related to one common referential index (Heim, 1982, p. 283), and their lifespan by properties of logical form (Heim, 1982, pp. 254-263).

In summary, discourse referents are formal semantic or formal pragmatic devices to capture explicitly and linguistically triggered information required for coreference. Crucially, they are *not* to be confused with inferred or entailed representations, or with the entities themselves semantic representations seek to describe, such as actual persons or objects. It is discourse referents in such a sense that are used by Lascarides & Stone (2009) in order to formalize coherence in gestural semantics.

2.2 DRT

The work on gestural discourse referents by Lascarides & Stone (2006, 2009) is based on the formalism of SDRT (Asher & Lascarides, 2003), which in turn is based on Discourse Representation Theory (DRT; Kamp & Reyle, 1993a, 1993b). In DRT discourse referents are semantic representations of individuals (Kamp & Reyle, 1993a, p. 60), not the individuals themselves, in line with their treatment by Karttunen (1971) and Heim (1982). Discourse referents are not directly evaluated against the model’s predicate extensions but have to be mapped to the model’s universe of individuals first (Kamp & Reyle, 1993a, pp. 94-95, 112). DRT uses syntactic trees as an input to rules to build a Discourse

Representation Structure (henceforth DRS): A DRS K consists of a set of discourse referents U_K and a set of conditions Con_K which are constructed according to syntactic and lexical properties of the sentence. For example, one rule is triggered by an indefinite NP in subject or object position, requires that a discourse referent u be added to U_K , that a predicate provided by the meaning of the syntactic head be applied to u as an additional condition in Con_K , and that grammatical gender specifications for u are included as a condition in Con_K (Kamp & Reyle, 1993a, p. 122). Together with a similar rule for proper names (Kamp & Reyle, 1993a, p. 121) the sentence (7a) can be represented as (7b)¹:

(7) a. Billie has a cat.

b.	x y
	<i>Billie</i> (x) <i>cat</i> (y) x has y

The top line in (7b) containing the discourse referents x and y represent U_K , while the lower part is occupied by the conditions of Con_K .

DRSs can be conditions of other DRSs by recursive definitions, in which case the unembedded DRS is called the *principal DRS* for disambiguation (Kamp & Reyle, 1993a, pp. 110-111). Often if a DRS K is subordinate to another DRS K' discourse referents will be members of the set U_K of the subordinate DRS K without being members of $U_{K'}$. This circumstance is used in order to formulate constraints on anaphoric binding: Depending on a discourse referent's location in some DRS within the principal DRS (Kamp & Reyle, 1993a, pp. 230-231), anaphoric expressions like pronouns can identify the discourse referent as an antecedent, and thus as coreferential (Kamp & Reyle, 1993a, pp.104-106). This coreference is formalized by introducing a second discourse referent for the pronoun and a condition to identify this second discourse referent with the antecedent (Kamp & Reyle, 1993a, pp. 238-239). Consider the following example discourse:

(8) Jones owns Ulysses. It fascinates him.

(Kamp & Reyle, 1993a, p. 60)

After processing of the first sentence *Jones owns Ulysses* the representation looks like in (9):

¹Gender specifications are omitted here.

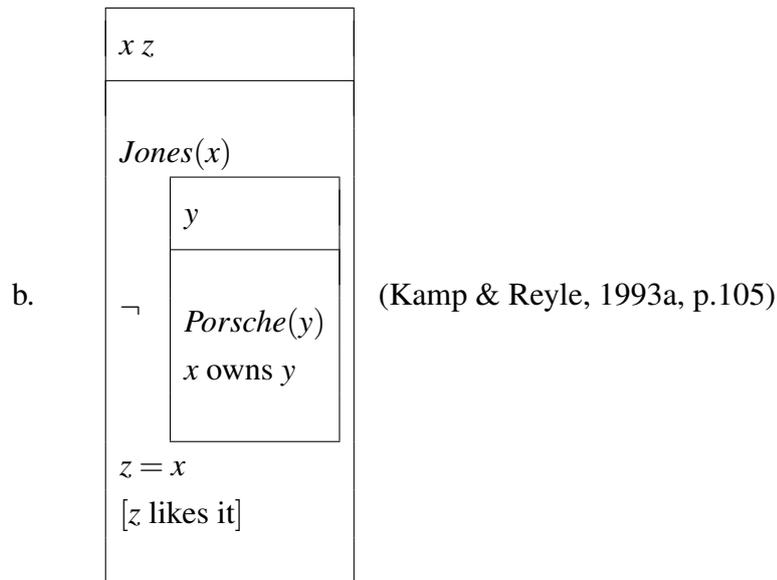
	$x y$	
(9)	<i>Jones(x)</i> <i>Ulysses(y)</i> <i>x owns y</i>	(Kamp & Reyle, 1993, p. 64)

In (9) the verb *owns* relates the discourse referents x and y . The names of their referents are formalized by conditions like *Jones(x)*. The DRS in (9) may be paraphrased by a wording such as *There is an x such that x is Jones and owns y , and there is a y such that y is Ulysses and is owned by x* . Now consider the DRS after processing of the second sentence *It fascinates him*:

	$x y u v$	
(10)	<i>Jones(x)</i> <i>Ulysses(y)</i> <i>x owns y</i> <i>$u = y$</i> <i>$v = x$</i> <i>u fascinates v</i>	(Kamp & Reyle, 1993, p. 69)

The rules for converting syntactic structure with pronouns to DRSs (Kamp & Reyle, 1993a, pp. 104-106) require separate discourse referents (u and v here), which are then identified with the previously established ones by explicit conditions like $u = y$ and $v = x$. It is therefore not the repeated immediate usage of the same discourse referent that ensures coreference, but rather the access statements such as $u = y$ have to the set of discourse referents U_K of some DRS K . This becomes important in cases of negation such as (11a), which is analyzed by the structure in (11b):

- (11) a. Jones does not own a Porsche. He likes it.



The negated VP in (11a) causes a negated DRS to appear as a condition of the principal DRS. The final line $[z \text{ likes it}]$ is meant to represent the fact that *it* cannot be resolved by introduction of some discourse referent and adding an equality condition like it was done for *He* and *Jones*. The negated DRS is said to be *subordinate* to the principal DRS, which is represented by the nesting of the boxes (Kamp & Reyle, 1993a, p. 119). Kamp & Reyle (1993a, pp. 119-120) define *weak subordination* as the transitive and reflexive versions of nesting by negation like in (11b), write $K \leq K'$ for K is weakly subordinate to K' and provide the crucial definition for constraints on coreference as given in (12)²:

- (12) Let K be a DRS, x a discourse referent and γ a DRS-condition. We say that x is *accessible from γ in K* iff there are $K_1 \leq K$ and $K_2 \leq K_1$ such that x belongs to U_{K_1} and γ belongs to Con_{K_2} .

By the definition of accessibility in (12), the discourse referent x in (11b) is accessible from the nested condition $x \text{ owns } y$, but the nested discourse referent y is not accessible from the condition $[z \text{ likes it}]$ in the principal DRS. The construction rule for pronouns however requires that a condition of the form $u = v$ is triggered, where u is a newly introduced discourse referent and v is accessible to the condition $u = v$ (Kamp & Reyle, 1993a, p. 122). Therefore the nesting of DRSs constrains coreference by the notion of accessibility. It is precisely this notion that is used as the fundament for formalizing antecedents to anaphoric conditions in SDRT (Asher & Lascarides, 2003, pp. 44, 149) and therefore by Lascarides & Stone (2009, p. 425).

²Kamp & Reyle (1993a, pp. 154-155) mention another case of subordination, used to represent material implication, and modify the definitions of subordination and accessibility accordingly. They also make use of a disjunction symbol \vee later on. In SDRT however these are replaced by the rhetorical relations *Consequence* and *Alternation* (Asher & Lascarides, 2003, p. 169), which is why only the simplified definition with negation is introduced.

2.3 SDRT

SDRT takes DRSs to be the formal representatives of clauses (Asher & Lascarides, 2003, p. 138). While nesting of DRSs is possible with negation (see an example in Asher & Lascarides, 2003, p. 150), in most cases DRSs will be connected in *Segmented* Discourse Representation Structures or SDRSs. This is done by assigning a label, usually a Greek letter, as a representative to each DRS to be connected inside the SDRS and by specifying a *rhetorical relation* between the labels (Asher & Lascarides, 2003, p. 138). Those relations are characteristic of SDRT and can be illustrated by example (13):

- (13) a. (i) Kim entered the room. (ii) It was pitch black.
b. (i) Kim turned out the light. (ii) It was pitch black.
(Asher & Lascarides, 2003, p. 210)

In both (13a) and (13b) clause (ii) is the same. The way it relates to the respective first clause (i) however is different: The event described in (13b) by clause (i) can be inferred to cause the one of clause (ii), which is not the case for (13a). Here clause (ii) rather seems to provide background information to the circumstances where Kim entered, without postulating that entering the room made it pitch black. In SDRT terms this can be analyzed as different rhetorical relations: Assuming the labels for the clauses to be *i* and *ii* respectively, in (13a) *Background(i,ii)* holds, in (13b) *Result(i,ii)* is the correct relation. Two rhetorical relations, *Elaboration* and *Narration*, are of particular importance to the structure of SDRSs, which is why these will be explained at some length.

The relation *Elaboration*(α, β) is involved in a variety of constraints and default inferences. The main characteristic is inference by a *subtype_D*(β, α) predicate (Asher & Lascarides, 2003, p. 205). Take the following example from Asher & Lascarides (2003, p. 282):

- (14) α Max ate a lovely meal.
 β He devoured lots of salmon.

In β the main eventuality is introduced by the verb *devoured*, in α it is introduced by *ate*. By an appropriate lexical type hierarchy devouring events are of the same type as eating events or a subtype of them, in the sense that devouring is usually considered eating. Furthermore, the patient theta role in β is occupied by (*lots of*) *salmon*, while it is occupied in α by (*a lovely*) *meal*. According to Asher & Lascarides (2003, pp. 282-283), viewing salmon as a subtype of food or meals together with the relation between *devoured* and *ate* allows for inferring a relation *subtype_D*(β, α), which is crucial for *defeasibly* inferring *Elaboration*(α, β) (Asher & Lascarides, 2003, p. 205). Defeasible inference is

a crucial component of the Glue Language (Asher & Lascarides, 2003, pp. 184-211), the logic that is used for choosing the correct rhetorical relation: Certain semantic properties of clauses suggest certain rhetorical relations, but some inferred relations (the ones based on defeasible inferences) can be cancelled by more specific or monotonic ones. For example, $subtype_D(\beta, \alpha)$ monotonically excludes $occasion(\alpha, \beta)$ (Asher & Lascarides, 2003, p. 207), which would be a criterion for inferring $Narration(\alpha, \beta)$ (which will be discussed below), such that evidence for $Elaboration(\alpha, \beta)$ overrides evidence for $Narration(\alpha, \beta)$.

In addition to the $subtype_D$ relation, certain aspectual characteristics $Aspect(\alpha, \beta)$ which are not explicitly specified in their entirety (Asher & Lascarides, 2003, p. 206), are a condition for inferring $Elaboration(\alpha, \beta)$. Note a general complication of the usage of $Elaboration$ in SDRT: It is frequently used with an SDRS as a second argument containing several DRSs linked by $Narrations$ or $Elaborations$, but $subtype_D(\beta, \alpha)$ is only discussed for clausal eventualities (Asher & Lascarides, 2003, pp. 282-283). While defining $subtype_D(\beta, \alpha)$ in such a way that all events in all DRSs inside the SDRS β have to fulfill the subtype requirement is an easy conceivable solution to this, the matter is more complicated if rhetorical relations in β such as for example $Result(\gamma, \delta)$ are to be related to the eventuality of α in terms of fulfilling $subtype_D(\beta, \alpha)$.

Turning to $Narration(\alpha, \beta)$ (Asher & Lascarides, 2003, pp. 162-165, 199-204), this relation represents a sequence of sentences where the event of β follows α in a narration with a common topic. One way to defeasibly infer $Narration(\alpha, \beta)$ is by world or script knowledge captured by $occasion(\alpha, \beta)$ (Asher & Lascarides, 2003, p. 201). This formula is kept vague, but the idea is that evidence that certain events exemplified by β generally tend to follow others of the kind exemplified by α . Asher & Lascarides (2003, p. 201) provide one example among many possible, namely that if α states that some individual x fell and β states that some individual y helped x up, such a frequent course of events $occasion(\alpha, \beta)$ can be inferred. So the two sentences in (15) would be related by $Narration(\alpha, \beta)$:

- (15) α John_i fell.
 β Max helped him_i up.
 (drawn from Asher & Lascarides, 2003, p. 201)

The SPATIOTEMPORAL CONSEQUENCE OF NARRATION (Asher & Lascarides, 2003, p. 163) provides a constraint to ensure the spatiotemporal properties of such sequences of events. Since sequences of events under a certain topic are the phenomenon to be captured by $Narration$, a second explicit way to infer $Narration(\alpha, \beta)$ is introducing the sentence β with the words *and then*, and further such axioms are conceivable and intended by Asher

& Lascarides (2003, p. 202). The inference by *and then* is monotonic and thus cannot be overridden. For example, (16) is related by $Narration(\alpha, \beta)$ as well, even though it narrates no intuitively common sequence of events:

- (16) α Professor Doe went outside the office
 β and then suddenly the lights went out.

The TOPIC CONSTRAINT ON NARRATION (Asher & Lascarides, 2003, p. 164) captures the intuition that a logically contingent common denominator of the narrated events is required in SDRT interpretation. Asher & Lascarides (2003, p. 164) use the notation $\neg \square(K_\alpha \sqcap K_\beta)$, where $K_\alpha \sqcap K_\beta$ denotes a summarizing topic of the sentences α and β . The example they provide for such a summarizing, logically contingent topic is given in (17):

- (17) α My car broke down.
 β Then the sun set
 γ and I knew I was in trouble.
 $K_\alpha \sqcap K_\beta$ I was in trouble.
 (example from Asher & Lascarides, 2003, p.164; formatting modified)

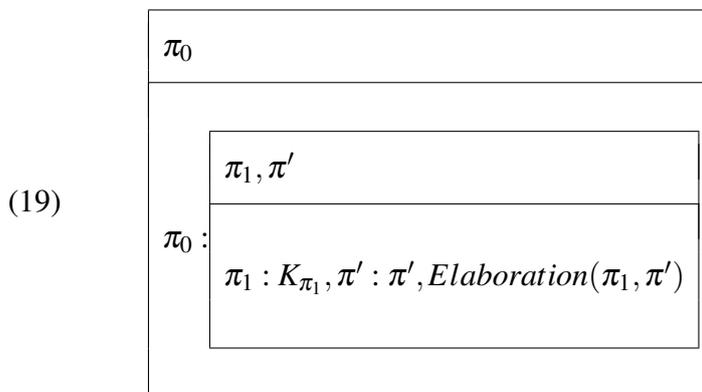
In (17) sentence γ provides evidence for the main point of the sequence of sentences α, β and allows for constructing a common topic $K_\alpha \sqcap K_\beta$ that is not a logical necessity like *Something happened*. A major complication arises from the UPDATE TOPIC CONSTRAINT FOR NARRATION (Asher & Lascarides, 2003, p. 219), which prescribes the influence of the TOPIC CONSTRAINT ON NARRATION on SDRSs: The summarizing topic $\alpha \sqcap \beta$ is to be represented explicitly in the SDRS by some label δ , which in turn stands in the relation $\Downarrow(\delta, \lambda)$, with λ being the label of the SDRS with the $Narration(\alpha, \beta)$ relation. \Downarrow is no independent relation, it is only introduced when discourse update infers other relations, but serves to capture binding data (Asher & Lascarides, 2003, p. 167). $\Downarrow(\delta, \lambda)$ means that δ is the label of a summarizing topic for the $Narration$ relation containing SDRS labeled λ . Since such topic information is capable of changing binding facts, but is not predictable, this poses severe problems for testing binding data in SDRT, which will be a major concern later in this thesis. These complications however can be avoided: $Elaboration(\alpha, \beta)$ entails $\Downarrow(\alpha, \beta)$ (Asher & Lascarides, 2003, p. 227), and $Narration$ relations elaborating on some sentence or discourse structure do not introduce additional structural complexity than the part already evident from $Elaboration$. Any argumentation related to binding in $Narration$ segments therefore needs to specify an explicit clause the $Narration$ related segment elaborates on.

Now consider example (18) for an illustration of SDRS notation and structure for

more complex discourses discussed by Asher & Lascarides (2003, pp. 224-229):

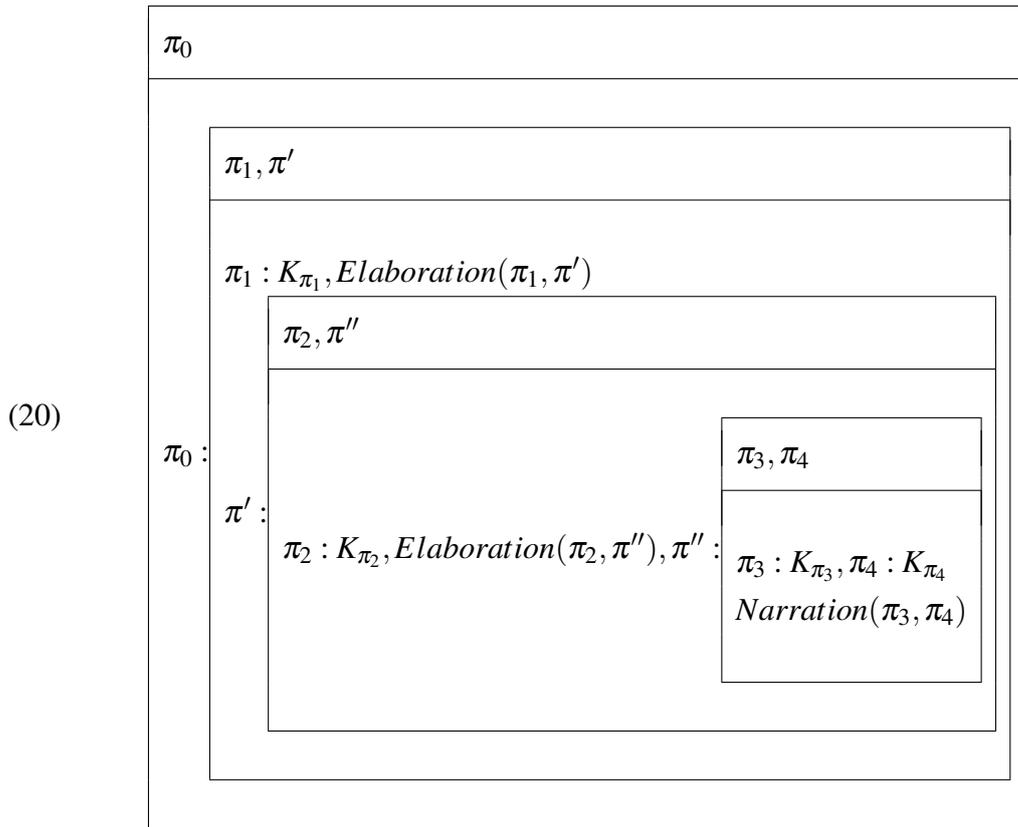
- (18) a. π_1 Max experienced a lovely evening last night.
 b. π_2 He had a fantastic meal.
 c. π_3 He ate salmon.
 d. π_4 He devoured lots of cheese.
 e. π_5 He won a dancing competition.
 (Asher & Lascarides, 2003, p. 8)

The idea behind (18) is that clause (a) is elaborated on by all the information in (b-e). Representing the information of (a) by the DRS abbreviation K_{π_1} and (b-e) by a label π' , the structure can be illustrated as in (19):

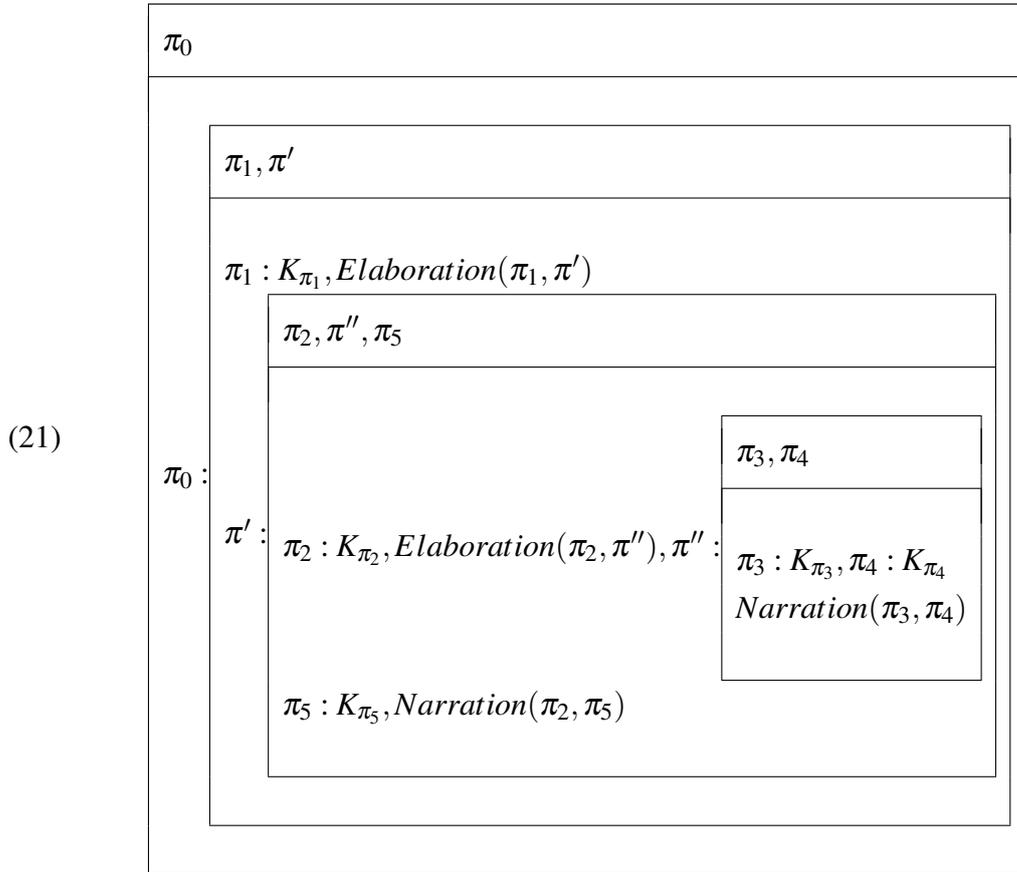


Each (S)DRS is given with its label followed by a colon followed by the (S)DRS itself. The upper part of an SDRS lists all (S)DRS labels immediately subordinate to it, for example π_1 and π' are immediately subordinate to the root SDRS π_0 . The only exception is the outermost box, which is drawn around the root SDRS π_0 and which lists π_0 in its upper part. At this point π' is not specified further in the example, which is why the uninformative notation $\pi' : \pi'$ is employed in (19).

Clauses (c) and (d) of (18) specify further the meal mentioned in (b), but among (c) and (d), none specifies the other in more detail - (c) describes an earlier part of the meal, (d) a later one. This can be integrated into the structure of (19) to yield (20):



In (20) again K_{π_2} represents the DRS of clause (18b), and (c) and (d) are represented the same way. The information contributed by (b-d) is labeled π'' and consists of π_3, π_4 and the *Narration* relation between them indicating their temporal succession. So far π'' is taken to be all that elaborates on π_1 , but clause (e) is not integrated yet. Since it temporally follows all meal-related information in (18) and elaborates on (a) as well, it is integrated on the same level as the meal segment π'' inside π' :



Note that while all relations are defined to hold between exactly two (S)DRSs, multiple relations and more than two (S)DRSs can occur as part of the same superordinate SDRS, just like π' in (21) contains three (S)DRSs and two relations.

2.4 The application of SDRT by Lascarides & Stone (2006, 2009)

Lascarides & Stone (2006, 2009) proposed an analysis of co-speech gesture where gestures are embedded into the logical forms of SDRT in a similar way as clauses. Gestures introduce discourse referents and are linked to other discourse units by rhetorical relations. To this end Lascarides & Stone (2009, p. 424) allow existing relations to connect gestures to other (S)DRSs, but also introduce three new discourse relations: *Overlay*, *Depiction* and *Replication*. While *Overlay* is beyond the focus of this thesis, the other two are presented below.

Lascarides & Stone (2009, p. 405) introduce *Depiction* to account for the intuition that gesture can provide more or less the same information as concurrent speech without being redundant. They hold that “*Depiction*(π_1, π_2) holds only if π_1 labels the content of a *spoken* unit, π_2 labels the content of a *gesture*, and K_{π_1} and K_{π_2} are non-monotonically equivalent” and “omit formal details because it requires a modal model theory for \mathcal{L}_{sdrs} as described in Asher & Lascarides (2003)” (Lascarides & Stone, 2009, p. 424). The modal model theory of SDRS semantics builds on the one for DRS semantics (Asher

& Lascarides, 2003, p. 156). Using modal models for indicative clauses is motivated by the usual necessity and possibility operators, but also a special conditional $>$ whose interpretation relies on normality (Asher & Lascarides, 2003, pp. 47-48), a barely defined notion. Asher & Lascarides (2003, p. 47) use as an example the sentence *Tasha is a cat* to illustrate the normality function it is based on: It provides those world assignment pairs where Tasha fulfills typical properties of cats such as having four legs or meowing. Since Lascarides & Stone (2009) mention non-monotonic equivalence and the modal theory, one can arguably conclude that it is normality or defeasibility of inferences they allude to, which renders more concrete definitions hardly possible at this point. Therefore recourse to their most explicit example is necessary in order to get an intuitive understanding of the intended phenomenon. The example (1) is repeated as (22) with labeling and with its semantics provided in (23):

- (22) a. π_1 : and took his hatchet and
 π_2 : Speaker's right hand grasps left hand, with wrists bent.
 π_3 : Speaker lifts poised hands above right shoulder.
 b. π_4 : with a mighty sweep
 π_5 : Hands, held above right shoulder, move back then forward slightly.
 c. π_6 : SLICED the wolf's stomach open
 π_7 : Arms swing in time with sliced; then are held horizontally at the left

(cited as in Lascarides & Stone, 2009, p. 425;
 originally from Kendon, 2004, p. 136)

- (23) π_1 : $\exists hw[took(e_1, w, h) \wedge hatchet(h)]$
 π_2 : $[\mathcal{G}](\exists lra[left_hand(l, w) \wedge right_hand(r, w) \wedge handle(h, a) \wedge grab(e_2, w, a) \wedge instrument(e_2, l) \wedge instrument(e_2, r)])$
 π_3 : $[\mathcal{G}](\exists d[right_shoulder(d, w) \wedge lift(e_3, w, h) \wedge goal_location(e_3, d) \wedge instrument(e_3, l) \wedge instrument(e_3, r)])$
 π_4 : $\exists s[sweep(s) \wedge mighty(s) \wedge with(e_4, s)]$
 π_5 : $[\mathcal{G}](coiled_backswing(e_5, s) \wedge instrument(e_5, l) \wedge instrument(e_5, r))$
 π_6 : $\exists ft[slice - open(e_4, w, t) \wedge stomach(t, f) \wedge wolf(f)]$
 π_7 : $[\mathcal{G}](slice - open(e_4, w, t))$
 π_0 : $Elaboration(\pi_1, \pi) \wedge Narration(\pi_1, \pi_5) \wedge Explanation(\pi_4, \pi_5) \wedge Replication(\pi, \pi_5) \wedge Narration(\pi, \pi_5) \wedge Background(\pi_4, \pi_6) \wedge Depiction(\pi_6, \pi_7) \wedge Replication(\pi_5, \pi_7) \wedge Narration(\pi_5, \pi_7)$
 π : $Replication(\pi_2, \pi_3)$
 (Lascarides & Stone, 2009, pp. 425-426)

Here w represents a previously mentioned woodsman. While different parts of this

extensive example will be discussed one after the other, for *Depiction* only some elements of (23) are relevant. The specification of π_0 provides exactly one *Depiction* relation, the one between π_6 and π_7 . The $[\mathcal{G}]$ symbol in the semantics of π_7 (and all gesture formulae in general) restricts discourse referents introduced in gesture to gesture modality so that they cannot serve as antecedents for speech (Lascarides & Stone, 2003, p. 422). Apart from this operator common to all gesture interpretations π_7 is presented as providing part of the information π_6 does. Note that separate DRSs introduce separate discourse referents and equate them even if coreferent, so $[\mathcal{G}](slice - open(e_4, w, t))$ is shorthand for a formula where some discourse referents for w^3 and t are introduced and connected by equality.

The formula of π_7 itself is not equivalent to π_6 even when abstracting from technicalities like the $[\mathcal{G}]$ operator and equality conditions: It provides no information about a wolf or a stomach, which is a plausible formal rendition of π_7 as described in (22). Taking up a comment on *Depiction* by Lascarides & Stone (2009, p. 405), it seems that any iconic gesture which in itself bears some information but does not contribute anything informative to the concurrent speech can be considered *Depiction*. But note that this analysis is by no means the only correct one of (22c), since trajectory and position of the axe are in fact not entailed by π_6 . Compare this to the *Elaboration* of π_1 by the π (the *Replication*(π_2, π_3) sequence): Lascarides & Stone (2009, p. 426) note that they provide the analysis in (23) because π_2 and π_3 demonstrate the way the woodsman took his hatchet. Therefore the diagnosis of *Depiction* is no formal property, but a rule of thumb, in that gesture semantics and speech semantics overlap and that the additional information in gesture is probably not considered relevant. If it is considered relevant, the relation is *Elaboration*, and whether it is judged as such or not is highly ambiguous.

Note informally the general pattern in (23) illustrating which kind of representations Lascarides & Stone (2009) have in mind: imaginary objects and body parts are ordinary discourse referents, their movements are provided as events, and next to predicates stemming directly from what the gesture looks like, multiple connections in terms of thematic roles are given (e.g. *instrument* in π_2, π_3, π_5 and *goal_location* in π_3).

The second common yet hardly defined rhetorical relation, *Replication*, connects two gesture units and indicates entities evoked by the first gestures being re-used in the second one in the same way, or that semantically similar entities are used in the second gesture (Lascarides & Stone, 2009, p. 424). It connects π_2 and π_3 in (23) since the discourse referents l and r for the woodsman's left and right hand continue to be referred

³To be exact, for the last discourse referent that was introduced in π_6 that amounts to the woodsman w by a chain of new discourse referents and equality conditions.

to in the gestures. The same holds for π and π_5 . Lascarides & Stone (2009, p. 426) consider even π_7 to refer back to l and r although they do not appear in the formula of π_7 in (23), since the gesture of π_7 depicts them just the same.

Semantic similarity is not formalized as a diagnostic, but the comments of Lascarides & Stone (2009, pp. 406, 424) suggest that it may not be only about similarity in interpretation of two gestures, but rather as well in form. The treatment of gesture form-meaning mapping by Lascarides & Stone (2006, 2009) will play a crucial role in this thesis.

One property of their analysis is the decomposition of gestures into feature structures. The example in (24) is analyzed as having the feature structure given in (25).

- (24) There are these very low level phonological errors that tend not to get reported.
The right hand is held in a fist and positioned below the mouth, where the previous gesture was performed; the hand iteratively moves in the sagittal plane (i.e., vertically outwards) in clockwise circles (as viewed from left).
 (Lascarides & Stone, 2006, p. 64)

- (25)
- | | | |
|---|--|---|
| [| hand-shape : <i>asl-s</i> |] |
| [| finger-direction : <i>down</i> |] |
| [| palm-direction : <i>left</i> |] |
| [| trajectory : <i>sagittal-circle</i> |] |
| [| movement-direction : { <i>iterative, clockwise</i> } |] |
| [| location : <i>central-right</i> |] |

(Lascarides & Stone, 2006, p. 66)⁴

The values to each attribute are first assigned by gesture form only, e.g. the value *asl-s*, a fist used in ASL to sign the letter *S*, for *hand-shape* can be read off the gesture's fist shape directly. These values however are part of a type hierarchy containing much more specific predicates, to which they can be resolved by contextual information. Lascarides & Stone (2006, p. 67; 2009, pp. 434-435) propose that the predicate derived from the value *asl-s* (or *asl-a*, see footnote 4), *hand_shape_asl – s(i)*, can be resolved to the more specific *something_held(x)* or *event_of_holding(e)*. This illustrates the fact that the type of discourse referent introduced by a gestural feature is underspecified according to Lascarides & Stone (2006, 2009) - the variable i of *hand_shape_asl – s(i)* may

⁴The same example is given a slightly modified attribute-value matrix in Lascarides & Stone (2009, p. 430), where the feature names bear the prefix *right-* in order to indicate properties of the right hand only, and the value *asl-s* (the letter *s* in the ASL alphabet, i.e. a fist) is given as *asl-a* (which has a different position of the thumb).

resolve to an event e or an entity x . The *event_of_holding*(e) predicate may be resolved further to a literal or a metaphorical interpretation, where *metaphorical_holding*(e) may ultimately resolve to *sustains*(e), which is an interpretation Lascarides & Stone (2009, p. 442) consider plausible for example (24). The type hierarchy of all subtypes of all feature values is not a tree, one subtype may be derived from various feature value supertypes. Note that by adopting the underspecification language Robust Minimal Recursion Semantics (RMRS; Copestake, 2006, 2007) even the number of arguments to a predicate is underspecified (Lascarides & Stone, 2006, p. 66).

Note also that Lascarides & Stone (2006, p. 67; 2009, p. 435) allow feature values like those in (25) to be resolved to mere logical validity \top in order to account for values not meant to contribute to meaning. The general contextual resolution algorithm of gesture feature values is not spelled out by them, and resolution to more concrete meaning will be partially covered by this thesis. Some of the six features as specified by Lascarides & Stone (2006, 2009) have a different status than others in contextual resolution. While *hand – shape* crucially helps to determine the kind of object referred to, *palm – direction* is likely too general to determine the kind of object irrespectively of the values of other features such as *hand – shape*. It will therefore be assumed that either the features interact for contextual resolution or that the six feature system has to be revised in terms of linguistic significance. The correct way to capture this is however beyond the scope of this thesis.

An adequate type hierarchy has to be large enough to cover all the meanings iconic gestures can convey. Lascarides & Stone (2006, 2009) only outline a general approach and provide an example. A more extensive hierarchy might be obtained by an AI analyzing gestural corpora or by taking into account empirical studies about gestures (among these McNeill, 1992; Müller, 1998; Kendon, 2004). For example Müller (1998, pp. 115-123) establishes four ways of mapping between a gesture and its meaning: The gesture can perform a movement which is similar to the movement of the event, it can portray objects as being manipulated or held by the hands, it can outline the contours of an object or the hands are meant to represent the object itself. If these four mappings are an exhaustive list or if they are expanded to such a list they can provide additional structure to the type hierarchy.

The remainder of this thesis will repeatedly refer to this hierarchy. There will be a distinction between unresolved gestures and their resolved meaning. For example (24) *asl – s* is such a feature value only obtained by looking at the gesture, not at the message it conveys, which makes *asl – s* unresolved. Since the linguistic significance of the exact six specific features of Lascarides & Stone (2006, 2009) is taken into doubt, all feature

values will be treated as one predicate which may resolve to a predicate about one or more *entities* (type *e* objects excluding events), a predicate about an *event* (possibly including thematic role predicates) or both. It will tacitly be assumed that this is achieved by a conjunction of all feature values while remaining agnostic which features are to be encoded in the matrix, and with the requirement that they influence the resolution of each other. The terms *gesture form*, *form* or *form predicate* will refer to the unresolved form of the joint predicate of all feature values. In contrast the term *subtype (predicate)* will refer to the contextually resolved predicates about entities or events. For example, if the unresolved feature values of the gesture (25) are called *rotate – fist*, then *rotate – fist* is the form predicate, while the contextually resolved meaning *sustain(e)* from the discussion above is the subtype.

Returning to the contextual resolution mechanism, the idea behind the algorithm assumed in this thesis can be illustrated with example (26) below. Most examples in this thesis will be provided in colloquial German (with possible regional influences) in order to provide natural native speaker judgments:

- (26) Ich hab die [Tür aufgemacht] _{turn-with-thumb} und sie reingelassen.
 I have the door opened and her let_in
 ‘I opened the door and let her in.’

In (26) the form predicate *turn – with – thumb* indicates the hand shape and orientation usually used when turning a key, in a location in front of the speaker like in a usual distance for unlocking something, with two turning movements to the left. With the one-sentence discourse in (26) there are very few salient entities: the speaker, whoever *sie* (‘her’) refers to, and the door. A possible argumentation about how to obtain a specified semantic reading of *turn – with – thumb* is the following: The gesture alignment indicates that the movement takes place before the referent of *sie* is inside. Interaction with the person in question by the indicated movement is therefore implausible, since a door separates the speaker and that person. There is no salient or conventional movement the speaker would perform with themselves or an unmentioned object.

The door as a whole is subject to the same reasoning, but one of very few well-known actions performed with a door, unlocking it, is compatible with hand shape and orientation, position, and movement features. Other actions involve turning a door knob or a door handle, but a door knob is not turned twice⁵ and door handles in addition are rather used with a tilt rather than by turning it around the forearm axis. Since unlocking is

⁵Introspectively judging, the visually most accurate hand shape would be one that is considerably more open, but fist shapes seem to fulfill the requirements to similarity sufficiently as well. Since the special orientation of thumb and index finger are easily overlooked, the frequency of turning movements is most likely to be responsible for resolving the potential ambiguity.

the only salient action for all entities in the discourse, and resolving the feature values to predicates involves arguments of underspecified number and type, an adequate semantic representation of the gesture will either contain a discourse referent entity for the key, a discourse referent event for the unlocking, or both.

The underspecification language RMRS allows for underspecified predicates, scope, arity and argument type (Copestake, 2006, pp. 5-6; Lascarides & Stone, 2009, pp. 432-433). Each feature can provide a separate predicate, but they can also resolve to validity \top , and there is no explicit pragmatic constraint against a conjunction of equivalent predicate-argument formulae. One can for example assume a representation where one feature value (the value of *hand-shape*) resolves to the key, one (the one of *movement-direction*) resolves to unlocking and the rest resolves to \top . If agent and patient are expressed by the event predicate directly and the instrument role is given separately (as in π_2 in (23)) resolving all underspecification results in the formula $[\mathcal{G}](\exists xyz[e[*speaker*](x) \wedge y = *door* \wedge *key*(z) \wedge *unlock*(e,x,y) \wedge *instrument*(e,z)])$. In this notation *door* represents the discourse referent of the door introduced in speech. The whole formula is outscoped by the gesture operator $[\mathcal{G}]$. Due to the argument structure of (26) the relationship between unlocking and opening something and the informative component about a key, the gesture is most likely⁶ related to the speech of (26) by *Elaboration*.

The proposal by Lascarides & Stone (2009, p. 411) requires that “entities introduced in gesture must be bridging related to entities introduced explicitly in accompanying speech”. By their reference to the article of Clark (1977) one may take this to mean that said entities do not need to be coreferent with those in accompanying speech, but there has to be one of possibly many relations between the two. This serves to establish coherence between speech and gesture. Such relations frequently license the usage of definite descriptions without coreference. One example for such a licensing relation called *probable parts* by Clark (1977, p. 416) is provided in (27). There *the room* licenses the usage of previously unmentioned windows as *the windows* because rooms are likely to have windows:

(27) I walked into the room. The windows looked out to the bay. (Clark, 1977, p. 416)

In example (26) such a process arguably applies to the key associated with the door mentioned. Note also that bridging relations, albeit systematic in relations such as *part-whole* or *probable parts*, can be creatively licensed by contextual knowledge of

⁶Note the discussion about the ambiguity between *Depiction* and *Elaboration* above.

particular interlocutors rather than being strictly encoded by conventional bridging sequences. For example Hawkins (1978, p. 100) claims that among linguists mentioning *transformational grammar* can license the definite description *the deep structure*. A full example presenting both Hawkins's claim and a gestural application is given in (28):

- (28) a. (i) Die Konstruktion ist ein echtes Problem für eine
 this construction is a real problem for a
 Transformationsgrammatik.
 transformational_ grammar
 'This construction is a real problem for a transformational grammar.'
- (ii) Die deep structure würde überhaupt keinen Sinn machen.
 the deep structure would at all no sense make
 'The deep structure wouldn't make any sense at all.'
- b. Hast du-s mal mit [QR]_{fist-over-fist} probiert?
 have you-it once with QR tried
 'Did you try QR?'

In (28b) the gesture *fist – over – fist* represents a gesture where both hands are in fist shape. The left fist remains in a fixed position while the right fist starts below the left one, moving in a semicircle around and above the left fist. With the concurrent mention of Quantifier Raising (May, 1977, 1985) the whole gesture can be regarded as representing the movement operation where one node moves above all others, which in English would be a quantifier phrase above the topmost IP. The quantifier and IP represented by the fists can be regarded as an instance involving Clark's (1977) *part-whole* bridging.

Lascarides & Stone (2006, p. 68) make further claims about the bridging mechanism of gestures: One way to fulfill the need for a bridging relation is equality and bridging relations are not obtained by the syntax-semantics interface. Since they assume SDRT as a framework, which in turn is based on Discourse Representation Theory (DRT; Kamp & Reyle 1993a, 1993b), introducing a new variable for the same individual cannot be avoided in the case of coreference. Therefore if a bridging constraint like the one mentioned above is adopted, coreferent gestural variables merely depicting the simultaneous speech referent are subject to it. Satisfaction of such a bridging constraint by equality conditions is therefore an important component of the analysis.

Having established the background, the framework of Lascarides & Stone (2009) can now be put to use to investigate a pronominal approach to coreference with gestures.

3 Gestural discourse referents as pronouns

While they do not explicitly make this claim, comments by Lascarides & Stone (2009, pp. 414, 417, 426) seem to suggest that anaphoric dependencies between gestural discourse referents and antecedents are constrained by discourse structure to some degree. The most explicit constraints on anaphoricity by discourse structure are those for pronouns. As a first point of investigation this thesis is therefore concerned with the question of if there is a connection between constraints on pronoun resolution and gesture. The question is specified in three different versions, given in decreasing strength below:

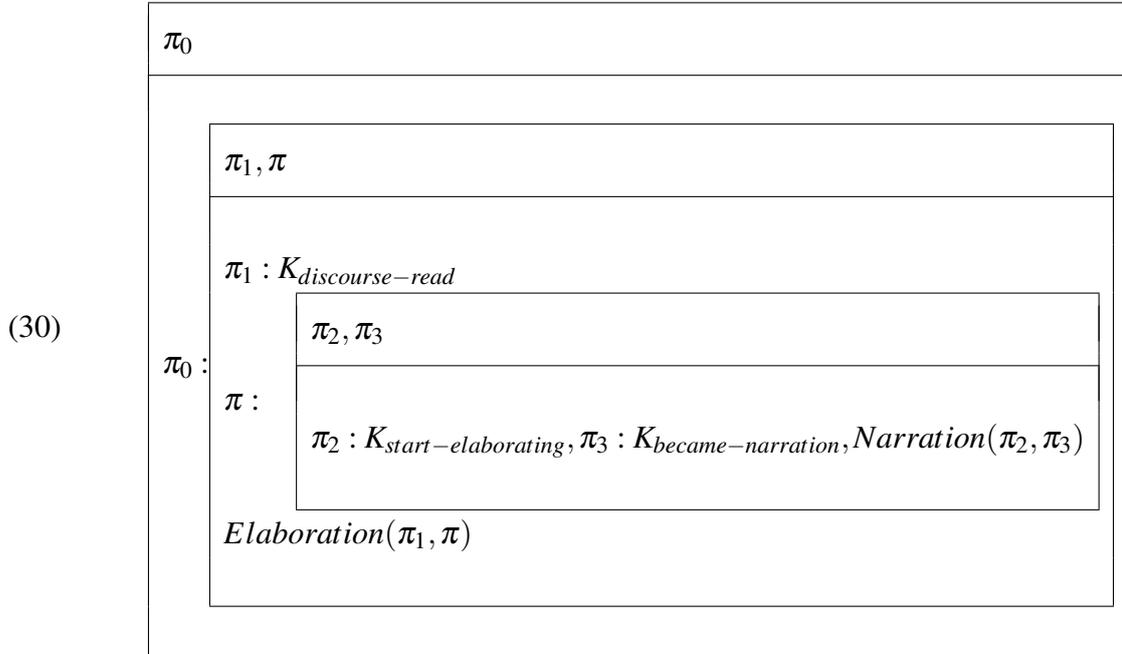
1. Can the discourse referents evoked in iconic co-speech gestures be treated the same way as pronouns are treated in SDRT?
2. Can the discourse referents evoked in iconic co-speech gestures be treated in a similar way as pronouns are treated in SDRT?
3. Can the discourse referents evoked in iconic co-speech gestures be treated the same way as pronouns?

These three degrees of the pronoun hypothesis are discussed simultaneously in the present section. It will be demonstrated that none of these versions can be upheld. The methodology used to achieve this is as follows: Structurally rich examples are provided alongside their SDRT representation. When these representations make wrong predictions about the available antecedents (such that version 1 has to be rejected) even version 2 will be rejected since the examples are chosen to violate the constraints to a considerable degree. Finally the weakest formulation, that the gestural referents behave like pronouns even though SDRT cannot capture pronoun behavior, is disproven by minimal pairs with speech pronouns.

The constraint on available antecedents of SDRT relies on the notion of *outscoopes*. Asher & Lascarides (2003, p. 138) define that “a label π *immediately outscoopes* a label π' if $\mathcal{F}(\pi)$ includes as a conjunct a formula of the form $R(\pi', \pi'')$ or $R(\pi'', \pi')$ for some discourse relation R ”, where \mathcal{F} is a function from SDRS labels to the well-formed SDRS formulae they represent. Then *outscoopes* is defined as the transitive closure of *immediately outscoopes* (Asher & Lascarides, 2003, p. 138). In the DRT style box notation of SDRSs an (S)DRS immediately outscooped by another SDRS will be nested inside that superordinate SDRS (Asher & Lascarides, 2003, p. 140). Consider the following sequence:

- (29) π_1 A discourse was read.
 π_2 This started in an elaborating way,
 π_3 and then it became a little narration.

Assuming a plausible axiom schema in the incomplete Glue Language of Asher & Lascarides (2003) where demonstratives such as *this* in π_2 resolved to a whole propositional antecedent license *Elaboration*⁷ and by means of the *and then* axiom schema already mentioned above we can infer the following discourse structure, where the DRSs have keywords of the sentences as subscripts so that they are more transparent:



In (30) π immediately outscopes π_2 and π_3 since there is a rhetorical relation in π (i.e. $Narration(\pi_2, \pi_3)$) which is a conjunct⁸ in π and takes π_2 and π_3 as arguments. The root SDRS π_0 immediately outscopes π_1 and π by the same definition via $Elaboration(\pi_1, \pi)$. By the transitivity of *outscopes*, all *immediately outscopes* relationships are *outscopes* relationships and in addition the root SDRS π_0 *outscopes* (but not *immediately outscopes*) π_2 and π_3 . With the definition of *outscopes* at hand now consider the constraint on available antecedents to anaphora itself:

(31) **Antecedents to Anaphora** (ignoring structural relations⁹)

Suppose that β labels a DRS K_β that contains an anaphoric condition φ . Then the *available antecedents* to the anaphoric condition are the discourse referents that are:

1. in K_β and DRS-accessible to φ [...]
2. in K_α , DRS-accessible to any condition in K_α , and there is a condition

⁷The same holds for *it* in π_3 . At several points Asher & Lascarides (2003) note that the Glue Language can be enriched by further axiom schemas to enhance detection of rhetorical relations, so for their diagnosis a rather informal understanding of the main intuition is best suited, although ideally each application should be explained.

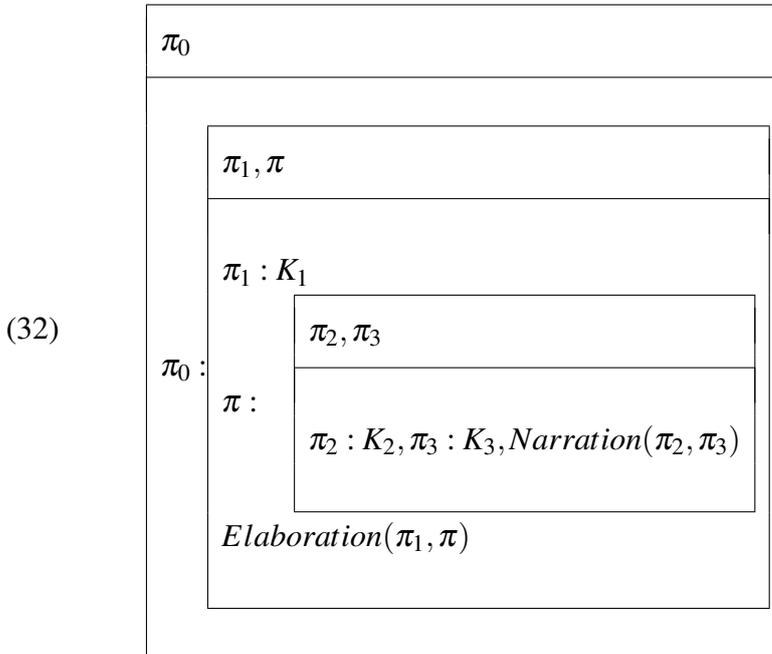
⁸In an SDRS multiple contained (S)DRSs and relations are treated as conjoined formulae (Asher & Lascarides, 2003, pp. 138-139, 156).

⁹Structural relations are *Contrast* and *Parallel* (Asher & Lascarides, 2003, p. 148).

$R(\alpha, \gamma)$ in the SDRS such that $\gamma = \beta$ or *outscopes*(γ, β) (where R isn't a structural relation).

(taken from Asher & Lascarides, 2003, p. 149)

The definition references DRS-accessibility which is handled very similarly as the definition in (12) above: Immediate subordination is given when some DRS K_2 contains another DRS K_1 or its negation $\neg K_1$ as a condition, and all discourse referents in the DRSs which are superordinate by transitive closure of immediate subordination are DRS-accessible (Asher & Lascarides, 2003, p. 44)¹⁰. Therefore DRS-accessibility boils down to the idea that discourse referents under the scope of negation are excluded by (31). Point 1 in (31) only considers DRS-internal antecedents, so for discourse structure point 2 is the relevant one. It states that whenever the DRS is the second argument (γ) of a rhetorical relation $R(\alpha, \gamma)$ or nested inside an SDRS γ that is the second argument of R , non-negated discourse referents from the first argument α DRS of R are available antecedents. Consider the following SDRS for illustration:

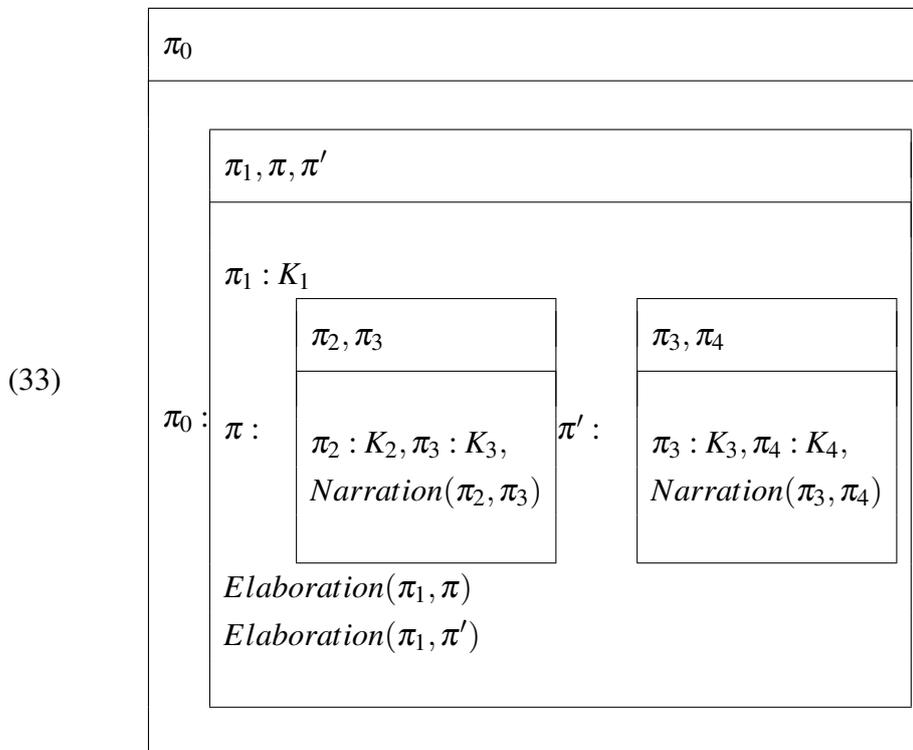


In a structure like (32) there is a π_0 that immediately outscopes a DRS π_1 and an SDRS π , which in turn immediately outscopes π_2 and π_3 . The *Narration*(π_2, π_3) licenses pronouns in π_3 to corefer with antecedents in π_2 according to (31) if we assume that $\gamma = \beta$ from the definition are equal to π_3 . π_2 has no access to discourse referents from π_3 in *Narration*(π_2, π_3) since the order of arguments is important. However both π_2 and π_3 can be considered the outscoped β of the definition, where the outscoping SDRS π from (32) is γ in definition (31). Then *Elaboration*(π_1, π) is the $R(\alpha, \gamma)$ from the definition,

¹⁰They also include a clause for quantifiers, which are beyond the scope of this thesis, and for conditionals, which are replaced by *Consequence* as previously mentioned.

which is why pronouns in both π_2 and π_3 have access to antecedents in π_1 .

By definition (31) and its equivalent in Lascarides & Stone (2009, p. 425) the structure in (33) does not allow π_4 to have access to other discourse referents other than those in π_2 and π_3 , since π is not a DRS but an SDRS:



Note however that Lascarides & Stone (2009, p. 426; see (23) above) license coreference of gestural discourse referents by a *Replication* relation from an SDRS to a DRS, so a modified definition where in (33) the DRS-accessible discourse referents in π_2 are considered DRS-accessible to all conditions in π (for all conditions inside π have access to π_2) may be considered an alternative. Generally speaking an agenda for testing the pronoun hypothesis results from this: constructing a complex SDRS where the antecedent to a gesture is deeply embedded in the left argument α of some relation $R(\alpha, \gamma)$ in the sense of definition (31), and not more easily available by some other R .

3.1 Speech-only antecedents

In order to force deeply hierarchical structures in SDRT which generally prefers interconnected but flat structures (Asher & Lascarides, 2003, p. 234), it is useful to resort to *Narration* sequence elaborating on a clause, where one clause of the *Narration* sequence in turn is elaborated on and so forth. The discourse in (34) is an example for this. It may seem excessive, but this is useful to investigate whether something close to

the definition (31) constrains antecedents to iconic co-speech gestures.¹¹

(34) (Context: The speaker is telling a detailed story about what they did today.)

- π_1 Ich hab heute gekocht.
I have today cooked
'Today I cooked.'
- π_2 Erst hab ich alle Zutaten und Utensilien bereitgelegt.
first have I all ingredients and utensils ready_put
'First I put all ingredients and utensils to the right places.'
- π_3 Dazu hab ich erst Gemüse geschnippelt
for_that have I first vegetables cut
'For that I first cut vegetables'
- π_4 und dann hab ich den weniger aufwändigen Kram bereitgestellt.
and then have I the less time_consuming stuff ready_put
'and then I prepared the less time-consuming stuff.'
- π_5 Also ich hab unter anderem Dosenmais abgeseibt,
well I have among others canned_maize sifted
'That is, among others I sifted canned maize,'
- π_6 dann hab ich Gewürze und ne Dose Tomatensoße aus dem Schrank
then have I spices and a can tomato_soup out_of the cupboard
geholt,
taken
'then I took spices and a can of tomato soup out of the cupboard,'
- π_7 und dann hab ich gemerkt, dass mein Mitbewohner unsere
and then have I realized that my flatmate our
Rührlöffel total versifft zurückgelassen hat
mixing_spoons totally dirty left_behind has
'and then I realized that my flatmate had left behind our mixing spoons
completely dirty'
- π_8 und hab deswegen halt stattdessen nen Teelöffel zum Rühren
and have because_of_that PRT instead a teaspoon for stirring
neben den Topf gelegt.
next_to the pot put
'so because of that I just placed a teaspoon for stirring next to the stove.'
- π_9 Dann hab ich den Herd angemacht,
then have I the stove turned_on
'Then I turned on the stove,'
- π_{10} und dann hab ich alle Zutaten nach und nach zusammengemischt.
and then have I all ingredients after and after mixed_together
'and then I combined all the ingredients one after the other.'
- π_{11} Dazu hab ich erst den Reis in der Tomatensoße gekocht,
to_that have I first the rice in the tomato_soup boiled
'That is I first boiled the rice in the tomato soup,'

¹¹In π_8 of this monologue and in later glossed examples PRT as a non-standard gloss represents the various German discourse particles that do not have a contextually accurate translation in English.

- π_{12} dann hab ich das Gemüse zugegeben,
then have I the vegetables added
'then I added the vegetables,'
- π_{13} danach hab ich-s gewürzt
afterwards have I-it seasoned
'then I seasoned it'
- π_{14} und am Ende hab ich-s noch mal [gut umgerührt] π_{15}
and at_the end have I-it yet again well stirred
'and in the end I stirred it well once more'
- π_{15} *two – fingers – pinch – stirring*
- π_{16} und dann war-s fertig.
and then was-it done
'and then it was done.'

First some explanations to (34) are due. The gesture form of π_{14} , called *two – fingers – pinch – stirring* here, has the hand shape of thumb and index finger extended and holding or pinching a light object downwards. The other three fingers are extended horizontally in a demonstrative way, suggesting that the object in question is light enough to be handled by two fingertips. The hand traces two or three circles such that thumb and index finger tip rotate around a vertical axis and is supposed to indicate the stirring movement by that pinched light object, the teaspoon in question.

Such a usage of gesture seems to be more acceptable if the speaker occasionally includes some other gestures which are omitted from (34). In this regard (34) should be considered simplified, but it seems that the gestures in question are quite exchangeable. Such supporting gestures would include e.g. pointing left with the hands in π_3 and right in π_4 as to indicate the slight contrast of the clauses. Counting gestures for the three steps described in π_5 , π_6 and π_8 are useful as well: Signing number one (putting the right index finger on the left thumb) in π_5 , signing the number two (extending thumb and index finger of the left hand, putting the right index finger on the left one) in π_6 and holding it until mentioning the teaspoon in π_8 , upon which number three (middle finger of the left hand extended as well, right index finger on the left middle finger) is signed improves acceptability. This ensures that π_{15} is not completely unexpected by virtue of being a gesture, but adding little to no information to (34). π_{11} and π_{12} can be accompanied by *Depiction* style gestures like a swiping hand movement and a shaking movement of an imaginary held object, respectively. Note that the acceptability of π_{15} is not due to being easily overlooked: If the speaker signs it demonstratively and adds a sarcastic smile, the gesture is highlighted and can be considered equally or even more coherent.

The point of (34) is its complex structure combined with the fact that π_{15} can be resolved to involve a teaspoon, or more specifically, the teaspoon mentioned in π_8 . But

according to Lascarides & Stone (2006, 2009) such gestures can be resolved to event predicates or individual predicates, so it is necessary to ask if it is not only something like $\exists ey[stir(e, speaker, y) \wedge y = food]$ which π_{15} denotes (where *food* represents some discourse referent for the food mentioned in the discourse). But the teaspoon is relevant to the felicity of π_{15} - consider replacing π_8 in (34) by (35):

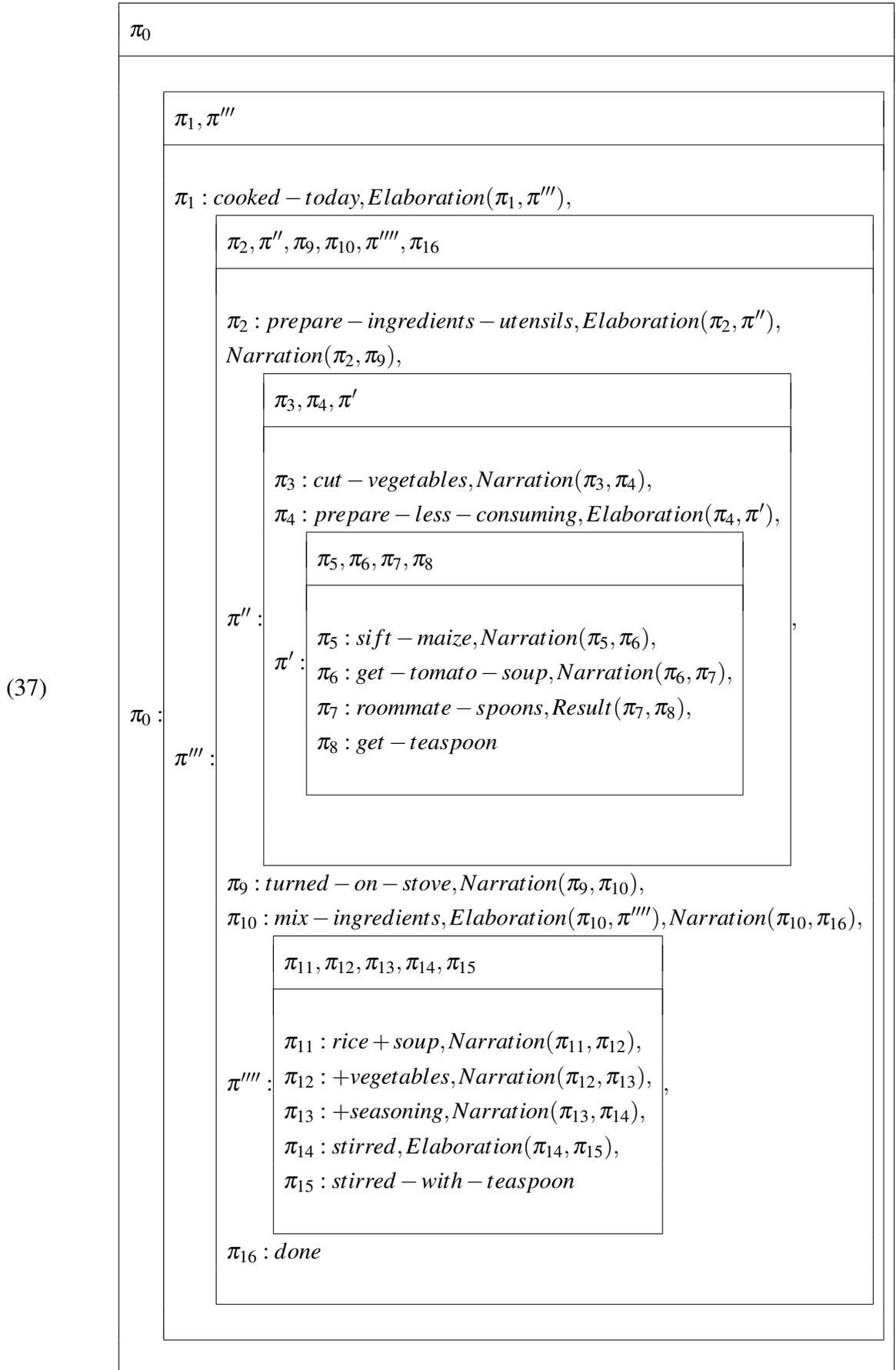
- (35) π_8 und hab deswegen halt stattdessen ne vollkommen
 and have because_of_that PRT instead a completely
 überdimensionierte Kelle zum Rühren neben den Topf gelegt.
 oversized ladle for stirring next_to the pot put
 ‘so because of that I just placed a completely oversized ladle for stirring
 next to the stove.’

If the narrator solves the dilemma of not having an adequately sized mixing spoon by taking a bigger one instead of a tiny teaspoon, the gesture π_{15} cannot be resolved to an antecedent and lacks context. It can therefore be assumed that π_{15} involves a representation with a subformula such as *teaspoon*(*x*). The discourse structure of (34) is regulated by the constraint (36) on building discourse structures:

- (36) **Availability for Multimodal Discourse** (Lascarides & Stone, 2009, pp. 424-425).

Let $S = \langle A, F, last \rangle$ be an SDRS for multimodal discourse (and so [...] *last* is a non-empty set of at most two labels). Where $\pi, \pi' \in A$, we say that $\pi > \pi'$ iff either π immediately outscopes π' or there is a label $\pi'' \in A$ such that $F(\pi'')$ contains the literal $R(\pi, \pi')$ for some subordinating rhetorical relation R (e.g. *Elaboration* or *Explanation* but not *Narration*). Let $>^*$ be the transitive closure of $>$. Then $\pi \in A$ is available in S iff $\pi >^* l$, where $l \in last$.

In (36) A is the set of labels for SDRSs, F the function assigning the semantics to members of A and *last* contains the label of the last processed clause DRS, and if that clause was accompanied by a gesture, the label of that gesture. Newly processed discourse units β have to stand in a relation $R(\alpha, \beta)$ with an available label α . Assuming in line with the previously mentioned *and then* axiom schema of SDRT that trigger phrases such as *(und) dann* (‘(and) then’) or *danach* (‘afterwards’) license *Narration*, the following representation is obtained by (36) for the structure of (34), where each atomic DRS (those with number subscripts with their labels) is indicated by keywords symbolic of its semantics:



By definition (31) the following labels of clauses can provide antecedent discourse

referents to the teaspoon: π_{15} itself, π_{14} since it is the left argument of $Elaboration(\pi_{14}, \pi_{15})$, π_{10} since it is the left argument in $Elaboration(\pi_{10}, \pi''')$ where π''' outscopes π_{15} , and π_1 by being the left argument in $Elaboration(\pi_1, \pi''')$, where π''' outscopes π_{15} . Assuming the more liberal constraint discussed above which involves SDRSs as well as left arguments instead of only DRSS in definition (31) does not change the predictions in (37) since no SDRS α is the left argument in a relation $R(\alpha, \beta)$ such that β outscopes or is equal to π_{15} . Therefore only the enumerated DRSSs will be discussed.

Depending on the judgment on a context just like (34) without the teaspoon segment π_7 and π_8 one might argue that stirring with some small object may be accommodated upon hearing/seeing the simultaneous π_{14} and π_{15} . The idea that it is precisely a teaspoon however could not be retrieved without explicit mention. Furthermore a theory postulating anaphoric dependencies to discourse referents cannot predict the felicity of indicating a small object with the gesture π_{15} since the verb *umrühren* ('to stir') does not evoke a discourse referent, which was introduced above to be triggered only by explicit linguistic means. This can be demonstrated easily by true pronouns. Consider the following minimal pair:

- (38) a. Alex_i rührt gerade die Suppe_j um. #Er/Sie/Es_k ist sehr groß.
 Alex stirs currently the soup around 3SG.M/F/N is very big
 intended: 'Alex is stirring the soup. It is very big.'
- b. Alex_i rührt gerade die Suppe_j mit einem Löffel_k um. Er_k ist sehr groß.
 Alex stirs currently the soup with a spoon around 3SG.M is very big
 big
 'Alex is stirring the soup with a spoon. It is very big.'

In (38a) the pronouns can refer to all discourse referents in the first clause matching their gender - *er* ('he') may refer to Alex if Alex is a man, *sie* ('she') to the grammatically female soup or Alex if Alex is a woman, and *es* ('it') cannot refer to anything since no neuter antecedent is available. Explicitly introducing a grammatically male spoon changes the situation for *er*. An analogous example can be constructed for the verb *zusammengemischt* ('combined') in π_{10} , where the discourse referents only include the representation of the speaker and the representation of all ingredients, and *gekocht* ('cooked') in π_1 . Since in definition (31) only an outscoping hierarchy on the right edge of the discourse can be considered and in the potential liberalization for an $R(\alpha, \beta)$ only DRSSs accessible from all conditions in all DRSSs of α are added, the accessibility of the teaspoon from π_8 in π_{15} is far from being a minor deviation from predictions. Inside π' the discourse referent teaspoon from π_8 by either version of the available antecedents constraint is only accessible to conditions in π_8 itself since it is the rightmost DRS within π' . The same can be said for the DRS conditions transitively nested within π'' - π' is the

right argument to a relation, but not a left one and thus not accessible to any condition of π_3 or π_4 .

Even if one assumed that the left argument α from definition (31) may just be some SDRS *outscoping* the DRS providing the antecedent, (31) could not be saved. π'' , which outscopes π_8 , cannot be relevant to π'''' , which outscopes π_{15} , since there is no relation $R(\pi'', \pi''')$, not even a chain of relations of a form such as $R(\pi'', \beta), R(\beta, \gamma), R(\gamma, \pi''')$. Therefore any revision of (31) to make the pronoun hypothesis work in SDRT would require making the left argument α such that it outscopes the DRS with the correct discourse referent antecedent transitively somehow *and* would require accounting for the lack of direct or indirect relationship between π'' and π'''' . If there was such a constraint, it would be far too liberal to capture basic intuitions about the phenomenon (31) was originally designed for: pronouns. Consider a variation on discourse (34) where π_{14} and the gesture π_{15} are replaced by (39):

(39) π_{14+15} #und am Ende hab ich-s noch mal gut mit ihm umgerührt
 and at_the end have I-it yet again well with him stirred
 intended: ‘and in the end I stirred it well with it once more’

By phi features *mit ihm* (‘with it’) should be able to pick up the teaspoon from π_8 just like the gesture does, but it cannot refer to the teaspoon mentioned long ago. This is not a question of word order or the PP in (39), since the following sentence is a perfectly acceptable replacement for π_{14} and π_{15} in (34):

(40) π_{14+15} ‐und am Ende hab ich-s noch mal gut mit dem Teelöffel
 and at_the end have I-it yet again well with the teaspoon
 umgerührt
 stirred
 ‘and in the end I stirred it well with the teaspoon once more’

This Section demonstrated the failure of a pronoun-based account on previously given gestural information for an antecedent discourse referent evoked in speech only. The next part investigates the possibility of re-using discourse referents evoked by gesture alone.

3.2 Gesture-only antecedents

Two of Lascarides & Stone’s (2009) new rhetorical relations, *Overlay* and *Replication*, rely on the fact that gestures can build on information evoked by previous gestures. In particular *Replication* should connect gestural discourse units in order to account for the phenomenon that the same hand shapes and movements are used to convey meaning about the same entities or events (see Lascarides & Stone, 2009, pp. 406, 424). If *Replication* is taken to be a necessary licensing condition for coreference between gestures this should

constrain possible antecedents further than just the antecedent constraint (31) because attachment by *Replication* is also subject to the availability constraint (36). Once more the inadequacy of a pronoun-based account is demonstrated by a large monologue involving complex nested structures:

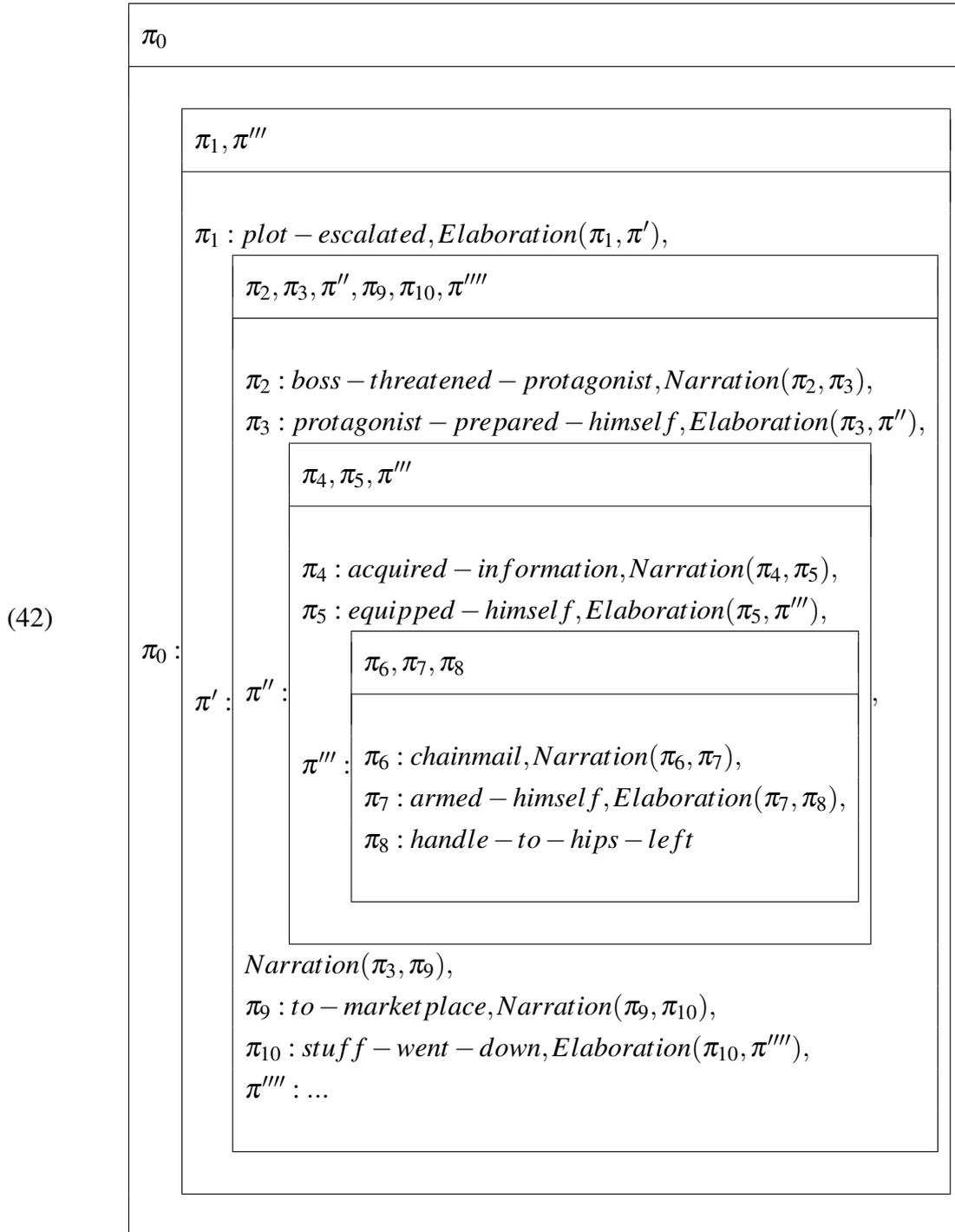
- (41) (Context: The speaker and the addressee started watching a movie together, but the addressee had to stop watching it in between. The movie is an action movie in a medieval setting. After finishing the movie, the speaker is now informing the addressee about the plot occurring after they had left the room.)
- π_1 Ey in dem Film ist die Handlung gegen Ende so richtig eskaliert.
 hey in the movie is the plot towards end so really escalated
 ‘Man, the plot of the movie escalated really hard towards the end.’
- π_2 Erst hat die Chefin vom Hauptcharakter ihn richtig offensichtlich
 first has the boss of_the protagonist him really obviously
 bedroht,
 threatened
 ‘First the boss of the protagonist threatened him pretty obviously,’
- π_3 und dann hat der sich natürlich ordentlich vorbereitet.
 and then has he_in_turn himself of_course adequately prepared
 ‘and then of course he in turn prepared himself accordingly.’
- π_4 Er hat sich erst nach ihren Plänen bei diesem Informanten
 he has himself first after her plans at that informant
 umgehört
 acquired_information
 ‘First he acquired information about her plans from that informant’
- π_5 und hat sich dann ausgerüstet.
 and has himself then equipped
 ‘and then he equipped himself.’
- π_6 Der hat sich son Kettenhemd unauffällig unter seinen normalen
 he has himself some chain_mail inconspicuously under his normal
 Klamotten angezogen
 clothes dressed
 ‘He put on some kind of chain mail stealthily under his normal clothes’
- π_7 und dann hat er sich bewaffnet.
 and then has he himself armed
 ‘and then he armed himself.’
- π_8 *handle – to – hips – left*
- π_9 Dann ist er zum Marktplatz gegangen
 then is he to_the marketplace gone
 ‘Then he went to the marketplace’
- π_{10} und dann ging-s richtig los.
 and then went-it really off
 ‘and then stuff really went down.’

- π_{11} Also π_{12} [da war dann der Hauptcharakter,
so there was then the protagonist
'So the protagonist was there,']
- π_{12} *right – hand – middle*
- π_{13} π_{14} [da links von ihm ne Kirche,] π_{12}
there left of him a church
'there to his left there was a church,']
- π_{14} *left – hand – left*
- π_{15} π_{16} [da n Haus] π_{14}
there a house
'there was a house'
- π_{16} *right – hand – right*
- π_{17} π_{18} [und da n Haus.] π_{16}] π_{18}
and there a house
'and there another house.'
- π_{18} *left – hand – front*
- π_{19} Dann springt halt einfach so-n Typ [mit ner Waffe] π_{20} hinter der
then jumps PRT just such-a guy with a weapon behind the
[Kirche] π_{21} raus.
church out
'then this guy with a weapon just comes out from behind the church.'
- π_{20} *right – holding – vertical*
- π_{21} *left – swipe – from – left*
- π_{22} Der Hauptcharakter hat halt [gleich reagiert] $\pi_{23,24}$,
the protagonist has PRT immediately reacted
'The protagonist reacted immediately'
- π_{23} *left – pull – from – hip*
- π_{24} *left – thrust – left*
- π_{25} aber damit war-s halt noch nich getan
but with_that was-it PRT yet not done
'but at that point it wasn't over yet.'

The idea of π_{25} is to indicate that the monologue continues afterwards with story elements irrelevant to the present linguistic discussion. Gesture π_8 , indicated by the notation *handle – to – hips – left*, is the left hand grasping an imaginary object in front of the speaker and leading the closed fist to the left side of the speaker's hips, representing a dagger or a knife, possibly being attached to a belt. In π_{11} , π_{13} , π_{15} and π_{17} the opening *and* closing brackets bear subscripts indicating gesture timing since the co-speech gestures π_{12} , π_{14} , π_{16} and π_{18} overlap in temporal alignment, with one gesture of each hand being executed at the same time. π_{12} is the open right hand doing a small downwards movement with the palm downwards and then staying in front of the speaker, fingers slightly bent, as if an object was placed on a table. π_{14} is the same hand shape and movement, but with the left hand left of where the right hand is lingering. π_{16}

represents the same with the right hand taken from the front location of π_{12} and instead being placed to a spot to the right of the original location. Finally π_{18} moves the left hand further to the front middle of the speaker, to a greater distance than where π_{12} was signed. The imaginary map drawn in this sequence indicates that the protagonist was situated between a church and a house, with yet another house forming a triangle with the other two buildings. This kind of spatial mappings receives an explicit semantic treatment by Lascarides & Stone (2009) and involves connecting the gestures by *Overlay*, but the details of the formulae are not important, unlike the structure they cause.

The gesture π_{20} is a fist with the knuckles facing right, therefore allowing for the interpretation that it depicts the weapon (due to the setting, possibly a sword) held vertically in the hand of the attacker. π_{21} is the left hand index finger pointing to the left, then swiping such that it points to the right and indicates the movement from left (the location of the church) towards right (the relative location of the protagonist to the church by the perspective taken in the map gesture sequence starting with π_{12}). π_{23} is the movement of the left fist drawn abruptly from the left hip towards in front of the speaker, whereby the fist slightly is rotating from the knuckles facing left at the hips, but facing the top in front of the speaker. This indicates the dagger drawn from the hip location. The last gesture π_{24} is a thrust to the left whereby the knuckles of the fist stay oriented upwards and the imaginary dagger moves to the left, in a straight line parallel to the speaker, the speaker's upper body slightly tilting to the right. With all the explanations at hand the assumed overall structure is provided in (42), while π'''' is presented separately:



Again a combination of *Elaboration* and *Narration* relations force a deeply nested discourse structure in (42). The relevant antecedent gesture is π_8 , where a dagger is evoked by the gesture, but not as a discourse referent in π_7 since a verb is usually observed not to introduce discourse referents (see Kamp & Reyle, 1993a, 1993b or the minimal pairs discussed by Heim, 1982). If gestures can only depict the same entities in a similar way as pronouns refer, by definition (31) above no DRS condition outside of π_8 itself is allowed to indicate the dagger. No DRS is in a rhetorical relation with π_8 as the left argument, and no SDRS outscoping π_8 is. Again consider the idea that “accessible from all conditions” in an SDRS α might amount to discourse referents in the leftmost

DRS inside α - the discourse of (42) is constructed such that this idea is not applicable, since π_8 is the last DRS inside its immediately superordinate SDRS π''' . The monologue (41) forces the deep embedding in (42) by *Narration* trigger words like *erst* ('first') or *dann* ('then') and by suitable candidates for a subtype relation for *Elaboration*: if the protagonist arms himself in the way depicted by π_8 , this is a sufficient condition to him arming himself, which in turn is sufficient for equipping himself, which is a subtype of preparing himself. The structure becomes more complicated with π'''' , whose content is abbreviated with ... in (42), but specified in (43) below:

$\pi_{FT}, \pi_{FT^{bg}}, \pi_{FT^{fg}}$	
$\pi_{FT} : FT, Background(\pi_{FT}, \pi_{FT^{bg}}),$	
$\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14}, \pi_{15}, \pi_{16}, \pi_{17}, \pi_{18}$	
$\pi_{11} : protagonist - there, Continuation(\pi_{11}, \pi_{13}),$ $\pi_{12} : right - hand - middle, Overlay(\pi_{12}, \pi_{14}),$ $\pi_{13} : church - left, Continuation(\pi_{13}, \pi_{15}),$ $\pi_{14} : left - hand - left, Overlay(\pi_{14}, \pi_{16}),$ $\pi_{15} : house - there, Continuation(\pi_{15}, \pi_{17}),$ $\pi_{16} : right - hand - right, Overlay(\pi_{16}, \pi_{18}),$ $\pi_{17} : other - house - there,$ $\pi_{18} : left - hand - front$	
$\pi_{FT^{bg}} :$	
$Elaboration(\pi_{FT}, \pi_{FT^{fg}}),$	
$\pi_{19}, \pi_{20}, \pi_{21}, \pi_{22}, \pi^{ana}, \pi_{25}$	
$\pi_{19} : guy - with - weapon, Depiction(\pi_{19}, \pi_{20}),$ $\pi_{20} : right - holding - vertical,$ $Depiction(\pi_{19}, \pi_{21}),$ $\pi_{21} : left - swipe - from - left,$ $Narration(\pi_{19}, \pi_{22}),$ $\pi_{22} : reacted - immediately, Elaboration(\pi_{22}, \pi^{ana}),$	
$\pi_{FT^{fg}} :$	
π_{23}, π_{24}	
$\pi_{23} : left - pull - from - hip,$ $Replication(\pi_{23}, \pi_{24}),$ $Narration(\pi_{23}, \pi_{24}),$ $\pi_{24} : left - thrust - left$	
$\pi^{ana} :$	
$Narration(\pi_{22}, \pi_{25}),$ $\pi_{25} : not - over - yet$	

The present analysis deviates from the usual SDRT standard adopted from Asher & Lascarides (2003) by assuming a different structure for *Background* relations, which is

inferred if one of the arguments contributes a stative event and the other does not (Asher & Lascarides, 2003, p. 207). *Background* entails temporal overlap between the events, where the stative event should provide the less important information to a common topic (Asher & Lascarides, 2003, pp. 165-166). According to an alternative structural analysis by Asher, Prévot & Vieu (2007) multiple backgrounded clauses like from π_{11} to π_{18} in (41) are contained in one SDRS, and *Background*(α, β) does not hold between a background α and foreground β , but rather between the common topic or *framing topic* α and the background β . In (43) the background is labeled π_{FTbg} , the framing topic is labeled π_{FT} and the foreground is labeled π_{FTfg} . A foreground β is taken to be related to a framing topic α by *Elaboration*(α, β) (Asher et al., 2007, p. 14). Since background structures have special binding properties, certain discourse referents may percolate (Asher et al., 2007, p. 17) from π_{FTfg} or π_{FTbg} to π_{FT} in (43), a possibility that has to be considered when testing the hypothesis at hand.

The content of the clauses of π_{FTbg} with the mapping gestures was included in (41) in order to test potential modifications to discourse structure for gestures. While Lascarides & Stone (2009, p. 417) hold that l in the availability constraint (36) can only be the last gesture if the last processed speech clause was accompanied by a co-speech gesture, conceivable modifications of anaphoricity constraints could relax this. A possibility would be to allow l to be the last overall gesture or even earlier gestures connected to that last gesture by sequences of *Replication* or *Overlay*, since they indicate a common way of depicting entities. In (43) however when processing the multimodal unit π^{ana} the last processed gesture would then be π_{21} , a pointing gesture relying on the map drawn in π_{FTbg} . This spacial dependence cannot be captured by *Overlay* however since π_{21} is simultaneous with the foregrounded clause π_{19} while the mapping is done within the backgrounded π_{FTbg} . This demonstrates a further issue with Lascarides & Stone's (2009) structural account: The phenomenon *Overlay* is supposed to formalize is possible across SDRS boundaries and thus cannot be captured by a relation like *Overlay* in all cases. Note also the intervening gesture π_{20} , which either contains no spatial component or could at most be taken to switch perspective towards the attacker's viewpoint from his left location near the church. Neither the gestures within π_{FTbg} nor π_{20} nor π_{21} bear any relation to the dagger indicated by π_8 , which is naturally taken to be the same dagger as the one represented by $\pi_{23,24}$. As a consequence the felicity of π_{23} and π_{24} cannot be saved by modifications on what it takes to be the last processed gesture or by multiple *Replication* relations allowing for access to π_8 by π_{23} . Note at this point that *Replication* faces the same problem as *Overlay*: π_8 and π_{23} depict the same object (the dagger) in the same way (indicating the handle with a fist) but cannot be linked directly by *Replication*(π_8, π_{23}) or indirectly by a chain of *Replication* relations. This means no less than that either rhetorical structure has to be worked around, or that

rhetorical relations cannot capture the phenomenon of coreference between two gestures.

The intervening gestures therefore reduce the question whether gestures necessarily behave like pronouns to the pronoun constraint (31). Whether the discourse referent serving as the anaphor is located within π_{FT} or π_{23} , by means of outscoping relations the ‘closest’ right argument γ to a relation $R(\alpha, \gamma)$ in the sense of definition (31) will be π''' , any higher SDRS outscopes both the antecedent in π_8 and the anaphor in π_{23} . Then the same kind of criticism applies as with the speech antecedent example (37): In (42) the most embedded SDRSs contained within the same SDRS and outscoping the antecedent and the anaphor are π' and π''' , which are not related rhetorically, and even they were, the antecedent containing DRS π_8 is embedded several levels inside π' , and even to the right edge within π' . Every briefly discussed version or modification of the formal system in order to save the pronominally constrained analysis of gesture felicity within SDRT therefore strongly undergenerates.

Note that a simple speech paraphrase of $\pi_{22,23,24}$ including the NP/DP *mit dem Dolch* (‘with the dagger’) or *mit ihm* (‘with it’) is odd:

- (44) $\pi_{22}/\#$ Der Hauptcharakter hat gleich mit ihm / dem Dolch reagiert.
 the protagonist has immediately with him / the dagger reacted
 intended: ‘The protagonist reacted immediately with the dagger.’ or ‘The protagonist reacted immediately with it.’

This is however due to the circumstance that *reagieren* (‘to react’) is hardly acceptable with instrumental adjuncts with *mit* (‘with’). Less direct paraphrases in the very same discourse structural position allow for definite descriptions referring to the dagger, but not for the corresponding pronoun:

- (45) π_{22}'' Der Hauptcharakter hat den gleich mit dem/seinem Dolch
 the protagonist has him immediately with the/his dagger
 abgewehrt.
 fended_off
 ‘The protagonist immediately fended off that guy with the/his dagger.’
 $\pi_{22}/\#$ Der Hauptcharakter hat den gleich mit ihm abgewehrt.
 the protagonist has him immediately with him fended_off
 intended: ‘The protagonist immediately fended off that guy with it.’

This section demonstrated that the pronominal hypothesis for iconic co-speech gestures cannot be upheld, neither in SDRT nor in general, neither for antecedents evoked in speech nor for those evoked in gesture. Gestures indicate the usage of entities which were mentioned or signed in completely unrelated discourse units, but still manage to convey the idea that the entity signed later is the exact same as the one mentioned or

signed before. An SDRT specific problem arises if *Overlay* and *Replication* are taken to be necessary conditions on similarly depicted equal locations or entities across gestures. This means that a theory which aims to account for all anaphoric dependencies requires an mechanism, since *Overlay* and *Replication* are not sufficient for dependencies between gestures. Note however that paraphrases with definite descriptions, unlike those involving pronouns, seemed to parallel the behavior of the corresponding gestures. Such an approach also seems to be appealing for the reason that it allows uniform treatment of anaphoric gestures irrespectively from the modality of the antecedent, which in turn captures best the behavior of the two large monologues of this section. Because of these parallels the next section is devoted to the idea that all gestures involving referents given before by discourse or context have access to such referents by the same rules as definite descriptions.

4 Gestures with definite description semantics

This section departs from SDRT alone and takes a more empirical approach, since SDRT has no theory of definite descriptions that could account for the phenomena motivating this thesis. The bridging theory of Asher & Lascarides (1998) allows to specify some anaphoric dependence of definite descriptions to some accessible antecedent, but recurring to accessibility allows for short-distance dependencies only. The observation by Lascarides & Stone (2006, 2009) that gestures have to be related to entities or events in simultaneous speech was suggested to involve a bridging constraint of this kind. This however only accounts for the immediate coherence of the gesture, not for coreference with entities introduced in a distant discourse segment. So the constraints on such long distance dependencies are to a considerable degree independent of accessibility and bridging as described by Asher & Lascarides (1998) or Lascarides & Stone (2006, 2009).

The structure of this section is as follows: First an overview of the literature on definite descriptions will be provided in order to establish criteria to test the definite description hypothesis. A brief discussion on definite descriptions in comparison to pronouns and the definite/indefinite distinction follows as the second part. The largest part is devoted to the criteria of existence and unique maximality which are usually assumed for definite descriptions, irrespectively of the choice of formalization. It commences with a discussion of the number properties of the gesture anaphor, where it is argued that iconic gestures may be anaphorically related to an antecedent representing multiple entities, but any subset of the antecedent renders the gestural contribution felicitous. Afterwards it is argued that despite of this, some notion of unique maximality can still be upheld for determining the antecedent, and that the introduction of relevant partitions in the set of entities suitable as antecedent can lead to a violation of that maximality

condition. The rest of this part is devoted to establishing the correct denotation of a gesture for the maximality effect to work best, where the contextually unresolved form predicate will be advocated as the right choice. Finally some ideas how to capture the vague notion of relevance with SDRT and ingredients of the QUD framework are presented. The main findings are summarized in a gesture resolution algorithm.

After having discussed how to account for antecedent existence, maximality and relevance the findings are integrated formally into SDRT. Finally the idea of an attributive/referential distinction raised by Donnellan (1966) is discussed with respect to iconic gestures.

Throughout the remainder of the thesis the following terminology will be adopted: An *antecedent* is a contextually given entity definite descriptions can pick up, whereby the antecedent is not necessarily a discourse referent. Likewise the term *coreference* will imply identity between the antecedent and the semantic contribution of the description or gesture without necessarily implying the presence of a discourse referent. The term *anaphoric* in turn will be restricted to cases where the antecedent is established by a discourse referent.

4.1 Formal literature on definiteness

The observation that definite descriptions carry some sort of uniqueness requirement is already noted by Mill (1843, p. 33). Another early, still influential account on definite descriptions is outlined by Russell (1905). He holds that the logical form of a sentence with a definite description as grammatical subject does not reflect the grammatical status as a proper name would do (Russell, 1905, p. 488), but rather develops an account overlapping with the treatment of quantifier phrases (Russell, 1905, p. 481). A definite description results in a threefold composition of sentence meaning: an assertion of the *existence*, an assertion of *uniqueness*, and the predication provided by taking the remainder of the sentence into account (Russell, 1905, pp. 481-482). Assuming an example sentence such as *The dog is happy* an example formula could therefore be written as $\exists x[\text{dog}(x) \wedge \forall y[\text{dog}(y) \rightarrow y = x] \wedge \text{happy}(x)]$: there is an x such that it is a dog, there is no other entity y such that it is a dog (formalized as equality with x for all dogs y), and this bound variable x is the argument to the VP semantics *happy*.

Russell (1905, pp. 482, 485, 490) observes that the consequence of this analysis is that in every sentence with a definite description where the NP is true for more than one entity or zero, such as in *The present king of France is bald*, makes the proposition false, a conclusion he supports by the law of the excluded middle. For sentences like

George IV. wished to know whether Scott was the author of Waverley (Russell, 1905, p. 489) he essentially develops an account of relative scope. He allows the existential and universal quantification to take wide scope over the modal *wished to know* to account for the problem that a simple term x as the meaning of both *Scott* and *the author of Waverly* would result in the proposition that George IV. was interested in knowing whether $x = x$, a tautology, is true.

While Russell advocates an analysis which relies on scope ambiguities, Strawson (1950) rejects it on the grounds of pragmatic observations. He stresses the difference between the conventional meaning of a definite description and its use or reference, providing the example of *the king of France* being uttered during the reign of different kings (Strawson, 1950, p. 325). Importantly, Strawson (1950, pp. 330, 333) supports the uniqueness and existence inferences observed by Russell (1905), but rejects the idea that they are being asserted rather than inferred in a different way. This kind of inference is identified by some authors as presupposition (e.g. Karttunen, 1973; Heim, 1983; Heim & Kratzer, 1998, p. 81) which may be analyzed by context models based on the Common Ground (Stalnaker, 1973, 1978).

The referential analysis of definite descriptions is unable to account for some phenomena by itself, but may be supplemented with additional assumptions. For example Russell (1905, p. 490) attributes an ambiguity to *The king of France is not bald*, with one reading where for an existing unique king of France baldness is denied and one where the negation takes scope even over existence and uniqueness. The first reading amounts to the paraphrase *There is exactly one king of France, and he is not bald*, while the second one amounts to *It is not the case that there is exactly one king of France which is bald*. Horn (1985) however argues that a mechanism of *metalinguistic negation* is necessary by virtue of independent data, e.g. phonetic corrections, which then would allow a scopeless account of definite descriptions supplemented with a metalinguistic presupposition denial for existence and uniqueness.

Whether a quantifier scope account in the spirit of Russell (1905) is sufficient or even valid is the subject of an ongoing debate. While for example Bach (2004), Salmon (2004) and Nunberg (2004) do not consider a referential approach very convincing, Abbott (2010, p. 152) and Devitt (2004) contribute arguments in favor of it. Another matter of research is the question whether admitting a referential approach should be encoded in semantic or pragmatic terms. In addition the approach of Heim (1982) requires definite descriptions to be *familiar*, that is, having a discourse representation motivated by observations about discourse referents, thus omitting uniqueness as such. While examples of unique discourse-new definite description referents are easily found

discourse referents will often be mentioned separately throughout this thesis, since they provide clearer judgments than contextual information from the Common Ground.

Complications for testing scopal predictions of Russell's (1905) theory for gestures arise by their modality itself. Informative identity statements such as *Scott is the author of Waverley* of the form *x is y* could be constructed with co-speech gestures, but then the speech content would prevent attributing the data directly to the gestures themselves rather than to the speech content. Replacing *x* or *y* with a single gesture would constitute a *pro-speech gesture* rather than a co-speech gesture, which Schlenker (2018b, p. 886) argues to be at-issue in contrast to the co-speech data discussed in this thesis. Therefore it is unclear if observations on scope would be meaningful for the present purposes. While this thesis assumes a presuppositional account to existence and uniqueness for definite descriptions, it should therefore be noted that this approach is taken for the sake of concreteness rather than motivated independently. Note also that the not-at-issue status of co-speech gestures is argued for by Ebert & Ebert (2014) and Schlenker (2018a) and should be generally kept in mind when constructing minimal pairs with speech definite descriptions - additional pragmatic constraints may apply for linking the potential definite description gesture to the speech content.

It will be assumed that contextually salient uniqueness is sufficient for definite descriptions (Birner & Ward, 1994), while the exact scope of uniqueness will be treated neutrally. For example Löbner (1985) provides an account based on situation semantics (Barwise & Perry, 1983), a degree of detail of which this thesis abstracts away. Note also that work on uniqueness as such often ignores additional complications with plural definite descriptions, which is why Sharvy (1980) introduces a notion of a *unique maximal* antecedent, which will be discussed in more detail at a later point.

4.2 Definite descriptions, pronouns and the definite/indefinite distinction

In order to determine the definite description status of gestural discourse referents they have to be separated from other kinds of meaning. One potential alternative analysis as pronouns was already challenged by distributional data in Section 3. A further difference between pronouns and definite descriptions is that pronouns only carry very restricted kinds of grammatical information in order to determine suitable antecedents, while definite descriptions contain lexical information provided by the NP as well. For example, the German pronoun *ihn* ('him') carries information about case (accusative), number (singular) and gender (masculine). Assuming that the accusative is more or less irrelevant for determining a suitable antecedent, such an antecedent to *ihn* therefore

has to be some singular referent introduced with grammatical masculine gender. The definite description *den Hammer* ('the hammer') carries the very same case, number and gender information, but in addition provides the lexical information that the suitable antecedent has to be a hammer. Such additional information arguably allows the antecedent resolution over larger distances within discourse as demonstrated with the pronominal and definite description paraphrases in Section 3. This difference in usage of definite descriptions and pronouns has been postulated before (see Prince, 1981; Gundel, Hedberg & Zacharski, 1993), but there is no clear-cut boundary in felicity of usage of one over the other. Over large distances in discourse however, as in the discourse examples of Section 3, judgments become much clearer.

Preiss, Gasperin & Briscoe (2004) compared the results of an algorithm modified from Lappin & Leass (1994) for resolution of pronominal anaphora in a given corpus with the results obtained when definite descriptions were replaced by pronouns. While the algorithm correctly resolved 62% of genuine inter-sentential anaphora, it only resolved 23% or 29% (depending on the weighting mechanism) of replaced definite descriptions correctly (Preiss et al., 2004, p. 1501). In addition a simple comparison algorithm choosing a suitable antecedent of a definite description only if the antecedent NP token is equal to the one of the definite description provided an accuracy of 62%, which suggests that the lexical NP information plays a crucial role for determining the antecedent (Preiss et al., 2004, p. 1501). With iconic co-speech gestures, it is arguably only the iconic information that helps determining a suitable antecedent, since they do not contain grammatical information such as gender. It seems plausible to draw an analogy between the iconic component of a gesture (excluding mere deictic gestures) and the lexical NP predicate information of a definite description.

Differentiating definite descriptions and pronouns by other properties is complicated by the fact that both require an existing and unambiguous antecedent and both are frequently claimed to come with a uniqueness requirement (see e.g. Kadmon, 1990; Abbott, 2010, p. 216). The most striking difference is that pronouns always require discourse referent antecedents while definite descriptions can be licensed by contextual uniqueness alone (Birner & Ward, 1994). Without having established where to draw the boundary between indefinite and definite gestures uniqueness without a familiar discourse referent however cannot be used to argue for a possible definite description approach to gestures. This is due to the fact that unfamiliar yet unique referents can be claimed to be unique by accident while no linguistic difference can be found to non-unique referents. In Section 4.3.2 it will be argued that there is in fact a difference between gestures that are plausibly resolved by bridging to concurrent speech but are not saliently given in previous context or discourse and those which rely on given information to be felicitous. Since the latter

will be shown to obey certain characteristics of definite descriptions, these are the ones that will be regarded as *definite* under the account presented in this thesis. The gestures inferrable by basic bridging coherence can be regarded as *indefinite* under the present account, and they do not require given antecedents or unique or maximal resolution.

4.3 Existence and unique maximality

4.3.1 Partitives and general number

Since a uniqueness requirement in some sense underlies most approaches to definiteness, it is necessary to investigate if the usage of a coreferential iconic gesture implies that there is only one suitable antecedent entity. A first naive conception of uniqueness has to be rejected. Consider the following monologue, where *throw – smth – long* represents the movement and hand shapes performed when throwing an elongated object, like a stick or in this case a boomerang:

- (46) (Background information: Tom’s mother has four boomerangs.)
- a. Tom hat sich die Box mit den Bumerangs seiner Mutter genommen,
Tom has himself the box with the boomerangs of_his mother taken
um mit ihnen zu üben.
for with them to practice
‘Tom took the box with his mother’s boomerangs to practice with them.’
 - b. Er hat schon [angefangen]*throw–smth–long* gehabt, als er erfahren
he has already started had when he found_out
hat, dass er das nicht darf.
has that he that not is_allowed
‘He had already started when he found out he wasn’t supposed to do that.’

Note that the discourse structure is not as elaborated as those in Section 3 since the predictions of SDRT for anaphors are considered disproven to be relevant for the phenomenon at hand at this point. Here clause (46a) mentions multiple boomerangs of Tom’s mother, and the gesture *throw – smth – long* in clause (46b) indicates them in a certain way. However, a witness of the events narrated in (46) would not consider the utterance of (46) infelicitous if Tom had already practiced with more than one and less than all boomerangs. Even if Tom practiced with two of them and then he was informed his mother did not appreciate the action - the gesture still is felicitous. The first consequence of this is that uniqueness cannot be taken to mean that plural antecedents to the gesture are prohibited.

It is not a problem for a definite description analysis that the gesture may pick up more than one boomerang, but it is a problem that it does not pick up all boomerangs. Plurals are commonly assumed to be captured well in semantics by grouping the singular

entities into a complex one of the same type (see e.g. Link, 1983; Farkas & de Swart, 2010). Single objects can be referred to as *atoms*, combinations of atoms as *pluralities*. Sharvy (1980, pp. 615-616) observes that plural definite descriptions such as *the people in Auckland* comprise the maximal plurality of people in Auckland while there is more than one plurality of people that would satisfy the property of being people in Auckland, e.g. all women in Auckland or all men in Auckland. The definite description will fuse all atoms and pluralities until the antecedent is the *unique maximal* plurality which still satisfies the predicate *people in Auckland*. Applying this observation to the discussion of (46) the unique plurality of all four assumed atomic boomerangs would be the correct antecedent under a definite description approach. The felicity in a scenario with two of four boomerangs thrown in the events of (46) cannot be captured. Neither coreference with a unique atom nor with a unique sum can be maintained. The fact that coreference with parts of the antecedent is inferred will be called *partial coreference* from now on.

The only close paraphrase capturing partial coreference would be *(mit) ein paar von den Bumerangs* ('(with) some of the boomerangs'). This construction is referred to in the literature as a *partitive*. In the compositional analysis by Barker (1998, pp. 698-699) the phrase *of the boomerangs* receives a semantics of an $\langle e, t \rangle$ type predicate whose argument is required to be a (proper) part of the denotation of *the boomerangs*. Under the assumption that *some* is best formalized by a simple existential quantification of type $\langle \langle e, t \rangle, t \rangle$ the paraphrase of the gesture in (46b) would in fact provide adequate felicity conditions.

This is however not to say that a partitive paraphrase captures all usages of iconic co-speech gestures or that it is well motivated by now. Rather the partitive approach is required to uphold an analysis of (partial) coreference with gestures by recourse to definite descriptions. Sections 4.3.2 to 4.3.3 will motivate the approach, at this point it is only introduced as a necessary ingredient.

Note that if *mit den Bumerangs* ('with the boomerangs') in (46a) was replaced by *mit dem Bumerang* ('with the boomerang'), the gesture in (46b) would still be able to refer back to that one boomerang introduced as a discourse referent. The antecedent therefore is allowed to be an atom or a plurality for the same gesture. The number properties of the subgroup evoked by the gesture are the same: In (46) where there were four boomerangs, a witness of the narrated events would not object to the discourse if Tom had thrown an arbitrary number from one to four of the boomerangs. Rullmann & You (2006) however discuss such a phenomenon for bare nouns in Mandarin Chinese. They argue that such a noun without further indication of number can represent atoms and pluralities, a property they call *general number*, e.g. that the word *shu* can mean *book*

or *books* (Rullmann & You, 2006, pp. 179-180). The absence of overt number marking in most gestures may explain how gestures resemble Mandarin Chinese bare nouns.

Assuming a partitive approach with a (not proper, but reflexive) part-of relation \leq where the unique maximal antecedent is written $\mu x[P(x)]$, a q such that $q \leq \mu x[P(x)]$ will be referred to as a *subgroup*. The subgroup and the antecedent can be considered to have to be general number¹². The next Sections are devoted to provide evidence that the seemingly *ad hoc* partitive approach in fact is supported by existence and maximality conditions for the antecedent.

4.3.2 Felicity by existence condition or bridging alone

Reed (1991, 1996) advocates capturing partitives in terms of discourse referents. According to her they introduce subgroups of pluralities already present in the dynamic discourse model. Such a view on gestures would allow to capture the partial coreference intuition that although the entities evoked by the gesture in (46b) do not have to be identical with the maximal plurality of the boomerangs mentioned in (46a), they are identical with a selection of all those boomerangs. However, it is necessary to motivate the analysis by cases where the existence of a previously introduced discourse referent makes an otherwise infelicitous context felicitous¹³, and by cases where a unique maximality requirement holds for atoms and pluralities. In (46) the subgroup does not consist of arbitrary boomerangs instead of a subgroup, but this can hardly be considered evidence for an existence and uniqueness requirement, since the subgroup inference would be present in (46) without the gesture as well.¹⁴

The point of a genuine existence requirement of an antecedent for some gestures can be made, even though many gestures are felicitous without an antecedent. Some gestures need no context (such as the gesture of handling a knife when using the verb *cut* or holding a cup when saying something like *pour in*). In other cases, however, there is no obvious link between an event and a gesture, for example in (47), where the gesture *small – into – mouth* is a movement of the hand towards the mouth of the speaker, with the tips of thumb and index finger in pinching shape.

¹²General number may also be promising property for indefinite gestures lacking an antecedent.

¹³Kendon (2004, p. 11) notes that some gestures are perceived as accidental by the addressee and are largely ignored. An interesting research question would be to which extent definite gestures are exempt from felicity conditions if they are signed casually enough to be considered unintended.

¹⁴The example (46) still serves the purpose of establishing felicity conditions for the construction as a whole. Under a uniform treatment with gestures which are necessary to trigger partial coreference inferences this may ultimately mean that the gesture in (46b) points towards a behavior where gestures may be potentially informative but not triggering redundancy-based infelicity effects, a set of properties compatible with the findings of Lascarides & Stone (2006, 2009), Ebert & Ebert (2014) and Schlenker (2018a).

- (47) #Sie hat sich [entschieden] *small-into-mouth* ·
 she has herself decided
 intended: ‘She made up her mind.’

Such a construction can become felicitous by a suitable antecedent providing information about what the gesture is supposed to depict, which object may be involved and how this relates to the speech component. This will create a coreference inference rather than introducing arbitrary objects by a habitual connection between the speech predicate and the gesture:

- (48) a. Der Arzt hat Alex eine Tablette gegeben und ihr gesagt, dass die ihr
 the doctor has Alex a pill given and her told that it her
 helfen kann, aber auch heftige Nebenwirkungen haben kann.
 help can but also severe side_effects have can
 ‘The doctor gave Alex a pill and told her that it could help her, but that it
 could cause severe side effects as well.’
- b. Sie hat dann ne Weile drüber nachdenken müssen, aber dann hat sie
 she has then a while about_it reflect must but then has she
 sich [entschieden] *small-into-mouth* ·
 herself decided
 ‘She had to think about it for a while, but then she made up her mind.’

(48) demonstrates that the presence of a suitable antecedent can license gestures that would not be licensed by general co-speech bridging mechanisms alone. This assumes that bridging is in fact not constrained by accessible antecedent constraints of SDRT but is a local phenomenon: The content of a gesture has to be bridged to the content of the concurrent clause¹⁵, otherwise it lacks coherence. This holds for established referents just as for new objects. Without adequately rich context however bridging seems to be rather restrictive. For this reason it is proposed in this thesis that the bridging constraint in a local version is valid independently of the existence requirement of some co-speech gestures, and that fulfilling the existence requirement of a coreferential reading licenses adding new local bridging relations or makes them more plausible.

Note that this is not to say that the existence requirement can only be fulfilled by genuine discourse referents, but this is not an issue to many accounts on definiteness (e.g. Hawkins, 1978; Löbner, 1985; Birner & Ward, 1994). If the variation (49) is continued with (48b) above, the gesture is felicitous as well:

- (49) Der Arzt hat Alex vorgeschlagen, sich medikamentös behandeln zu
 the doctor has suggested Alex herself medicinally treated to

¹⁵This may be restricted even further by observations on temporal alignment. See Alahverdzhieva & Lascarides (2010) for some ideas.

lassen, und ihr gesagt, dass das ihr helfen kann, aber auch heftige
let and her told that it her help can but also severe
Nebenwirkungen haben kann.

side_effects have can

‘The doctor suggested to Alex to get herself medicinally treated and told her
that it could help her, but that it could cause severe side effects as well.’

While *sich medikamentös behandeln zu lassen* (‘to get oneself medicinally treated’) certainly is compatible with the meaning to ingest pills, it invokes no discourse referent. However taking context and discourse into account there is only one action and one contextually given object, and not just some medicine will satisfy the gesture (e.g. pills against headache due to the difficult decision), only the pills constituting the treatment. In the terminology of Lambrecht (1994, pp. 77-78) one can express this by saying that context renders the object depicted by the gesture *identifiable* to speaker and addressee alike¹⁶.

There is a class of contexts where drawing the line between bridging and definite uses is more complicated. The reason is that contextual information can also license speech-gesture combinations which seem not to involve very specific entities but rather kinds of objects. Consider (50), where the resolution of *am arbeiten* (‘working’) neither necessarily involves slicing cheese in the context nor is slicing cheese a very common interpretation of saying that a person was working out of the blue, but the context licenses this interpretation of the scrubbing movement of a piece of cheese against a grater since it is one way the subject is commonly working in that context:

- (50) (Context: The speaker is a kitchen assistant in a restaurant. The speaker usually has only three tasks: cleaning dishes, stirring soup and grating cheese.)
- a. Gestern war ich grad [am arbeiten]_{scrub} als plötzlich...
yesterday was I currently at working when suddenly
‘Yesterday I was just working when suddenly...’

Under the additional assumption that the kitchen has a large stock of cheese graters and the speaker always just picks one at random, the gesture *scrub* evokes cheese graters, but not a particular one or one among a particularly salient set. If instead one knows that the kitchen has a large stock of graters, but the speaker may only use specific ones, there will be the stronger inference that the speaker did not just pick a random grater from the stock. It is not entirely clear if one should account for the reading of picking a random grater by allowing a contextually informed indefinite use or by a definite use with a partitive, where the definite antecedent is contextually

¹⁶Note the difference to specific objects only identifiable by the speaker, such as in *I am looking for a specific teacher, his name is Mr. Doe.*

restricted to the large stock of the kitchen. Since stronger partial coreference inferences need to be accounted for anyways, the latter is more appealing by uniformity of the theory.

Another complication when allowing contextually informed indefinite bridging is posed by the scope of uniqueness of some theories of definiteness. If in a certain office setting *to finalize a document* always entailed finalizing by placing one stamp on it, a stamping gesture would allow the bridging constraint by Lascarides & Stone (2006, 2009) to be fulfilled by invoking the stamp associated with each document. As Löbner (1985) already observed, the uniqueness requirement of definite descriptions may be subject to severe contextual limitations in terms of with respect to what the referent has to be unique. This makes distinguishing indefinite uses from definite ones in gestures more complicated. In the stamp example one might argue that there is no contextually salient particular stamp, so it is indefinite. However by the contextual information one may paraphrase the stamping gesture as *the respective stamp of the/each document*, which is unique when restricting the situation adequately.

It was demonstrated that the bridging constraint can be contextually informed to license new speech-gesture combinations and readings of partial coreference. It will be assumed that all contextually licensed bridging instances are definite uses if there is at least a large contextually restricted antecedent, a choice of definition which is not obligatory in light of the data but appealing in terms of theoretic uniformity. The observations so far can be captured by the algorithm below:

(51) **Algorithm for gesture licensing by existence of antecedents** (to be expanded for gesture resolution)

1. Access all subtypes of the type hierarchy resolving gesture form to interpretation proposed by Lascarides & Stone (2006, 2009) for the present gesture
2. For all multimodal utterances (i.e., speech and gesture) in the previous discourse related to the same topic, determine
 - (a) discourse referents which are an argument to a predicate in the discourse representation equivalent to a subtype of the gesture form (e.g. x if a DRS contains $pill(x)$ like in (48) and the form predicate *small – into – mouth*(i) has the subtype $pill(x)$)
 - (b) entities whose existence is contextually entailed by the discourse material which are a subtype of the gesture form (e.g. stamps or stamping events in discourses containing *finalized a document* with a gesture *stamp*, where $stamp(x)$ is a subtype of the form predicate $stamp(i)$)

(c) entities or events whose existence is contextually entailed by some more specific common resolution of slightly vague information, under which interpretation they are a subtype of the gesture form (e.g. getting oneself medically treated in (49) may be taken to mean ingesting pills, which is a subtype of the gesture form *small – into – mouth(i)*, and in (50) *scrub(i)* is licensed by the fact that its gesture subtype *grate – cheese(e)* or *cheese – grater(x)* is entailed by a contextually salient way to resolve the vague information that the speaker was working)

3. Form the set of *compatible* or *suitable* referents consisting of all referents and events obtained by step 2. If it is empty, check whether bridging to concurrent speech is possible for an indefinite use. If yes, use the gesture indefinitely, if not, raise infelicity.

4.3.3 Unique maximality and relevance

Uniqueness in one sense or another is the central component of most approaches to definiteness (e.g. Mill, 1885; Russell, 1905; Strawson, 1950; Hawkins, 1978; Löbner, 1985). Recall from examples (47) and (48) that a sentence of the form *x [made up their mind]_{small–into–mouth}*, where *small – into – mouth* is the movement of the pinching fingertips of thumb and index finger towards the mouth, is infelicitous without context, but can be licensed by an appropriate contextual reference. Recall also from the boomerang example (46) that the antecedent may be explicitly introduced by a plural DP. The felicitous examples have in common that there is exactly one singular or plural entity the gesture can refer to in terms of compatibility with the form of the gesture and in terms of their interpretation. For example when the form *small – into – mouth* was signed there was only the salient idea that some object(s) was put into someone's mouth, which in turn could only be a pill in the scenario, and there was but one kind of pill. A unique maximality constraint should not allow ambiguity about the possibly plural entity of which one or more atoms are ingested. Consider the monologue in (52) where the speaker narrates a highly symbolic moment they witnessed when visiting someone in the hospital:

- (52) a. Also der Arzt hat dem Niko halt zwei Pillen gegeben, ne rote und
so the doctor has the Niko PRT two pills given a red and
ne blaue.
a blue
'So the doctor gave Niko two pills, a red one and a blue one.'
- b. Und er meinte halt, wenn er die blauen anfängt zu nehmen, lebt er
and he said PRT if he the blue starts to take lives he
noch 10 Jahre, aber die meiste Zeit im Krankenhaus,
still 10 years but the most time in_the hospital

‘And he said if he started to take the blue ones, he’d live another 10 years, but most of the time in the hospital,’

- c. und wenn er die roten nimmt hat er noch ein Jahr, aber er ist bis
and if he the red takes has he still one year but he is until
kurz vor Ende noch fit.
short before end still fit
‘and if he took the red ones he’d have one more year but he’d be fit until short before the end.’
- d. Dann hat der Niko halt ne Weile überlegt, und dann hat er [sich
then has the Niko PRT a while reflected and then has he himself
entschieden]_{small-into-mouth} .
decided
‘Then Niko thought about it for a while, and then he made up his mind.’

If the narration of the speaker is assumed to end with (52d), it ends in an unsatisfactory and infelicitous way. The gesture can be easily resolved to involve one of the pills mentioned in the previous discourse, so one might argue there is a part of the resolution process that turns out successful. This would be resolving the form predicate towards the much more specific *pill(x)* or some related interpretation. Still, any listener interested about the content of (52) would wish to know *which* pill Niko took. There is a meaningful contrast between the choices implying considerable consequences, and the color of the pills cannot be encoded by standard iconic gestural means.

Constructing paraphrases in accordance with a uniqueness or unique maximality constraint could take two different approaches. According to the analysis as partitives the gestural subgroup would amount to one or more of all antecedent pills. An alternative is the presence of a uniqueness constraint without the automatic formation of maximal antecedents. If the antecedent pill is taken to be either the red or the blue pill, then there would be one pill in both the antecedent and the gestural contribution, which is reducible to the paraphrase *the pill*. In this case the infelicity of the lack of information in (52) has to be attributed to the fact that there is more than one possible antecedent - either the blue pill or the red one - and it is impossible to determine which is the correct one. If it is assumed that the antecedent is formed by an implicit summation of the red and the blue pills and thus is equal to the plurality of both pills, the gesture can be paraphrased to *one of the pills* and the infelicity has to be attributed to a relevance effect which disallows this vague contribution in this context. Both approaches seem plausible by the observation that both speech-only paraphrases *can* be grammatical and felicitous given the right context, but simply are not in the context and discourse of (52). First consider a variation of the context (52) where (d) is replaced by either (53a) or (53b):

- (53) (Previous discourse equals (52a-c))
- a. #Dann hat der Niko halt ne Weile überlegt, und dann hat er die Pille
 then has the Niko PRT a while reflected and then has he the pill
 genommen.
 taken
 ‘Then Niko thought about it for a while, and then he took the pill.’
- b. #Dann hat der Niko halt ne Weile überlegt, und dann hat er eine von den
 then has the Niko PRT a while reflected and then has he one of the
 Pillen genommen.
 pills taken
 ‘Then Niko thought about it for a while, and then he took of the pills.’¹⁷

Then take the following narration into account, where (54d), (54d’) and (54d’’) are alternative continuations from (54c):

- (54) a. Also der Arzt hat dem Niko halt zwei Pillen gegeben, ne rote und ne
 so the doctor has the Niko PRT two pills given a red and a
 blaue, also eine für abends und eine für morgens.
 blue, so one for in_the_evening and one for in_the_morning
 ‘So the doctor gave Niko two pills, a red one and a blue one, that is one for
 the evening and one for the morning.’
- b. Und er meinte halt, wenn er die Pillen anfängt zu nehmen, lebt er noch
 and he said PRT if he the pills starts to take lives he still
 10 Jahre, aber die meiste Zeit im Krankenhaus,
 10 years but the most time in_the_hospital
 ‘And he said if he started to take the pills, he’d live another 10 years, but
 most of the time in the hospital,’
- c. und wenn nicht, hat er noch ein Jahr, aber er ist bis kurz vor Ende
 and if not has he still one year but he is until short before end
 noch fit.
 still fit
 ‘and if he took the red ones he’d have one more year but he’d be fit until
 short before the end.’
- d. Dann hat der Niko halt ne Weile überlegt, und dann hat er [sich
 then has the Niko PRT a while reflected and then has he himself
 entschieden]*small-into-mouth* .
 decided
 ‘Then Niko thought about it for a while, and then he made up his mind.’
- d’. Dann hat der Niko halt ne Weile überlegt, und dann hat er eine von den
 then has the Niko PRT a while reflected and then has he one of the
 Pillen genommen.
 pills taken

¹⁷The same judgments apply for the phrasing *und dann hat er sich für die Pille entschieden* (‘and then he chose the pill’) and its partitive counterpart. Therefore the parallels between (52) and (53) can be paraphrased more directly. The wording here is adopted to highlight the difference to example (54) below.

‘Then Niko thought about it for a while, and then he took one of the pills.’
 d”. #Dann hat der Niko halt ne Weile überlegt, und dann hat er die Pille(n)
 then has the Niko PRT a while reflected and then has he the pill(s)
 genommen.
 taken
 ‘Then Niko thought about it for a while, and then he took the pill(s).’

This time the discourse introduces two kinds of pills again, a red and a blue one, but the difference is irrelevant to the decision the story addresses and its implications. Both the gesture in (54d) and the paraphrase (54d’) are felicitous. The infelicity of (54d”) with the plural *Pillen* (‘pills’) is due to the previously established difference between morning and evening medication. The singular version with *Pille* (‘pill’) contrasting with the felicitous gesture in (54d) suggests that the gesture will take the maximal plurality consisting of the compatible antecedent entities as one antecedent and then choose a subgroup of it, in line with the judgment on the paraphrase (54d’).

Maximal antecedents are formed irrespectively of whether the salient entities are discourse referents or not. The felicity of (54d), which persists even if the discourse referent possibly evoked or bridged by the DP *die Pillen* (‘the pills’) in *wenn er die Pillen anfängt zu nehmen* (‘if he starts taking the pills’) in (54b) is avoided by instead using the wording *wenn er die Behandlung anfängt* (‘if he starts the (process of) treatment’). Definite descriptions and definite gestures compile the maximal antecedent irrespectively of the predicates suitable discourse referents were introduced with, like the blue and the red pill can be compiled by means of a phrase such as *the pills*¹⁸ or by *small – into – mouth*, and (54d”) would be felicitous if (54a) would claim the need to take both at the same time instead of differentiating them by day time. This behavior matches Sharvy’s (1980) observations on maximality¹⁹ perfectly. The usual tests for maximality of antecedents by trying to make statements about subgroups of a maximal entity derivable from the extension of a predicate cannot be applied under the partitive approach, but by exploiting the vagueness of maximal antecedents and by relevance phenomena maximality is still detectable.

Assuming a gesture $G(i)$ resolves to an event predication $event(e)$ and an entity predication $entity(x)$ with a thematic role $role(e,y)$, writing a maximal antecedent as $\mu x[P(x)]$ and representing the partitive relation by \leq , the definite use of $G(i)$ can be written $\exists e \exists y[event(e) \wedge y \leq \mu x[entity(x)] \wedge role(e,y)]$. This is also compatible with counterexamples to maximality raised by von Stechow, Fox & Iatridou (2014). All

¹⁸This is not only due to the modifier adjective with a common noun predicate, since a summarizing description such as *the snacks* can be used after a sentence like *I bought chips and small pretzels*.

¹⁹The status of maximality is disputed to some degree. See Schwarz (2013) for an overview and experimental evidence.

counterexamples provided by them involve predicates *within* the scope of the definite operator where non-maximal entities are most informative, e.g. *the amount of money John can live on* (von Fintel et al., 2014, p. 169). This means that in the formula presented for a gesture $G(i)$ the antecedent $\mu x[entity(x)]$ has to be most informative with respect to $entity(x)$, not with respect to the whole formula. Thus for gestures it makes no difference whether μ represents maximality or maximal informativeness. The formula itself however becomes increasingly vague with a larger antecedent since the partitive can choose among more potential subgroups of the antecedent, allowing for relevance felicity issues as demonstrated above.

The claims of this section are summarized as formulated below:

Maximality of plural definite descriptions

If in a plural definite description *the NPs* the predicate of *NP* is satisfied by more than one (possibly plural) entity, *the NPs* denotes all of these satisfying entities as one plurality. Therefore the uniqueness constraint found with singular definite descriptions cannot be observed with the presence of additional entities satisfying the predicate of *NP*.

Insufficiently informative maximality

If a definite denotation or use of an expression involves a maximal plurality and the context requires partitions of this maximal plurality by relevance conditions, maximal pluralities invoke infelicity.

Maximality of definite iconic co-speech gestures

An iconic co-speech gesture in definite use or denotation²⁰ evokes a subgroup of a maximal antecedent plurality. *Insufficiently informative maximality* applies to the gesture's antecedent.

The issue of relevance, which governs both the felicitous usage of the iconic gesture and the felicitous usage of the definite descriptions in (52)-(54), is highly vague and informal. While most of the vague nature of this notion of relevance will remain unclear, a sketch of an approach to tackle it will be presented in Section 4.3.5 below. First however the nature of the maximality operation will be explored further by establishing if the standard of maximality with gestures is the form predicate or the subtype.

²⁰The thesis is agnostic about the issue of whether definiteness is semantically meaningful or just a pragmatic use of the gesture.

4.3.4 Maximal by form or by subtype

It was established that different entities which are distinguished in a contextually irrelevant way may be grouped together as one antecedent. By the previous examples it was also established that uniqueness violations may occur if one possible subtype of the gesture (e.g. *pill(x)* as subtype for *small – into – mouth*) is satisfied by different entities where the distinction is relevant but not retrievable by the gestural form itself (in this example, color information). But can infelicity arise even if the gesture form has two distinct subtypes that are satisfied by a contextually unique entity each? An affirmative answer would mean that maximality does not hold for rather specific interpretations or subtypes of a gesture form, but for the way from the abstract form to the entities themselves. This affirmative answer is argued for below.

Consider the following example, where building on the previous examples involving *x* *made up their [mind]* _{*small–into–mouth*} four kinds of small objects are introduced:

- (55) a. Auf Kathis Geburtstag gabs diese kleinen Brezeln, Chips, und
at Kathi's birthday there_were those small pretzels chips and
Gummibärchen und so was.
gummy_bears and such something
'At Kathi's birthday they had those small pretzels, chips, and gummy bears
and stuff like that.'
- b. Und weil Tom ja momentan diese Magenprobleme hat
and because Tom PRT currently these stomach_problems has
'And because Tom currently has those stomach problems'
- c. und er auch nach jedem Essen diese Pillen nehmen muss,
and he also after every meal these pills take must
'and he also has to take those pills after every meal'
- d. war er sich erst nich sicher, ob er sich auch was nehmen
was he himself first not sure if he himself too something take
will.
wants
'he wasn't sure if he too wanted to take anything.'
- e. Aber dann hat er sich [entschieden] _{*small–into–mouth*} .
but then has he himself decided
'But then he made up his mind.'

Discourse (55) is felicitous. Unlike with the examples (52)-(54), there is no common nominal predicate for (55) that summarizes both kinds of chips, pretzels, gummy bears and pills. One might paraphrase the common denominator of the first three objects as *small snacks*, but trying to paraphrase the three kinds of snacks *and* the pills with one DP can only result in highly odd formulations such as *some of the small things to put in his mouth*. This basically amounts to the gesture form predicate itself once an event

interpretation is excluded in favor of an entity interpretation. Again the choice of whether Tom is depicted ingesting his pills or eating the snacks is irrelevant to the felicity for reasons of relevance: The consequences are the same, and one contextually implies the other. Likewise it is irrelevant which kind of chips Tom ate, if he ate any at all. Starting from this observation, (55c) - (55e) is now replaced with (56) below:

- (56) a. und er seine Pillen ja nur nehmen kann, wenn er zwei Stunden nichts
 and he his pills PRT only take can if he two hours nothing
 gegessen hat,
 eaten has
 ‘and he can only take his pills if he hasn’t eaten anything for two hours’
- b. war er sich erst nich sicher, ob er lieber was essen oder lieber
 was he himself first not sure if he rather something eat or rather
 seine Pillen nehmen will.
 his pills take wants
 ‘he wasn’t sure if he rather wanted to eat something or rather would take his
 pills.’
- c. #Aber dann hat er sich [entschieden] *small–into–mouth* .
 but then has he himself decided
 ‘But then he made up his mind.’

Now again the gesture is made unsatisfactory by the maximality operation of all entities compatible with the gesture by relevance factors. This means that the subtype *pill(x)* conflicts with something like *small – food(x)* or *small – snack(x)* in resolution, although deciding on either of these interpretations would lead to a unique maximal antecedent that is informative enough for the relevant question how Tom decided. If the unique maximal antecedent of a subtype does not fulfill the maximality requirement, but without relevant distinctions in (55) the gesture can be felicitous, this means that the unresolved form predicate *small – into – mouth(i)* is the denotation of the gesture and the maximal antecedent will be all entities satisfying this predicate. This implies that interpreting gestures relies on quite abstract concepts which may be represented as types located in a high position of Lascarides & Stone’s (2006, 2009) gesture type hierarchy, and that paraphrasing gestures with everyday speech predicates has to be taken with some caution. While the maximality operation of grouping red and blue pills together as one joint plurality can be easily paraphrased by simply omitting the adjectival contribution by saying *the pill(s)*, in some cases an abstract semantic representation such as $\mu x[\textit{small – into – mouth}(x)]$ is inevitable. If the form predicates are taken to be the lexical meaning of the iconic gesture and subtypes are only obtained by the contextual possibility to exclude certain interpretations, this fits well into the picture of a definite description: There is exactly one maximal *x* such that the predicate is true for *x*, the predicate is the gesture’s semantics, which is *small – into – mouth*. Note that this thesis did not devote much attention to the event reading component of gestures. Leaving open

the question of definiteness effects on event resolutions but also providing a concrete formula, it will be assumed that the form predicate splits into two versions resolved for argument type, *small – into – mouth*(*e*) and *small – into – mouth*(*x*), and that only the latter is within the scope unique maximality operator μ .

The next section returns to the problem of the notion of relevance invoked for the present purposes, offering a fragmentary direction of research.

4.3.5 Approaching formal relevance

According to Grice (1975) a relevance requirement pertains to the fundamental rules cooperative speakers try to observe, and certain implicatures can be drawn from apparent violations. If however there is no implicature that can be retrieved from a speaker under the assumption that conversational cooperation is maintained despite of the apparent violation of the Maxim of Relevance, the utterance should be considered odd. Stating it informally, whatever contribution the speaker makes, it should not just be a random assertion contrary to the established goal. Pinning down the goal of a conversation is however a similarly vague endeavor with a lot of leeway.

In the Question-under-Discussion (QUD) framework as spelled out by Roberts (2012) the issue of relevance is partly formalized in discourse moves. The interlocutors try to answer the Big Question *What is the way things are?* by pruning Stalnaker's (1978) Common Ground as much as possible, and they do so by establishing subquestions and answering them as well as possible (Roberts, 2012, pp. 6:4-6:6). In the semantics for questions proposed, the *q-alternatives* of a clause α , i.e., the propositions that can serve to provide answers to a potential question, are defined as all propositions p where for each variable introduced by a question word such as *who* or *what* there is some adequately typed semantic object which can be inserted for the variable. Formally the definition which covers both questions and assertions is written as follows:

(57) The *q-alternatives* corresponding to utterance of a clause α :

$$q-alt(\alpha) = \{p : \exists u^{i-1}, \dots, u^{i-n} \in D[p = |\beta|(u^{i-1}) \dots (u^{i-n})]\}$$

where α has the logical form $wh_{i-1}, \dots, wh_{i-n}(\beta)$, with $\{wh_{i-1}, \dots, wh_{i-n}\}$ the (possibly empty) set of *wh*-elements in α , and

where D is the domain of the model for the language, suitably sortally restricted, e.g., to humans for *who*, nonhumans for *what* (Roberts, 2012, p. 6:10)

For example if α is *Who did Mary invite?* and *who* is assumed to abstract over type e entities the logical form is $?(\text{who}(\lambda x.\text{Mary invited } x))$ (Roberts, 2012, p. 6:11). Then $|\beta| = \lambda x.\text{Mary invited } x$, and assuming that $D = \{\text{Mary}, \text{Alice}, \text{Grace}\}$ and a non-reflexive *invite* predicate, $q - \text{alt}(\alpha)$ equals the set of propositions $\{\text{Mary invited Alice}, \text{Mary invited Grace}\}$ (Roberts, 2012, pp. 6:11-6:12). This holds irrespectively of whether Mary actually invited Grace and/or Alice. By this semantics of the alternative set of a question a *complete answer* is defined as a proposition which for each $p \in q - \text{alt}(\alpha)$ contextually entails either p or $\neg p$ (Roberts, 2012, p. 6:11). Therefore potential complete answers to *Who did Mary invite?* would be *Mary invited Grace and Alice*, *Mary invited Alice but not Grace*, *Mary invited neither Grace nor Alice* etc. *Partial answers* in turn only entail p or $\neg p$ for *some* $p \in q - \text{alt}(\alpha)$, e.g. *I know that Mary invited Grace, but I don't know about Alice*.

A lot of the predictive power of the QUD framework stems from the idea of complete and partial answers and the notions derived from it, contextual entailment of questions and answers. Given a question under discussion, which is the last element of the stack of accepted questions (Roberts, 2012, p. 6:38), any following question has to be such that a complete answer entails a partial answer to the current question under discussion to be acceptable and to potentially become the new question under discussion (Roberts, 2012, pp. 6:16-6:18, 6:38). For example, a complete answer to *What did Hilary eat?* lists all things Hilary ate and all she did not eat, which is a partial answer to *Who ate what?*, so *What did Hilary eat?* may be accepted as a relevant question and the new question under discussion in order to pursue a strategy to answer the current question under discussion *Who ate what?*. Likewise any assertion has to provide a partial answer to the current question under discussion in order to be relevant (Roberts, 2012, p. 6:21). This relevance constraint severely restricts which questions may be raised and which assertions are considered relevant given a certain topic the discourse is intended to address.

Unfortunately using this formulation to explain all relevance phenomena is too restrictive for some storytelling monologues. Consider a story like (58) below:

- (58)
- a. So the doctor gave Niko two pills, a red one and a blue one.
 - b. And he said if he started to take the blue ones, he'd live another 10 years, but most of the time in the hospital,
 - c. and if he took the red ones he'd have one more year but he'd be fit until short before the end.
 - d. Then Niko thought about it for a while,
 - e. and he realized he was incapable of making that decision on his own.
 - f. So he asked his sister for advice,

- g. and he also asked two friends about their opinion.
- h. It took him quite some time,
- i. but in the end he chose the blue ones.

Trying to capture relevance in a monologue full of assertions requires inferring some implicit questions under discussion. The most general question to begin with would be *What happened in the story the speaker intends to tell?*. One could argue that after (58a) a subquestion is raised, *What is the difference between the blue and the red pill?* relative to the story setting. This is answered by (58b) and (58c). Now however the initial observation motivating the search for a framework of relevance was the idea that the addressee wants to know which pill Niko took. After the question what the difference between the pills was is popped off the QUD stack the question about Niko's choice is a valid subquestion to the *What happened...?* general question. Now however the framework is predicted to only allow subquestions and partial answers to the question which pills Niko chose. From (58d) to (58i) no subquestion to this is raised, and the answer is suspended until (58i). Instead one might argue that (58d) gave a partial answer to the *What happened...?* question and (58e) raised the question *How did he deal with the problem of not being able to choose on his own?*. This question is addressed at least by (58f) and (58g), and this implies that a raised question like the one for the choice of the pills can be ignored for some time. New assertions and questions are added and removed without fulfilling any entailment requirement to the topmost question on the QUD stack²¹.

However if the entailment requirements are not considered obligatory the QUD framework provides some means to deal with the relevance issues discussed above. Instead of choosing a stack with an ordering of questions, QUD could be viewed as an ordinary set. Then in certain settings, among these the storytelling monologues this thesis concentrates on, there is an active constraint that the QUD set ought to be empty. This would explain why (52) (the context where the choice between red and blue pill is relevant) is infelicitous when ended with the *small – into – mouth* gesture without indicating the color. It would also explain why (54) (a context where the choice is between taking the pills or not taking them) is felicitous with the gesture, but infelicitous without. In each infelicitous case the question how Niko chose is unanswered, in the felicitous cases the answer is provided at least indirectly²². The drawback of this is that nothing replaces the originally easily falsifiable and predictive QUD component of entailment requirements. However this modified QUD framework can still contribute some components to theories of discourse: It comes with a semantics for questions, and it can capture relevance by adding and removing questions to a set.

²¹This observation is also made by Hunter & Abrusán (2015).

²²Indirectly because iconic co-speech gestures are considered not-at-issue (Ebert & Ebert, 2014; Schlenker, 2018a).

Hunter & Abrusán (2015) discuss possible mappings between the SDRT approach and the QUD framework. They observe that certain rhetorical relations are hardly formulated as answering QUD questions that do not make use of formulations that paraphrase the SDRT relations. For example *Contrast* relations like in *Pat came to the party but Mel didn't* could only be paraphrased with QUD style questions of the form *What does x contrast with?* and *Narration* by *What happened next?*. This has the effect that SDRT diagnostics to infer rhetorical relations are necessary to infer the correct questions in QUD as well. For example, suppose the question under discussion is *What happened to x yesterday evening?*, and a sentence *x's car broke down* is already inferred to answer this question at least partially. In order to infer correctly that *x's car broke down* is not the complete answer and that an utterance *and x realized their brother couldn't pick them up* is relevant to the same question, it has to be inferred that both sentences elaborate on the same topic in a temporal successive way of narrating, that is, *Narration* has to be inferred.

This means SDRT seems to have important advantages over QUD²³, but the relevant questions in earlier examples have no straightforward equivalent in SDRT structure. It is suggested here that SDRT is supplemented with a QUD set which can store questions such as *Did Niko take the red or the blue pills?* but that the QUD set does not have to carry the burden of being a standalone discourse model. Having chosen the framework, it is now possible to investigate means of how to add the correct questions to the QUD set that are not equivalent with one rhetorical relation or an SDRS.

Consider example (58) again. There (58b) and (58c) both are conditional constructions. Therefore they can be analyzed in SDRT as of the form *Consequence*($\pi_{blue-pills}$, $\pi_{10-years}$) and *Consequence*($\pi_{red-pills}$, $\pi_{1-fit-year}$). One possibility is to detect multiple possible relevant outcomes by a structure where *Contrast*(α , β)²⁴ relation is present and both α and β contain a *Consequence* relation, the SDRT equivalent of conditional constructions. But choices can also be diagnosed with not seemingly mutually exclusive conditionals like in (59):

(59) (Speaker narrating how they designed a character in a game.)

a. So you can choose two of three advantages that help you during the game.

²³Onea Gáspár (2016, p. 356) however advocates a view on his QUD-based framework according to which discourse trees built from potential questions are either equally or more informative than SDRSs. While it is true that the close relationship between questions and rhetorical relations like with *What does x contrast with?* and *Contrast*(α , β) is no theoretical problem, SDRT's Glue Logic is necessary for the inference of the right relations or questions.

²⁴Asher & Lascarides (2003, p. 465) suggest that two discourse segments are in a *Contrast* relation if one allows to infer the negation of the other. Inferring this contradiction of the clauses is regulated by a theory Asher, Hardt & Busquets (1997) put forward. *contrast* is often signaled by the word *but*.

- b. If you make your character brave, they will be able to handle difficulties.
- c. If you give them good empathy, they will be better in social situations.
- d. If you make them wealthy they will be able to afford stuff.
- e. Well, I chose the empathy thing and the wealth thing.

But as (60) shows, conditionals have to be resolved even when they are not about choices. In (60) as part of an appropriate story about some party the speaker went to and centered about the complicated relationship with Sam (60b) is not optional:

- (60) a. If Sam had been there, that would have been really awkward.
 b. Luckily, he didn't show up.

If one opts for the plausible assumption that context (58) triggers the separate questions *Did Niko take the blue pills?* and *Did Niko take the red pills?*, it seems that all examples displaying relevance effects trigger the following pragmatic constraint:

- (61) (first attempt) If the SDRS of the whole monologue contains a condition of the form *Consequence*(α, β), the context and discourse combined have to entail either α or $\neg\alpha$ at the end of the monologue.

Another ingredient of QUD however is required to allow this inference. Sometimes a *Consequence* relation will just answer a salient question and will not trigger the constraint in (61):

- (62) a. In this case, choosing between a job and your studies boils down to one criterion.
 b. If you need money right now, you have to take the job, and if not, you should continue to study.

(62a) can be said to trigger a question (or *potential* question in terms of Onea Gáspár, 2016) *What is the criterion for choosing between a job and continuing studies?*. This is answered by (62b). The constraint (61) does not apply here since no question *Does the addressee need money right now?* seem to be necessarily raised - the addressee can go home and reflect on their situation and the conversation will not become infelicitous. However while the constraint of Roberts (2012) that each discourse move has to address the *last* question or pose a subquestion is too strong, it is certainly plausible that each discourse move has to either pose or address *some* question. Take for example (63):

- (63) a. I went on this car trip yesterday.
 b. I was passing through the woods.
 c. My motor had been making weird sounds from time to time.
 d. Then I got lost.

- d'. Then it broke down.
- e. But with the help of a passing driver I made it to the next town.

If the fourth utterance is (63d) and (63d') is omitted, the utterance of (63c) is irrelevant. If however (63d') is used instead of (63d), (63c) is relevant. This can be explained by the idea that (63c) does not answer the *What happened during the car trip?* question in a more central way than expository clauses like *The birds were chirping* and thus cannot be regarded as partially answering a relevant question²⁵. However by the Gricean (1975) cooperative principle one will infer a sensible contribution to the story. The general assumption to background information such as (63c) (which can be classified as an SDRT *Background* relation) will then be that it has consequences on the course of narrated events. Therefore (63c) can be said to raise the question *What consequence did the behavior of the motor have?* rather than addressing an existing question. While the research of relevance of such narrations is a potentially large research field, the following constraint will be proposed:

- (64) (tentative) For each assertion a , if a does not partially answer an open question, it raises a new question.

This can be supplemented with the observations about certain rhetorical relations:

- (65) a. (tentative, revised attempt) If the SDRS of the whole monologue contains an embedded SDRS π with a condition of the form $Consequence(\alpha, \beta)$, and π does not (partially or completely) answer an open question, the context and discourse combined have to entail either α or $\neg\alpha$ at the end of the monologue.
- b. (tentative) If the SDRS of the whole monologue contains a backgrounded²⁶ (S)DRS π , and π does not (partially or completely) answer an open question, the context and discourse combined have to entail $Result(\pi, \beta)$ or $Narration(\pi, \beta)$ at the end of the monologue, for some foregrounded (S)DRS β .

These constraints may be regulated by a QUD-like set of relevant questions or by a special structural topic in the SDRT model. The structure of the latter however would have to be motivated by possible attachment points of new discourse segments and binding facts. Note also that discourse structural constraints may be far from exhaustive in accounting for relevance violations not directly attributable to SDRT constraints on coherence, but

²⁵Admittedly there seems to be another component of relevance which excludes to answer *What happened?* questions with chirping birds, which are not at all central to the main plot but technically one possible answer. This ought to be explored further, but exceeds the scope of this thesis.

²⁶This abstracts away from the correct analysis of *Background* relations (see Asher, Aurnague, Bras, Sablayrolles & Vieu, 1995; Asher et al., 2007, for some discussion).

it is for the examples discussed so far. More complicated factors can be based on world knowledge. For example, consider context (66):

- (66)
- a. Das war in diesem furchtbar trockenen Sommer.
that was in that horribly dry summer
'That was during that horribly dry summer.'
 - b. Da war dieser Typ, dem ich im Wald begegnet bin, der mich
there was this guy who I in_the forest met AUX who me
wahllos angepöbelt hat.
randomly yelled_at has
'There was that guy I came across in the forest who just yelled at me
randomly.'
 - c. Später hab ich gesehen wie er eine geraucht hat,
later have I seen how he one smoked has
'later I saw him smoking.'
 - d. sich nen Kaugummi in den Mund gesteckt hat
himself a chewing_gum into the mouth put has
'putting a chewing gum into his mouth'
 - e. und dann meinte die Gegend [verschandeln]_{toss-away} zu müssen.
and then thought the area make_ugly to have_to
'and then apparently thinking he had to trash the area.'

The gesture of tossing away something may trigger a question of relevance to some attentive listeners (although not all): Did the person throw his possibly still glowing cigarette into the forest in the dry summer, causing a fire hazard? Such effects may be exploited or accidental, more or less obvious, and related to world knowledge about fire, dry summers and so forth. Such effects are too complicated and manifold (and most likely require psychological experimental work for explanation as well), so simple linguistic approaches may not capture them. But the relevance factors attributed to the constraints (64)-(65) are consistently and reliably triggered by linguistically diagnosable factors.

The crucial point of this discussion about relevant questions is that certain structures or conditions present in SDRT may be taken to trigger relevance effects, and that such triggers may be possibly identified and accounted for in a systematic and concise way. Constraints on discourse as such (like (64)) and those on specific rhetorical relations (like (65)) can be employed to achieve this. The ordinary unique maximal antecedents to (possibly) plural definite descriptions and gestures discussed in Section 4.3.3 only constitute a constraint in combination with open relevant questions. Relevance constraints such as (64)-(65) can plausibly be taken to be separately encoded from unique maximality, but still interacting with it. The algorithm presented in Section 4.3.2 can now be updated with the results of Sections 4.3.3 and 4.3.5:

Algorithm to satisfy the maximality and relevance requirements (to be appended to (51))

1. Form the set of *compatible* or *suitable* referents / events consisting of all referents obtained by (51). If it is empty, raise infelicity.
2. If the set of compatible referents is non-empty, form the maximal entity of all compatible entities. Do so by forming a plurality if there are more than two compatible entities.
3. Use this maximal entity as the antecedent and introduce a subgroup of it for the semantic contribution of the gesture to the concurrent speech.
4. If the set of relevant questions is non-empty after finishing the monologue (possibly because the antecedent plurality the gesture compiled is too general for the contributed subset to answer these questions, e.g. *some pills of the plurality of blue and red pills* is too general to answer *Did Niko take the blue pills?*), raise infelicity.

4.4 The definite approach in a dynamic framework

This section will discuss some aspects of how to interpret definite descriptions and iconic co-speech gestures in an SDRT framework. As already established definite descriptions do not necessarily require a discourse referent antecedent. If they are felicitous by virtue of a unique salient entity in the context, they introduce a discourse referent that can be coreferential to a subsequent pronoun:

(67) I saw [the geologist you told me about]_i today. She_i seemed very busy.

This means that an adequate interpretation of contextually given definite descriptions has to be able to treat them like existentials in terms of a newly introduced discourse referent, because contextual information alone cannot license coreferential pronouns. This discussion abstracts away from the right method of implementation of a Common Ground in SDRT, and since whether discourse referents are introduced or not is strictly regulated by triggering configurations (see Kamp & Reyle, 1993a, 1993b), they will receive special treatment by the analysis. Introducing new discourse referents in each case is sufficient and to be preferred, since in case of plurality antecedents frequently a plurality is formed by existing atomic referents.

In Lascarides & Stone's (2009, p. 422) approach there are two assignment functions: The domain of f includes all discourse referents introduced for speech, and the domain of g

in addition covers discourse referents present in gesture only. The role of *g* is to prevent pronouns in speech from being coreferential with gesture-only discourse referents. Such a restriction is desirable for pronouns, but not for definite descriptions, as can be seen in the example (68) below, where the gesture *swing – at – fist* introduces a chisel which is picked up by a subsequent definite description in speech. *swing – at – fist* is executed as the left fist standing still in front of the speaker while making quick hammering swings with an imaginary object held by the right fist at the left fist:

- (68) (Context: Two archaeologists are having a conversation. They are on an excavation site with student interns they have to supervise, Hansen is one of them.)
- a. Ey gestern is mir was richtig Ärgerliches passiert.
hey yesterday is me something really annoying happened
'So yesterday something really annoying happened to me.'
 - b. Der Hansen hat angefangen [nen Tonkrug freizulegen] _{swing-at-fist}.
the Hansen has started a clay_pot lay_bare
'Hansen started laying bare a clay pot.'
 - c. Ich bin dann nur kurz weggegangen, um mir nen Kaffee zu holen,
I am then just shortly gone_away for me a coffee to take
'I then just left for a short time to grab a coffee,'
 - d. und als ich wiedergekommen bin,
and when I returned am
'and when I came back'
 - e. war der Meißel drin in dem Krug!
was the chisel inside in the pot
'the chisel had ended up inside the pot!'

Under the assumption that the context allowed for brushing away sand to lay bare a clay pot the gesture *swing – at – fist* is necessary to not make (68e) odd. That the definite description *der Meißel* ('the chisel') is felicitous with the gesture and refers to the one Hansen used is most easily explained by the introduction of a discourse referent by the gesture picked up by the definite description. This means that definite descriptions should be evaluated with respect to the assignment function with the larger domain, *g*. By example (41) above and by uniformly analyzing the data of Lascarides & Stone (2006, 2009) this applies to definite uses of gestures as well. This data is sufficient for an implementation in SDRT. First the relevant interpretation definitions of Lascarides & Stone (2009) are provided:

- (69)
- a. Where *i* is a variable, $\langle f, g \rangle \llbracket i \rrbracket^M = f(i)$.
 - b. For a formula $P(i_1, \dots, i_n)$, $\langle f, g \rangle \llbracket P(i_1, \dots, i_n) \rrbracket^M \langle f', g' \rangle$ iff $\langle f, g \rangle = \langle f', g' \rangle$ and $\langle \langle f, g \rangle \llbracket i_1 \rrbracket^M, \dots, \langle f, g \rangle \llbracket i_n \rrbracket^M \rangle \in I(P^n)$.
 - c. $\langle f, g \rangle \llbracket \exists x \rrbracket^M \langle f', g' \rangle$ iff:

- (a) $dom(f') = dom(f) \cup \{x\}$ and $\forall \gamma \in dom(f), f'(\gamma) = f(\gamma)$ (i.e. $f \subseteq_x f'$);
- (b) $dom(g') = dom(g) \cup \{x\}$ and $\forall \gamma \in dom(g), g'(\gamma) = g(\gamma)$ (i.e. $g \subseteq_x g'$);
- (c) $f'(x) = g'(x)$
- d. $\langle f, g \rangle \llbracket [\mathcal{G}](\varphi) \rrbracket^M \langle f', g' \rangle$ iff $f = f'$ and $\exists g''$ such that $\langle g, g \rangle \llbracket \varphi \rrbracket^M \langle g'', g' \rangle$.

In these definitions felicity is encoded by successful transitions from one assignment pair to another. (69d) interprets gestures with an operator that ensures that other definitions like (69a) and (69c) are interpreted by substituting g for f . For example, (69a) would require a variable to be interpreted with respect to the first of the two assignments, normally the smaller assignment f , but if it is within the scope of $[\mathcal{G}]$, it will be interpreted with respect to the first assignment which is now g since (69d) substitutes $\langle g, g \rangle$ for $\langle f, g \rangle$. If $\exists x$ introduces a discourse referent, it will normally be introduced to both f and g , but with a $[\mathcal{G}]$ operator the condition $f = f'$ of (69d) and the substitution of $\langle g, g \rangle$ as the starting assignment pair will restrict this introduction to g .

For example, assume that the domain of both f and g initially is $\{\}$, and then a discourse referent x is introduced by *There was a cat* with the semantics $\exists x[cat(x)]^{27}$. Then (69c) will ensure that both $dom(f)$ and $dom(g)$ will be $\{x\}$ afterwards, and (69b) will ensure that whatever entity $\langle f, g \rangle \llbracket x \rrbracket^M$ denotes, it should be in the set of cats $I(cat)$ according to the model's information about which entities are cats and which are not. By (69a) we know that $\langle f, g \rangle \llbracket x \rrbracket^M$ denotes $f(x)$ which is defined and points towards the actual cat.

Compare this with a gesture *handle – to – hips* as in (41) above, where the semantics can be summarized as $[\mathcal{G}](\exists x, y, z[dagger(x) \wedge hips(y) \wedge z = protagonist \wedge place(z, x, y)])$, where *protagonist* is the discourse referent from previous discourse for the protagonist of the movie. Ignoring everything else than the portion $[\mathcal{G}](\exists x[dagger(x)])$, first (69d) will ensure that $dom(f)$ will not be expanded when f transitions to f' , and that all evaluations of $\exists x[dagger(x)]$ are done with respect to a starting assignment pair $\langle g, g \rangle$. Then (69c) will introduce x to the domain of both assignments and ensure they receive the same interpretation, but this is a trivial step since both assignments are the same assignment g inside an application of (69d). After introducing x to g $dagger(x)$ can be evaluated by (69b) with respect to the updated version of g and is defined. Then it will be true that $x \in dom(g)$ but not that $x \in dom(f)$. If after such a gestural introduction one tries to pick up the same x again in the style of Lascarides & Stone's (2009) discourse analysis (23) above, again $\langle g, g \rangle$ is substituted for interpretation and $g(x)$ is defined, but if one tries to refer to a gesture with a pronoun like in the minimal pair (45), no substitution of $\langle g, g \rangle$ will take place, and since $x \in dom(f)$ is false, it will be uninterpretable by rule (69a).

²⁷This abstracts away from events.

While for definite formulae $\mu x[\varphi[x]]$, where $\varphi[x]$ is a formula containing an occurrence of x ²⁸, introduction of pluralities should not be restricted to g without a $[\mathcal{G}]$ operator, since they can be picked up by speech pronouns (see (67) above), the contents within the μ operator should be able to draw from $dom(g)$ as demonstrated by (68). This means a definition should copy the substitution of $\langle g, g \rangle$ of (69d) without restricting introduction. The interpretation of definiteness irrespectively of modality proposed here will therefore be as follows:

- (70) **Dynamic interpretation of $\mu x[\varphi[x]]$**
 $\langle f, g \rangle \llbracket \mu x[\varphi[x]] \rrbracket^M \langle f', g' \rangle$, where $\varphi[x]$ is a formula containing an occurrence of x ,
iff:
1. $\langle f, g \rangle \llbracket \exists x[\varphi[x]] \rrbracket^M \langle f', g' \rangle$; and
2. $g(x)$ is the unique maximal entity formed of
(i) a , where a is the unique maximal entity of all entities $g(y)$ such that there is a discourse referent y and an assignment g'' such that $\langle g, g \rangle \llbracket \varphi[y] \rrbracket^M \langle g'', g' \rangle$; and of
(ii) all salient entities b in the Common Ground such that $b \in I(P^n)$.

This definition will make sure that definite descriptions will introduce a plurality which can draw from previously defined discourse referents irrespectively of modality. As it stands it will introduce the discourse referent for maximal entity x as a potential antecedent to both f and g by first producing the maximal entity a from all defined discourse referents such that the model can interpret a predicate with y as true. Then it will add discourse referents to the interpretation of x which are not defined in $dom(g)$ but which render the predicate true according to the model of the outer world by clause (ii). If a gesture is used as a definite, the formula will be $[\mathcal{G}](\mu x[\varphi[x]])$ and the $[\mathcal{G}]$ will prevent pronouns picking up the newly introduced referent.

As for the accessibility of antecedents, the definitions above allow accessing all defined discourse referents without structural conditions, which is desirable in light of the data presented in Section 3. f and g themselves are not restricted by pronominal constraints.

Then again note that the analysis of definite uses of gestures does not only include the construction of an antecedent but also the introduction of a subgroup. Does this discourse referent have effects on the interpretation of subsequent formulae? The answer is no. Since μ will introduce the maximal entity fulfilling the form predicate of a gesture, any definite description with a predicate equivalent to that form predicate

²⁸ φ can be multiple predicates in expressions such as *green cat* ($cat(x) \wedge green(x)$) and it need not be the only argument with relational nouns such as *John's mother* ($\mu x[mother(x, john)]$).

(whether in speech or another gesture) will automatically compile the same maximal entity. Since this compilation process draws from all defined discourse referents in $dom(g)$ irrespectively of structure and both are introduced within the same DRS, there is no structural configuration where the partitive subgroup is defined but not the maximal antecedent. Furthermore, since pronouns in speech cannot draw from gestural discourse referents and only definite descriptions can, there will never be an ambiguity between a non-maximal subgroup and a maximal entity. The introduction of the subgroup will always be vacuous. This is not necessarily an asset, since an important criterion of science is falsifiability. Note however that the partitive approach presented in this thesis *can* be falsified by maximal antecedents compiled incorrectly and thus under- or overgenerating felicitous gestures (whether by relevant distinctions as in Section 4.3.3 or by failing gestures as will be discussed below in Section 5.2) or by theoretically conceivable examples only allowing for the subgroup to be maximal as well, it just cannot be falsified by the *dynamic* effects of the subgroup.

4.5 Applying Donnellan's (1966) referential-attributive distinction

A notable issue with definite descriptions is observed and discussed by Donnellan (1966), who uses the sentence *Smith's murderer is insane* (Donnellan, 1966, p. 285) to illustrate the distinction. If a person of the name Jones was witnessed to act insane and to be the murderer of Smith, *Jones is insane* and *Smith's murderer is insane* express the exact same thing and the descriptive content of being a murderer of Smith is irrelevant, which Donnellan (1966) calls the *referential* use. If the sentence *Smith's murderer is insane* is uttered because the speaker saw a horrible crime scene but is unaware that Jones is Smith's murderer, it would not make sense to say *Jones is insane*. Then statements about the entity pointed out by *Smith's murderer* could only be made by virtue of the descriptive content that the entity is a murderer of Smith. This use is called the *attributive* use by Donnellan (1966).

Unfortunately it is not entirely clear what exactly the difference between the two uses is, since several authors have argued that none or some of the criteria Donnellan (1966) mentions are reliable (see e.g. Bach, 2004 or Abbott, 2010, pp. 140-152 for some discussion). For this reason this short discussion of which uses gestures can convey has to remain informal and potentially inconclusive, yet examples in the style of Donnellan (1966) can be constructed. A promising candidate for a referential use is (71), where the gesture *holding – back – forth* is a fist moving repeatedly back and forth as if one do when using a small saw, while the gesture *flat – back – forth* is the same movement but with a vertical flat hand instead of a fist. These two different gestures convey the same image of a saw by virtue of Müller's (1998, pp. 115-123) different kinds of iconic

portrayal, in this case the object-manipulating and the object-representing hand. However both are acceptable for the same entity, Linda's handsaw.

- (71) a. Linda will jetzt Holzskulpturen machen und hat sich so ne kleine
Linda wants now wood_sculptures make and has herself such a small
Handsäge gekauft.
handsaw bought
'Linda decided she's going to make wooden sculptures and bought a small
handsaw.'
- b. Dann hat se aber gemerkt, dass se mit der noch nich wirklich umgehen
than has she but realized that she with it still not really handle
kann.
can
'But then she realized she isn't very skillful with it yet.'
- c. Also hat se sich einfach wirklich 10 Kilo Holz gekauft und den
so has she herself just really 10 kilograms wood bought and the
[ganzen Tag geübt] _{holding-back-forth/flat-back-forth} ·
whole day practiced
'So she really just bought 10 kilograms of wood and practiced the whole
day.'

A potential candidate for an attributive use is (72), where *fist – into – flat* is the movement of hitting the left flat palm with the right fist.

- (72) a. Ey schau ma, da oben hat irgendwer mit irgendwas n Fenster
hey look PRT there up has someone with something a window
eingeworfen!
thrown_in
'Hey look, up there someone actually threw in a window with something!'
- (i) Alter haste ne Ahnung wie dermaßen stabil diese Fenster
dude have_you an idea how very stable these windows
sind?
are
'Dude do you have any idea how incredibly stable these windows
are?'
- (ii) Das Ding muss [hunderte] _{fist-into-flat} Kilo abbekommen
that thing must hundreds_of kilograms gotten
haben!
have
'That thing must have been hit by hundreds of kilograms!'

In (72) the fist represents the unknown heavy object, the flat hand the window. Classifying this gesture as an attributive use may be weakened by the fact that it is accompanied by the NP *hunderte Kilo* ('hundreds of kilograms') and the gesture may be attributed to representing the discourse referent of *hunderte Kilo*. This could however only be a

metaphoric use of *hundert Kilo* since a quantity of a physical unit does not have a shape. For plural antecedents the judgments become even less clear-cut, since there are cases where the exact composition of the antecedent plurality may be unknown but the gestures are interchangeable:

- (73)
- a. Stefan hat ja unzählige Handsägen in seinem Schuppen.
 Stefan has PRT countless handsaws in his shed
 ‘You know Stefan has countless handsaws in his shed.’
 - b. Ich hab keine, und ich will versuchen, damit Holzskulpturen zu machen.
 I have none and I want try with_that wood_sculptures to make
 ‘I have none, and I want to try to make wooden sculptures with them.’
 - c. Ich werd Stefan mal fragen, ob ich bei ihm [üben kann].
 I will Stefan once ask if I at him practice kann]*holding-back-forth/flat-back-forth* ·
 can
 ‘I’m gonna ask Stefan if I can practice at his place.’

Here both gestures take Stefan’s handsaws as an antecedent, which indicates a referential use. However the speaker is not necessarily aware of which saws make up the totality of the antecedent or even of the antecedent’s cardinality - they may have glanced over six saws, while Stefan has 43. This seems to indicate an attributive use as paraphrasable by *whatever saws Stefan has*. Note however that interchangeable definite descriptions in speech behave the same:

- (74)
- a. Ich werd Stefan mal fragen, ob ich mit seinen Sägen üben kann.
 I will Stefan once ask if I with his saws practice can
 ‘I’m gonna ask Stefan if I can practice with his saws.’
 - b. Ich werd Stefan mal fragen, ob ich mit seiner Sammlung üben kann.
 I will Stefan once ask if I with his collection practice can
 ‘I’m gonna ask Stefan if I can practice with his collection.’

Unfortunately the difficulties of pinning down the exact criterion for distinguishing attributive and referential uses do not allow to turn these examples into a compelling argument. This intuitive treatment however seems to suggest further shared properties between iconic co-speech gestures and definite descriptions in speech.

5 Gestural behavior unexplained by a definiteness account

While the previous section presented phenomena and analyses that can be captured well under a specific analysis involving definiteness, two instances of puzzling semantic-pragmatic behavior of iconic co-speech gestures remain a challenge.

5.1 Representative gestures

Sometimes the form of a gesture will allude to a combination of events and/or entities, among which not all of them satisfy the form predicate of the gesture. This holds for the proposed categories of indefinite and definite uses alike. Consider the following example, which so far only involves bridging but no salient antecedent:

- (75) (Context: The speaker and a bystander of the conversation observed a group of woodcutters cutting down many sick trees in a forest. The woodcutters used axes and chainsaws to cut down the trees. The addressee does not know about these events, the speaker is summarizing them.)
- a. Als wir beide heut im Wald waren, warn da n paar Holzfäller
when we both today in_the forest were were there a some woodcutters
unterwegs.
around
'When we two were in the forest today some woodcutters were around.'
- b. Die ham [die kranken Bäume gefällt]_{holding-slow/chop-twice} .
they have the sick trees cut
'They cut down the sick trees.'
- b'.??Die ham die kranken Bäume mit Kettensägen / mit Äxten gefällt.
they have the sick trees with chainsaws / with axes cut
'They cut down the sick trees with chainsaws / with axes.'

Here *holding – slow* indicates a gesture of both hands holding some object and moving it slowly horizontally from right to left, while *chop – twice* indicates a rapid horizontal movement of the right hand holding some object from right to left, followed by a less abrupt movement back to the right, executed twice. The notation in (75b) is meant to express that the discourse is felicitous with either *holding – slow* or *chop – twice* executed. The choice of gesture seems to be irrelevant. The observer may however object to the failing paraphrase (75b') since they may feel it conveys the idea that only the instruments mentioned explicitly (chainsaws or axes, but not both) were used. Hence the gesture seems to make a non-exhaustive contribution. The same applies for definite uses:

- (76) (Same context as (75).)

- a. Als wir beide heute im Wald waren, waren da ein paar Holzfäller
 when we both today in_the forest were were there a some woodcutters
 mit Äxten und Kettensägen unterwegs.
 with axes and chainsaws around
 ‘When we two were in the forest today some woodcutters with axes and
 chainsaws were around.’
- b. Die ham [die kranken Bäume gefällt] *holding-slow/chop-twice* .
 they have the sick trees cut
 ‘They cut down the sick trees.’
- b’??Die ham die kranken Bäume mit den Kettensägen gefällt.
 they have the sick trees with the chainsaws cut
 ‘They cut down the sick trees with the chainsaws.’
- b’’??Die ham die kranken Bäume mit den Äxten gefällt.
 they have the sick trees with the axes cut
 ‘They cut down the sick trees with the axes.’

Note that the judgments change when the context is altered. If (76) also involved contextual information that using axes was allowed, but using chainsaws was forbidden due to a lack of training, and the speaker had been asked by their supervisor whether the woodcutters had worked according to regulations, the observer may find the usage of *chop – twice* insincere or odd. If additionally the woodcutters were part of some rehabilitation program for especially disobedient criminal youth where the usage of permitted instruments by even only some participants was a huge success, the gesture *holding – slow* could be considered insincere or odd. The common denominator of both extended contexts with the original (76) is that the gesture may not be unrepresentative or misleading for the event according to contextual standards, but otherwise need not be exhaustive.

One popular way to analyze non-exhaustivity is the scalar implicature theory of Horn (1976). Along this theory one might claim that both the speech and the gesture version *denote* the non-exhaustive reading, but the speech version *con conversationally implicates* that no larger quantity was involved. This would then mean that some or all iconic co-speech gestures escape scalar exhaustivity implicatures. Although the exact definitions, tests and the pragmatic taxonomy involved with conversational implicatures are not always satisfactory (see Sadock, 1978; Grice, 1981; Scharfen, 1997), some tests allow to replicate the observations by Horn (1976) in order to establish a similarity in behavior.

The tests are most easily replicated by using contrastive focus constructions with contrasting gestures where the speech material seems redundant or contradictory without gesture, constructions discussed by Esipova (2018). First note that they are inherently

marked/degraded, but acceptable with prominent intonation under certain circumstances, as demonstrated in example (77) with the gestures from (76):

- (77) ?Ich habe keine [Bäume gefällt]_{chop–twice} , ich hab [Bäume
 I have no trees cut, I have trees
 gefällt]_{holding–slow} .
 cut
 ‘I didn’t cut down trees (with an axe), I cut down trees (with a chainsaw).’

As long as this degree of inherent degradedness is kept in mind and contexts are chosen such that both gestures can be considered representative by lack of information about asymmetric implications, Rullmann & You (2006) give an example of applying two tests to distinguish underspecification from scalar implicatures. The construction involving *um genau zu sein* (‘to be exact’) should be felicitous with underspecified items and not with scalar ones, while *wenn nicht (sogar)* (‘if not (even)’) should behave the opposite way (Horn, 1976; Rullmann & You, 2006):

- (78) a. (i) Ich habe Kinder. Um genau zu sein habe ich drei Kinder.
 I have children for exact to be have I three children
 ‘I have children. To be exact, I have three children.’
 (ii) #Ich habe zwei Kinder. Um genau zu sein habe ich drei Kinder.
 I have two children for exact to be have I three children
 ‘I have two children. To be exact, I have three children.’
 b. (i) #Alex hat Kinder, wenn nicht sogar drei.
 Alex has children if not even three
 ‘Alex has children, if not even three.’
 (ii) Alex hat zwei Kinder, wenn nicht sogar drei.
 Alex has two children if not even three
 ‘Alex has two children, if not three.’

Note that *to be exact* constructions require that the speaker knows about the correction, while *if not* constructions require them being agnostic. The gestures from (78) together with *auch* (‘also’) tend to match the behavior of scalar implicatures in equally representative contexts by being more acceptable with *wenn nicht sogar* (‘if not even’):

- (79) (Context: The woodcutters were strictly prohibited to work on Thursday because the total equipment was being counted. Axes and chainsaws are counted separately, so if some axes *and* some chainsaws had been missing during the counting, double the counting work would have to be repeated. The speaker is the watchman of the area, the addressee is the woodcutters’ supervisor who has asked in hindsight whether they had seen them going to work anyways.)
 (Assuming the speaker is certain:)

- a. ??Die ham am Donnerstag [Bäume gefällt]*chop–twice* . Um genau zu sein
they have on Thursday trees cut for exact to be
ham sie auch [Bäume gefällt]*holding–slow* .
have they also trees cut
‘They cut trees on Thursday (with axes). To be exact they also cut trees (with
chainsaws).’ (Assuming that the speaker is not sure about the chainsaws but
that they had heard something that sounded like chainsaws:)
- b. ?Die ham am Donnerstag [Bäume gefällt]*chop–twice* , wenn nicht sogar
they have on Thursday trees cut if not even
auch [Bäume gefällt]*holding–slow* .
also trees cut
‘They cut trees on Thursday (with axes), if not also cut trees (with
chainsaws).’

A scalar account which exempts gestures from scalarity therefore seems plausible to explain the lack of exhaustivity. However it does not explain the representativity requirement for non-exhaustive gestural contributions. Nakanishi & Tomioka (2004) provide a related account for the Japanese plural morpheme *-tati*, although there the notion of representativeness remains vague as well. They suggest that when *-tati* attaches to a common noun, it provides a predicate on pluralities that the common noun is representative of the plurality:

$$(80) \quad \llbracket -tati \rrbracket \in D_{\langle\langle e,t \rangle, \langle e,t \rangle\rangle} = \lambda P_{\langle e,t \rangle} . \lambda Y_e . |Y| \geq 2 \& P \text{ represents } Y \text{ (Nakanishi \& Tomioka, 2004, p. 129)}$$

In an example of a definite use combining the predicate *boy* with *-tati* and combined with the verbal predicate *be – playing* would provide that “[t]he unique Y, whose cardinality is two or more and which is represented by *boy*’, is playing” (Nakanishi & Tomioka, 2004, p. 129). Applying this to gestures is slightly more complicated due to the analysis that they introduce a subgroup of the antecedent. Let $\exists e \exists y [event(e) \wedge y \leq \mu x [entity(x)] \wedge role(e, y)]$ be the formula for definite uses of iconic co-speech gestures, where *role*(*e*, *y*) represents the thematic role of *y* in event *e*, where *entity*(*x*) is the form predicate only resolved to an entity type and *event*(*e*) is the event reading resolution as discussed briefly at the end of Section 4.3.4. Then one could demand an antecedent to which the gesture form predicate is representative²⁹, that is, the antecedent *x* would consist of atoms which satisfy *entity*, but also some which do not, but which are felicitously represented by those who satisfy *entity*. This would result in the underspecification of the antecedent. This solution is unappealing since it cannot be applied to non-definite uses. Both definite (partitive) and indefinite uses share the property of introducing a referent, whether it is a subgroup of

²⁹Alternatively, an antecedent to which the entities established by the definite resolution algorithm for iconic co-speech gestures are representative, since *-tati* can attach to type *e* proper names as well (Nakanishi & Tomioka, 2004).

an antecedent or entirely new. One could claim that the entities satisfying the predicate *entity* or the predicate *entity* itself are representative of this introduced referent, such that definite uses are to be analyzed either as in (81a) or in (81b):

- (81) a. $\exists e \exists y \exists z [event(e) \wedge y \leq \mu x [entity(x)] \wedge represents(y, z) \wedge role(e, z)]$
 b. $\exists e \exists y \exists z [event(e) \wedge y \leq \mu x [entity(x)] \wedge y \leq z \wedge represents(entity, z) \wedge role(e, z)]$

(81a) assumes that *represents*(*y*, *z*) includes a part-of relation $y \leq z$, while (81b) encodes $y \leq z$ separately since *entity* can be satisfied by more entities than those which make up *y* since contextual restriction is encoded in the μ operator, not in *entity* as a predicate. Theoretically one may analyze the examples above as involving entities which are representative of the range of entities fulfilling the thematic role *role*, but this option is disregarded in (81) due to the postulation of unique thematic role bearers in the literature (e.g. Chomsky, 1981, p. 36).

The exact choice between (81a) and (81b) is left open but may be a mere matter of definition of *represents* in the end. While the formulae capture the observations presented above, it suffers from the vague notion of representativity and from a lack of independent explanation. As it stands, the representativity phenomenon of iconic co-speech gestures is to be postulated separately from other semantic or pragmatic properties. Furthermore the correct definition of representativity may be well different from the factors governing the use of Japanese *-tati*³⁰. Another question is whether the important notion is if *y* represents *z* better than some standard or if *y* should not be misleading as a representative of *z*. In the first case in contexts where a representing *y* and a represented *z* seem to bear no salient connection but also *y* is not misleading as a proxy for *z* the formulae in (81) turn out false, while in the latter they turn out true. These open questions lead too far at this point. The observation that iconic co-speech gestures may involve non-exhaustive entities which are representative or not misleading for everything satisfying a thematic role of an event however has to be kept in mind, just as their being unexplained by a definite approach to gestures.

5.2 Overgeneration of antecedent members

Section 4.3.3 demonstrated that if relevance requirements enforce distinctions between identically portrayable entities, infelicity arises by the maximal antecedent. However there are other examples where the data is more challenging. Consider example the movie

³⁰Note however that such a morpheme is not only a peculiarity of Japanese. Leslau (2000, pp. 41-42) provides similar data for the Amharic proper name prefix *ənnä-*, and Nakanishi & Tomioka (2004, p. 124) refer the reader to den Besten (1996) for Afrikaans *-hulle*.

plot narration (82):

- (82) (Context: The plot of the movie ends with the narrated events. It is a generic action movie with a policeman called Dave as the protagonist. Dave investigated on his own after being suspended after an awful fight with his superior Copperton.)
- a. Also der Dave hat sich dann ne private Waffe geschnappt
so the Dave has himself then a private weapon grabbed
'So Dave took a private gun'
 - b. un is dann der Spur mitten in das Versteck von der Gang gefolgt.
and is then the trace directly into the hideout of the gang followed
'and then followed the trace right into the hideout of the gang.'
 - c. Die ham ihn erwischt und entwaffnet und er hätt fast den Löffel
they have him caught and disarmed and he had almost the spoon
abgegeben,
given_away
'They caught and disarmed him and he almost would've died,'
 - d. aber dann ist dann doch noch Copperton mit Verstärkung aufgetaucht
but then is then PRT still Copperton with reinforcement appeared
'but then Copperton showed up with reinforcement'
 - e. und hat ihm [geholffen] *fingergun–twice* .
and has him helped
'and helped him.'

The gesture *fingergun – twice* is a conventional fingergun gesture moving rapidly upwards twice in order to mimic the recoil of a shot. The context of an action movie setting together with the knowledge that Copperton and the reinforcement belong to a firearm-bearing unit licenses the idea that they might be likely to use firearms to help the protagonist. In the algorithm as stated in Section 4 the firearms of Copperton and his reinforcement can be treated as a case where vague information has a not entailed but likely specification (in this case helping as helping by using their firearms) which involves the entities depicted (see algorithm clause (51.2c) and its motivating examples (49)-(50)). The algorithm is however predicted to take an antecedent plurality consisting of all discourse referent and contextually salient guns - this involves Dave's private gun introduced in (82a). It can however safely be excluded that Copperton and his reinforcement used the gun of the man they were liberating instead of their own ones. In this the result of the algorithm is more vague than it ought to be.

One possible way to characterize the implausibility of Dave's gun as a member of the antecedent plurality of the gesture is a notion of *controversy*. Schlenker (2010, p. 10) notes that supplements, although they convey new information, may not do so if the supplemented information is controversial or surprising. Likewise Soames (1982,

p. 486) redefines presuppositions by appealing to a notion of being *uncontroversial* in face of presupposition accommodation data. While von Stechow (2008, p. 151) and Heim (1992, p. 212) do not endorse this proposal of definition, they also note that the degree of controversy regulates accommodation. The important observation of these three accounts is that pragmatics operates on a notion of controversy independently of gestures. Applying it to the members of the antecedent of a gesture would therefore not create entirely additional linguistic machinery that does not have to be investigated either way.

Like the problem of relevance and representativity, controversy is then a further notion involved in gestures which is hard to define. Another problem is that the notion of controversy Schlenker (2010) appeals to seems to be different from the one involved in co-speech gestures. He rules out casually conveying very meaningful content from appositive relative clauses such as the claim that former French president Sarkozy murdered his wife (Schlenker, 2010, p. 10). Section 4.3.3 however demonstrated that relevant distinctions between antecedent subgroups will clash with maximal antecedents, so arguing that the maximal antecedent in (82) does not contain Dave's gun because of a too meaningful contribution is unlikely. The kind of surprise encountered with the idea that the police uses Dave's gun involves the physical implausibility that the firearm can get into their, while Sarkozy murdering his wife is not physically implausible, just a very meaningful claim in terms of gossip and politics. In conclusion, there seems to be some pragmatic reasoning about unlikely antecedents which is plausibly involved, but its exact characterization is problematic at this point.

Finally by a naive conception of contexts licensing speech-gesture combinations one would expect the following story to be felicitous:

- (83)
- a. Luca hat Kim nen Antrag gemacht
Luca has Kim a proposal made
'Luca proposed to Kim'
 - b. und ihr nen richtig teuren Ring geschenkt.
and her a really expensive ring given
'and gave her a really expensive ring.'
 - c. Aber die hatten ja schon immer Probleme
but they had PRT already always problems
'But they've always had problems'
 - d. Und neulich ham die sich so gestritten,
and recently have they themselves so_much quarreled
'And recently they had such a bad fight'
 - e. ??dass Kim [richtig ausgerastet]_{pull-4-throw} is.
that Kim really lost_it is

‘that Kim really lost it.’

- f. Und jetzt zweifelt Luca halt stark.
and now doubts Luca PRT strongly
‘And now Luca has serious doubts.’

The gesture *pull – 4 – throw* represents holding the left hand in front of the body with slightly spread fingers and a movement of the right hand with the right fingertips sliding along the left ring finger, like when pulling something from that finger, succeeded by a gesture of tossing away something with the right hand. While it seems possible to resolve that *pull – 4 – throw* involves the ring mentioned in (83b), packaging this information as a gesture accompanying the verbally expressed event of Kim losing it seems odd. If (83e) is replaced by (84) the discourse is entirely acceptable:

- (84) a. dass Kim richtig ausgerastet is
that Kim really lost_it is
‘that Kim really lost it’
b. und den Ring auf den Boden geworden hat.
and the ring on the ground tossed has
‘and tossed the ring on the ground.’

If one shares the intuition that (83e) has a coreferential reading but is odd anyways, this implies that further restrictions on the combined usage of gesture and speech are at work. If not, it is plausible as well that the antecedent did in fact not license an unusual combination which is not licensed by the bridging condition alone, and then the licensing of certain bridging patterns by an antecedent would have to be restricted more accurately. The reasons for the infelicity of (83e) cannot be attributed to one particular reason at this point and will remain an open issue for the analysis of this thesis.

6 Conclusion

The analysis of Lascarides & Stone (2006, 2009) provides important contributions by giving examples of the semantics of gestures, how they may be related to discourse and by providing a formalism for their resolution. However the overall discourse structure does not seem to restrict them, and the *Replication* relation used to capture what in this thesis would be called (partial) coreference between two gestures obeys constraints that cannot be upheld. Analyzing the presumed definiteness in gestures as pronouns cannot be advocated by any predictive means from within and outside of SDRT, irrespectively of the modality (gesture or speech) of the antecedent.

Parallels between coreference with gestures and definite descriptions however are quite consistent. Both definite descriptions and definite gestures defy discourse structural

distance to a comparable degree by allowing coreference in the same positions. Although there is no difference between what is called definite and indefinite gestures in this thesis when looking at the gesture alone, some speech-gesture combinations require licensing context or discourse just like definite descriptions do. Licensing by context alone does not prove a definite description analysis. The context should make these uses felicitous only if a unique and maximal antecedent is provided by the context. But this seems to be the case.

Since many gestures seem to contribute entities to truth or felicity conditions which relate to an antecedent, but not necessarily to all of its subgroups, the definite description approach to gestures comes with the necessary ingredients of partitivity in the style of *some of the NP* and the requirement that the antecedent and its partitive subgroup are not specified for number. It is important to stress that partitives of definite descriptions obey certain pragmatic constraints which are perfectly derivable from a maximality treatment of definite descriptions. Just like for example *x took some of the pills* is not informative enough where a listener is interested in whether *x* took red or blue pills, the gesture of ingesting some small object is prohibited by the same constraints. Partitive definite descriptions and definite gestures can therefore be observed to behave very similarly.

These observations allow to say that the overall idea of definite gestures involving definite description semantics and pragmatics is a plausible approach to the partially coreferential data discussed. Building on the empirical observations this thesis contributes a semi-formal algorithm to resolve the unresolved form of gestures towards its antecedent (or the lack thereof), formula templates for definite use gestures, an integration into the SDRT framework and some considerations how to capture relevance phenomena with the help of discourse structure. The definite description approach allows for a uniform treatment of definite descriptions, definite gestures with speech antecedents and those with gestural antecedents, while Lascarides & Stone's (2009) *Replication* relation can only account for some cases of gestural antecedents.

The account of this thesis may not be the only plausible one. One alternative might be an approach where gestures are considered as extremely polysemous, and various contextual factors determine the predicate they are resolved to and the inferences about partial coreference. And there are still some open questions with the present account. Among these possibly the most problematic issue is whether all contexts which license new bridging relations as discussed in Section 4.3.2 can be regarded as providing an antecedent, and if not, how to draw the line between the two kinds of contextual licensing. Other questions are: Why do instruments provided by gestures seem to be non-exhaustive, while those by speech seem to be exhaustive? Certain contexts license gestures and partial coreference inferences, but what are the factors that limit licensing?

Can the bridging constraint be fleshed out to an exhaustive list? And can the literature on speech definite description projection be integrated to validate or invalidate the approach?

Many details of a successful approach remain vague. None of these unresolved issues however seems to constitute a fatal invalidation of the analysis rather than demanding further investigation to refine it. This thesis therefore has managed to flesh out one possible approach to a large data set in gesture pragmatics - that there is such a thing as gestures involving definite descriptions, and how to analyze this in context and discourse.

References

- Abbott, Barbara K. 2010. *Reference*. New York, NY: Oxford University Press.
- Alahverdzhieva, Katya & Alex Lascarides. 2010. Analysing speech and co-speech gesture in constraint-based grammars. *International Conference on Head-Driven Phrase Structure Grammar* 17. 6-26.
- Asher, Nicholas, Michel Aurnague, Myriam Bras, Pierre Sablayrolles & Laure Vieu. 1995. De l'espace-temps dans l'analyse du discours. *Sémiotiques* 9. 11-62.
- Asher, Nicholas & Alex Lascarides. 1998. Bridging. *Journal of Semantics* 15(1). 83-113.
- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation* (Studies in Natural Language Processing). Cambridge: Cambridge University Press.
- Asher, Nicholas, Daniel Hardt & Joan Busquets. 1997. Discourse parallelism, scope, and ellipsis. *Semantics and Linguistic Theory (SALT)* 7. 19-36.
- Asher, Nicholas, Laurent Prévot & Laure Vieu. 2007. Setting the background in discourse. *Discours* 1. 1-29. DOI: 10.4000/discours.301.
- Bach, Kent. 2004. Descriptions: Points of reference. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 189-229. New York, NY: Oxford University Press.
- Barker, Chris. 1998. Partitives, double genitives and anti-uniqueness. *Natural Language & Linguistic Theory* 16(4). 679-717.
- Barwise, Jon & John Perry. 1983. *Situations and attitudes*. Cambridge, MA & London: The MIT Press.
- Besten, Hans den. 1996. Associative DPs. *Linguistics in the Netherlands* 13(1). 13-24.
- Birner, Betty & Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. *Annual Meeting of the Berkeley Linguistics Society* 20. 93-102.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Clark, Herbert H. 1977. Bridging. In Philip N. Johnson-Laird & Peter C. Wason (eds.), *Thinking: Readings in cognitive science*, 411-420. Cambridge: Cambridge University Press.
- Copestake, Ann. 2006. Robust Minimal Recursion Semantics. <https://www.cl.cam.ac.uk/aac10/papers/rmrsdraft.pdf> (30 July, 2020).
- Copestake, Ann. 2007. Semantic composition with (Robust) Minimal Recursion Semantics. In Timothy Baldwin, Mark Dras, Julia Hockenmaier, Tracy Holloway King & Gertjan van Noord (eds.), *ACL 2007 Workshop on Deep Linguistic Processing*, 73-80. Prague: Association for Computational Linguistics.

- den Besten, Hans (see Besten)
- Devitt, Michael. 2004. The case for referential descriptions. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 280-305. New York, NY: Oxford University Press.
- Donnellan, Keith S. 1966. Reference and definite descriptions. *The Philosophical Review* 75(3). 281-304.
- Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Handout of a talk given at *Semantics and Philosophy in Europe (SPE) 7*, Berlin, June 28, 2014.
- Esipova, Maria. 2018. Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive focus. *Sinn und Bedeutung* 22(1). 385-402.
- Farkas, Donka F. & Henriëtte E. de Swart. 2010. The semantics and pragmatics of plurals. *Semantics and Pragmatics* 3(6). 1-54.
- Fintel, Kai von. 2008. What is presupposition accommodation, again?. *Philosophical Perspectives* 22. 137-180.
- Fintel, Kai von, Danny Fox & Sabine Iatridou. 2014. Definiteness as maximal informativeness. *MIT Working Papers in Linguistics* 70. 165-174.
- Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Atlantic Highlands, NJ: Humanities Press.
- Heim, Irene R. 1982. *The semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachusetts dissertation.
- Heim, Irene R. 1983. On the projection problem for presuppositions. *WCCFL (West Coast Conference on Formal Linguistics)* 2. 114-125.
- Heim, Irene R. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3). 183-221.
- Heim, Irene R. & Angelika Kratzer. 1998. *Semantics in generative grammar* (Blackwell Textbooks in Linguistics 13). Malden, MA & Oxford: Blackwell Publishers.
- Horn, Laurence R. 1976. *On the semantic properties of logical operators in English*. Bloomington, IN: Indiana University Linguistics Club.
- Horn, Laurence R. 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61(1). 121-174.
- Hunter, Julie & Márta Abrusán. 2015. Rhetorical structure and QUDs. *JSAI-isAI (JSAI International Symposium on Artificial Intelligence)*. 41-57.
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts* (Syntax and Semantics 3), 41-58. New York, NY: Academic Press.
- Grice, Herbert Paul. 1981. Presupposition and conversational implicature. In Peter Cole (ed.), *Radical pragmatics*, 183-198. New York, NY: Academic Press.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274-307.
- Kadmon, Nirit. 1990. Uniqueness. *Linguistics and Philosophy* 13. 273-324.
- Kamp, Hans & Uwe Reyle. 1993a. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory. Part 1.* (Studies in Linguistics and Philosophy 42). Dordrecht: Kluwer Academic Publishers.
- Kamp, Hans & Uwe Reyle. 1993a. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory. Part 2.* (Studies in Linguistics and Philosophy 42).

- Dordrecht: Kluwer Academic Publishers.
- Karttunen, Lauri. 1971. Discourse referents.
<http://semantics.uchicago.edu/kennedy/classes/s08/semantics2/karttunen71.pdf> (10 August, 2020).
- Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4(2). 169-193.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. New York, NY: Cambridge University Press.
- Lambrech, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press.
- Lappin, Shalom & Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics* 20(4). 535-561.
- Lascarides, Alex & Matthew Stone. 2006. Formal semantics for iconic gesture. *Workshop on the Semantics and Pragmatics of Dialogue* 10. 64-71.
- Lascarides, Alex & Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics* 26(4). 393-449.
- Leslau, Wolf. 2000. *Introductory grammar of Amharic*. Wiesbaden: Otto Harrassowitz.
- Link, Godehard. 1983. The logical analysis of plural and mass nouns: A lattice-theoretic approach. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 302-323. Berlin & New York, NY: Walter de Gruyter.
- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279-326.
- May, Robert C. 1977. *The grammar of quantification*. Cambridge, MA: MIT dissertation.
- May, Robert C. 1985. *Logical Form: Its structure and derivation* (Linguistic Inquiry Monographs 12). Cambridge, MA & London: The MIT Press.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL & London: The University of Chicago Press.
- Mill, John Stuart. 1843. *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*, vol. 1. https://archive.org/details/bub_gb_y4MEAAAAQAAJ (23 August, 2020).
- Müller, Cornelia. 1998. *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich*. Berlin: Berlin Verlag.
- Nakanishi, Kimiko & Satoshi Tomioka. 2004. Japanese plurals are exceptional. *Journal of East Asian Linguistics* 13(2). 113-140.
- Nunberg, Geoffrey. 2004. Descriptive indexicals and indexical descriptions. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 261-279. New York, NY: Oxford University Press.
- Onea Gáspár, Edgar. 2016. *Potential questions at the semantics-pragmatics interface* (Current Research in the Semantics/Pragmatics Interface 33). Leiden: Koninklijke Brill.
- Preiss, Judita, Caroline Gasperin & Ted Briscoe. 2004. Can anaphoric definite descriptions be replaced by pronouns? *LREC (International Conference on Language Resources and Evaluation)* 4. 1499-1502.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223-255. New York, NY: Academic Press.
- Reed, Ann M. 1991. On interpreting partitives. In Donna Jo Napoli & Judy Anne Kegl (eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila*

- Gleitman*, 207-223. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Reed, Ann M. 1996. Partitives, existentials, and partitive determiners. In Jacob Hoeksema (ed.), *Partitives: Studies on the syntax and semantics of partitive and related constructions*, 143-178. Berlin & New York, NY: Mouton de Gruyter.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5. 1-69.
- Rullmann, Hotze & Aili You. 2006. General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese. In Klaus von Heusinger & Ken Turner (eds.), *Where semantics meets pragmatics* (Current Research in the Semantics/Pragmatics Interface 16), 175-196. Oxford: Elsevier.
- Russell, Bertrand. 1905. On Denoting. *Mind* 14(56). 479-493.
- Sadock, Jerrold M. 1978. On testing for conversational implicature. In Peter Cole (ed.), *Pragmatics* (Syntax and Semantics 9), 281-298. New York, NY: Academic Press.
- Salmon, Nathan. 2004. The good, the bad, and the ugly. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 230-260. New York, NY: Oxford University Press.
- Scharten, Rosemarijn. 1997. *Exhaustive interpretation: A discourse-semantic account*. Nijmegen: Katholieke Universiteit Nijmegen dissertation.
- Schlenker, Philippe. 2010. Supplements within a unidimensional semantics II: Epistemic status and projection. <https://ling.auf.net/lingbuzz/002579> (12 November, 2019).
- Schlenker, Philippe. 2018a. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295-365.
- Schlenker, Philippe. 2018b. Iconic pragmatics. *Natural Language and Linguistic Theory* 36(3). 877-936.
- Schwarz, Florian. 2013. Maximality and definite plurals - Experimental evidence. *Sinn und Bedeutung* 17. 509-526.
- Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89(4). 607-624.
- Soames, Scott. 1982. How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry* 13(3). 483-545.
- Stalnaker, Robert. 1973. Presuppositions. *Journal of Philosophical Logic* 2(4). 447-457.
- Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Pragmatics* (Syntax and Semantics 9), 315-332. New York, NY: Academic Press.
- Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320-344.
- von Fintel, Kai (see Fintel)