

Representing *De Se* Thoughts and their Reports

Hans Kamp

University of Stuttgart, University of Texas at Austin

August 31, 2011

1 Introduction

- This talk is about the representation of *de se* and *de nunc* and about their relation to the use of first and second person pronouns in English.
- It distinguishes between
 - (i) the representation of *de se/de nunc* thoughts and
 - (ii) the representation of the reports of such thoughts.
- What I will present is part of a larger enterprise: to develop a viable account of the structure and content of thought, which can *then* serve as the basis of an account of the semantics (and pragmatics) of attitude reports.
- Important requirements on a theory of the structure and content of thought:
 - (i) The theory must be able to handle combinations of connected thoughts, not just single thoughts.
 - (ii) Thought complexes will typically not only consist of propositional attitudes, but also involve *entity representations*. The theory must be able to handle these as well.
- Important requirements on an account of the semantics of attitude reports:
 - (i) Attitude reports that people use are very diverse, vastly more so than most current theories of attitude reports acknowledge. We often report on complexes consisting of several connected attitudes, on the ways in which attitudes change over time (using verbs like *learn*, *forget*), on the ways in which attitudes are shared between different agents; and sometimes on all of that at once. The theory must be able to deal with such reports as well, and not just with reports of individual attitudinal states, as given in sentences of the sort *John believes that...*, *Mary wants that...*

(ii) Attitude reports are often highly underspecified. The theory must be in a position to explicate that intuition and to get the interpretation of underspecified reports right.

(iii) A significant number of problems in linguistic semantics and pragmatics require for their solution that we distinguish sharply and carefully between the production of an expression and its interpretation.

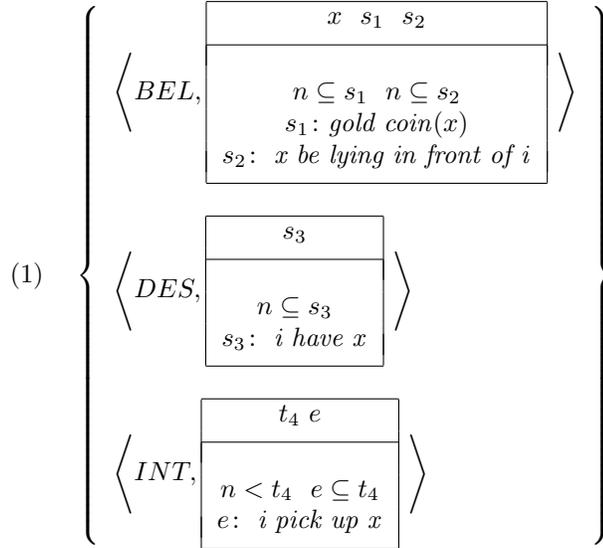
In such cases the theory must be able to articulate not only the ‘participant-neutral’ content of the expression, but also the thought or thoughts of the speaker that she chooses her words to express and the thought(s) that the addressee obtains by interpreting her words.

N.B. The distinction between the thoughts of the speaker and those of the hearer are especially important when the words contain elements that invite interpretations *de se*.

- Both the theory of thought and the semantics of thought reports I will present parts of are heavily indebted to *Discourse Representation Theory* (Kamp & Reyle (1993), Beaver & Geurts (2005), Genabith et al. (2011)):
 - The focus of the theory of thought on complexes of structurally connected propositional attitudes was inspired by the nature of incremental interpretation that is central to DRT.
 - Moreover, DRSs serve as representations of the contents of individual attitudes, and thus as the principal building blocks of the attitudinal state structures that are postulated by the theory.
 - The syntax-semantics interface used in assigning logical forms to attitude reports (and other sentences) is formulated in DRT terms.
- The first part of the paper presents the motivation for the formalism and, by presenting a number of examples, shows what it is like and how it works.
- The second (and longer) part deals with the relation between *de se* and pronouns *I* and *you*.

2 Some examples of representations of complex thoughts

1. Mental state of an agent who perceives (or thinks she perceives) a gold coin in front of her, wants to have it and intends to pick it up:

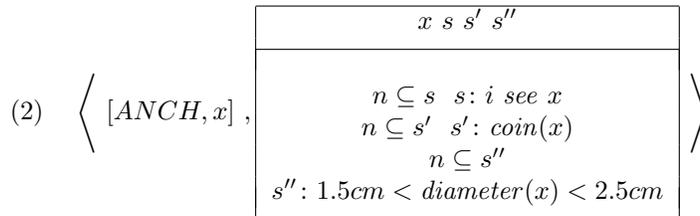


Semantics for (1): Find proposition-like values for the content of belief, desire and intention in (1) such that the ‘internal coreference relations’ between the three attitudes (indicated by the occurrence of the same discourse referent x in the three content characterisations) are accounted for.

This can be accomplished by treating the belief in (1) as resolving referential presuppositions imposed by the desire and the intention.

Note well: the propositions that get assigned this way to the components of (1) are general propositions as far as the role of x is concerned. (However, they are singular with respect to i and n .)

2. Anchored entity representation of the coin for the agent in (1):



3. Representation of the agent's mental state with entity representation for the coin:

$$(3) \left\{ \begin{array}{l} \left\langle [ANCH, x], \begin{array}{|l} x \ s_1 \ s_2 \\ \hline n \subseteq s_1 \quad n \subseteq s_2 \\ s_1: i \ see \ x \\ s_2: x \ be \ lying \ ifo \ i \end{array} \right\rangle \\ \left\langle BEL, \begin{array}{|l} s_3 \\ \hline n \subseteq s_3 \\ s_3: gold \ coin(x) \end{array} \right\rangle \\ \left\langle DES, \begin{array}{|l} s_4 \\ \hline n \subseteq s_4 \\ s_4: i \ have \ x \end{array} \right\rangle \\ \left\langle INT, \begin{array}{|l} t_5 \ e \\ \hline n < t_5 \ e \subseteq t_5 \\ e: i \ pick \ up \ x \end{array} \right\rangle \end{array} \right\}$$

The problem that (3) presents for an account of propositional content is different from the one presented by (1). In (3) the DRSs of each of the three propositional attitudes determine propositions that are singular with respect to the entity represented by the anchored entity representation.

However, these propositions are *defined* only if the anchored entity representation whose discourse referent occurs in their DRSs does have a referent. (If there is a referent, it is also called the *external anchor* of the entity representation). The referent is an entity that stands to the representation and the agent in the relation described by the entity representation's DRS (the so-called *internal anchor* of the entity representation).

Even if this condition is not fulfilled, there nevertheless remains a 'purely descriptive core' of the representation that is semantically well-defined. This core is obtained by exchanging the entity representation for the belief that an entity of the kind described by the internal anchor exists; see (4).

$$(4) \left\{ \begin{array}{l} \left\langle \begin{array}{l} BEL, \\ n \subseteq s_1 \quad n \subseteq s_2 \\ s_1: i \text{ see } x \\ s_2: x \text{ be lying ifo } i \end{array} \right\rangle \\ \left\langle \begin{array}{l} s_3 \\ BEL, \\ n \subseteq s_4 \\ s_3: \text{gold coin}(x) \end{array} \right\rangle \\ \left\langle \begin{array}{l} s_4 \\ DES, \\ n \subseteq s_4 \\ s_4: i \text{ have } x \end{array} \right\rangle \\ \left\langle \begin{array}{l} t_5 \ e \\ INT, \\ n < t_5 \ e \subseteq t_5 \\ e: i \text{ pick up } x \end{array} \right\rangle \end{array} \right\}$$

Internal and external anchors

- No internal anchor without an external anchor:

Internal anchors come with a presupposition to the effect that there is a corresponding external anchor: If x is the discourse referent of an internally anchored entity that is part of an agent's mental state, but x has no external anchor corresponding to the representation's internal anchor (i.e. there is no entity to which agent and representation are causally related in the way the internal anchor describes), then the internal anchor is 'ungrounded'.

In that case every thought of the agent in which x figures as a constituent fails to define a propositional content.

- No external anchor for x without an internal anchor for x :

An external anchor for a discourse referent x figuring in the internal (structural) characterisation of part of an agent's mental state is meaningless unless x is internally anchored.

Moreover, in case x is internally and externally anchored, the actual relation between its external anchor, its internal anchor and the agent must fit the specification provided by the internal anchor.

Question 1: What are the preconditions for the formation of anchored entity representations and what is the repertoire of internal anchor specifications?

Question 2: How closely should the actual relation fit the internal specification lest the internal anchor should be declared incoherent?

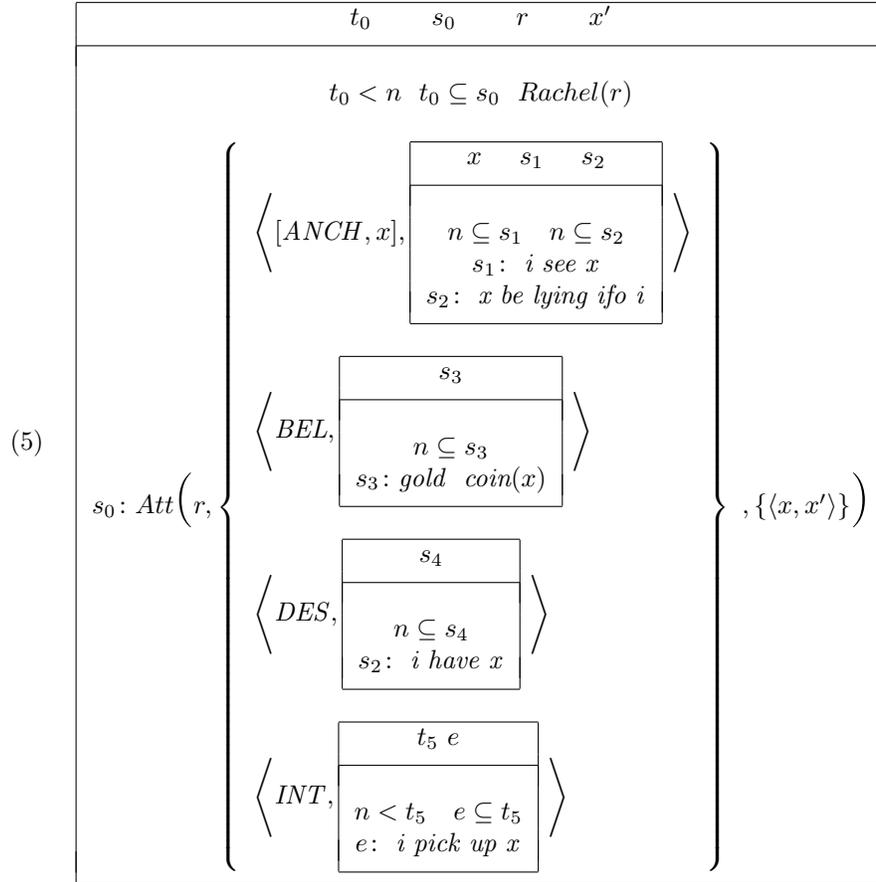
Representing Attitude Attributions

Representations like those in (1)–(4) can also be used as constituents of representations of attributions of the mental states they describe, as shown in (5) for (3).

The central condition of (5) is the one which characterises the state s_0 . This condition is a predication involving the static predicate *Att*. This predicate is used to ascribe simple and complex mental states to subjects capable of entertaining such states. It has three non-referential argument slots. The first is for the agent whose mental state is being described. The second slot is for the characterisation of the ascribed mental state; it can be filled by state descriptions like that in (3). The third argument slot is for external anchors for discourse referents of internal arguments belonging to the second argument of *Att*; it can be filled with terms that denote sets of external anchors for discourse referents of anchored entity representations occurring in the second argument of *Att*; we assume that these external anchors are given as pairs $\langle x, x' \rangle$, where x is the discourse referent of some anchored entity representation in *Att*'s second argument and x' is some other discourse referent, which does not occur in the second argument of *Att*.

(It follows from the principle above that an *Att*-predication will make a coherent attribution to the first argument of *Att* only if its third argument contains an external anchor for each of the discourse referents of internally anchored entity representations in the second argument.)

The state characterised by an *Att*-predication, such as s_0 in (5), can be located in time in the usual way. Thus in (5) s_0 is represented as holding at the utterance time n .



3 *De Se* and *De Nunc*

1. The role of *i* and *n*

The present account assumes that self-attributing thoughts take the form of predications of a special singular term that 'represents the self' to the agent *as* her self. We use the symbol *i* to play this part.

In the same way we use the symbol *n* to represent the 'psychological present' of the agent. More precisely, let T be any thought entertained by an agent at a time t and suppose that *n* is a constituent of T. (This is nearly always the case.) Then at that time t the occurrences of *n* in T will represent t, and will represent t as the agent's psychological present.

N.B. Since *n*, as it occurs in the representations of mental states, always refers

to the time at which the represented thoughts are entertained, this has the following consequence for when an agent entertains a thought with a given representational form over a given period of time: if this form contains n , then the content of the thought it represents will change as time goes on, as n will refer to different instants within this period as the period goes by.

Example: Suppose that the thought content in (6) is a belief that agent A entertains at time t . (In other words, $\langle \text{BEL}, (6) \rangle$ is one of the attitudes that make up A's mental state at t .)

$$(6) \quad \boxed{\begin{array}{c} s \\ n \subseteq s \\ s: \text{hungry}(i) \end{array}}$$

(6) is the belief that A herself could express in the words '(At this moment) I am hungry.'

Furthermore, in its role as the representation of a thought that A entertains at t the propositional content of (6) is the doubly singular proposition (relative to A and t) that A is at t in a state of being hungry.

(This effect – the incessant change of the propositional content of many of an agent's thoughts – can be mitigated by the agent's understanding of the occurrences of n in her thought as representing some interval that properly includes the utterance time (an 'extended now'). If the 'now' that agent associates with some particular thought is temporally extended in the direction of the future, then the content of the thought can remain constant for as long as this 'extended present' lasts. Note well, however, that as things stand, our formalism does not have the means to formalise this adequately.)

2. Comparison with D. Lewis' 'Attitudes De Dicto and De Se'

One of the central intuitions connected with the concept of *de se* is that our access to ourselves is completely direct; it is not conveyed by, and therefore unencumbered by, any form of descriptive content. (Perry (2000)).

Lewis' account of *de se* attitudes in his 'Attitudes *de dicto* and *de se*' (Lewis (1979)) offers what is arguably the purest way to account for this aspect of *de se*:

All propositional attitudes of creatures like ourselves are in last analysis self-attributions, in that all of them are reflections of our efforts at 'self-location':

They serve us not only to locate ourselves in logical space, but also with regard to who we are in this world, and where we are in it, in the sense of real time.

Lewis' theory and the one presented here are related as follows:

To the thought which in the present approach is assumed to have the structure given in (6) corresponds in Lewis' theory the relation that in standard notation could be represented either as (7.i) or as (7.ii).

- (7) i. $\lambda x.\lambda t.\text{hungry}(x,t)$
ii. $\lambda w.\lambda x.\lambda t.\text{hungry}(x,t,w)$

As far as these representations/logical forms are concerned, there doesn't seem to be any difference of substance between the two accounts of the *de se* and the *de nunc*. In fact, there is a straightforward way of translating between them

More specifically, they share:

- (i) Setting self- and now-attributions clearly apart from attributions to others and to times externally conceived.
- (ii) Saying nothing 'positive' about what self-attribution amounts to. (That will reveal itself, to the extent that that is possible, in the role that self-attribution plays 'in the mental life' of the agent.)

There are, however, some differences between Lewis' account and the one I have outlined:

- (i) The present account acknowledges thoughts with *de re* content as a category that is distinct both from thoughts *de se* and from thoughts *de dicto*.
- (ii) I favour a haecceitist ontology.
(Nothing in this talk, however, depends on the choice between haecceitism and counterpart theory.)

(For more about the problems that arise for Lewis' reduction of *de re* to *de se* see Ninan (2011*b*).)

3. Classical *de re* vs. *de se* puzzles

The 'classical' *de re* vs. *de se* puzzles – seeing yourself in a mirror with your pants on fire, or spilling sugar from your shopping cart, etc. without realising it is you (Perry (2000), Kaplan (1989), etc) are naturally dealt with in the present framework (as they are in Lewis' theory); likewise for similar puzzles involving *de nunc*.

4 Attitude reports involving the pronouns *I* and *you*

Consider the following dialogue (8):

- (8) **A:** Why do you dislike me?
B: Why do you think I dislike you?
A: Well, why do YOU think I dislike YOU?
B: I dont understand. Why do you think I think you dislike me?

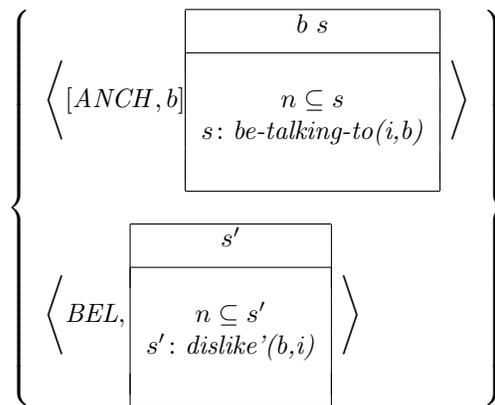
(**A:** Well, dont you? **B:** No! And you know it, and you know perfectly well I that like you.)

We focus on the statements that are the topics of the why-questions in the first four utterances of (8), enumerated in (9):

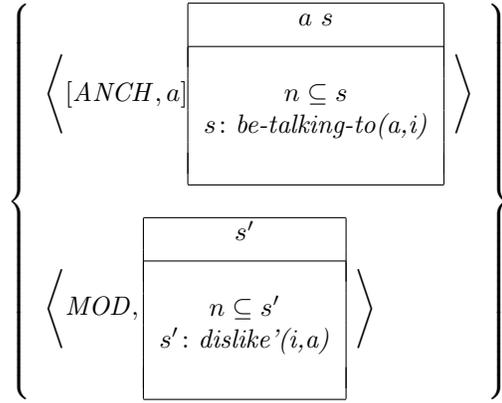
- (9) **A1':** You dislike me.
B1': You think I dislike you.
A2': YOU think I dislike YOU.
B2': You think I think you dislike me.

First, I will, in (10), give some representations for these sentences. After that we will go through the construction of a few of those representations in some detail.

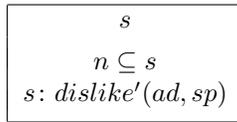
- (10) a. Representation of the thought complex in the mind of A that leads A to use the words in (9.A1').



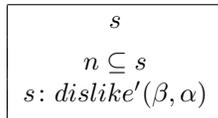
- b. Representation of the thought that B extracts from the words in (9.A1').



- c. Participant-neutral representation of the content of (9.A1').

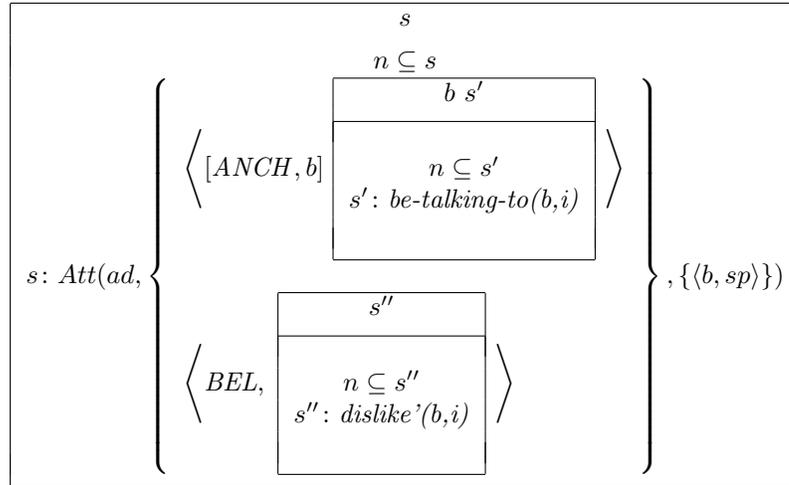


- d. Alternative participant-neutral representation of the content of the words in (A1') of (9), using discourse referents α and β to represent A and B, respectively, instead of sp and ad .

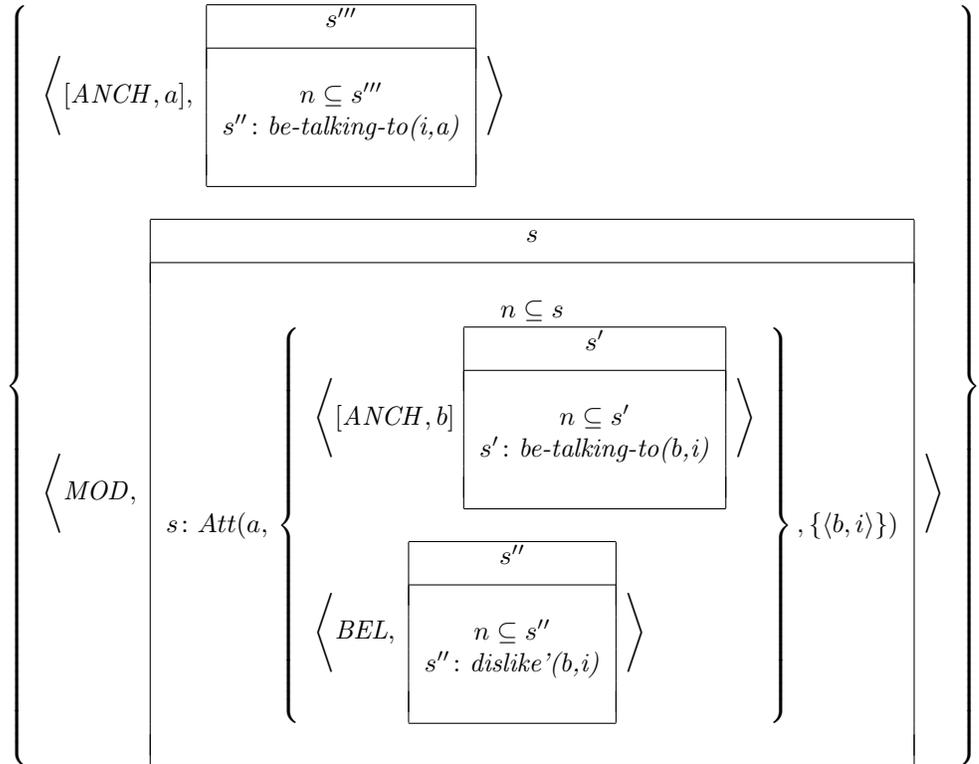


Note well: The occurrences of n have a different force in (10.c) and (10.d) from the one they have in (10.a) and (10.b). In (10.a) n represents the psychological 'now' of the speaker A – a constituent of As thoughts at the time of the utterance (8.A1); likewise for the occurrence of n in (10.b). In (10.c,d) n stands for the – publicly accessible – utterance time.

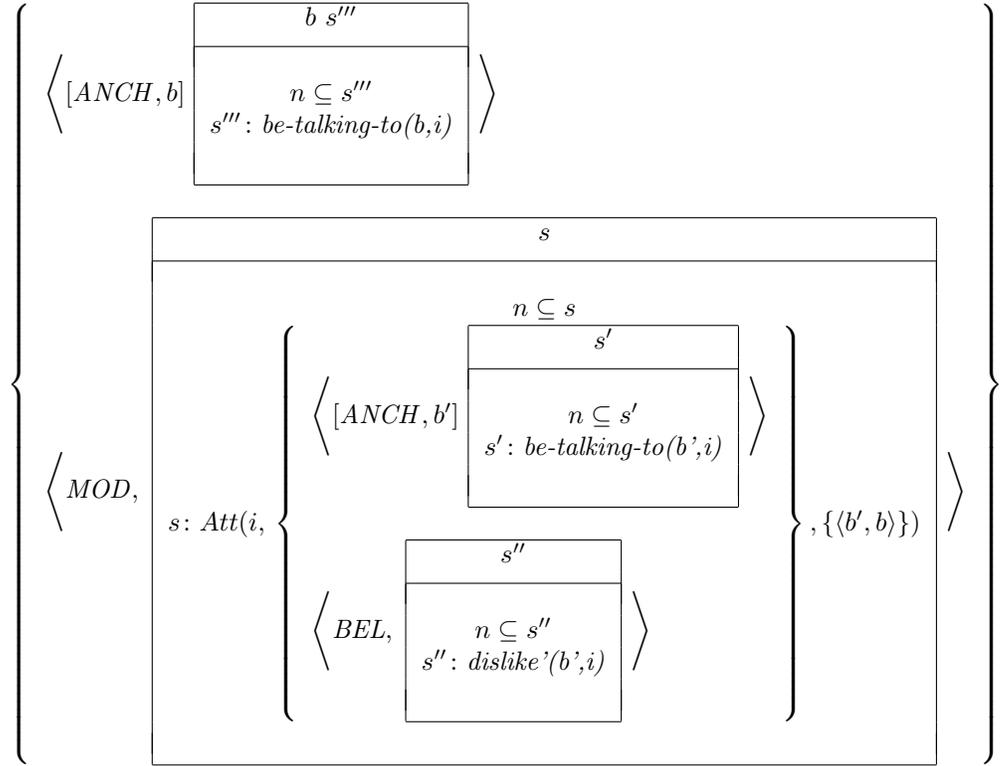
e. Participant-neutral representation of (9.B1').



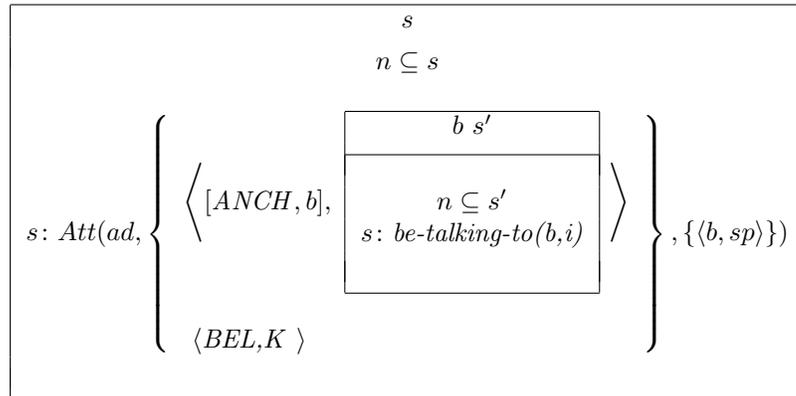
f. Thought complex of B that motivates the words in (9.B1').



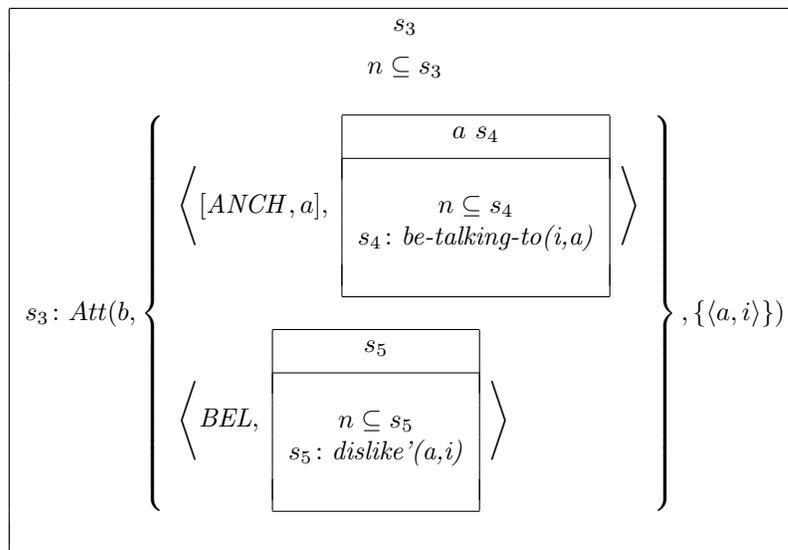
g. Thought that A extracts from (9.B1').



h. Participant-neutral representation of (9.B2').

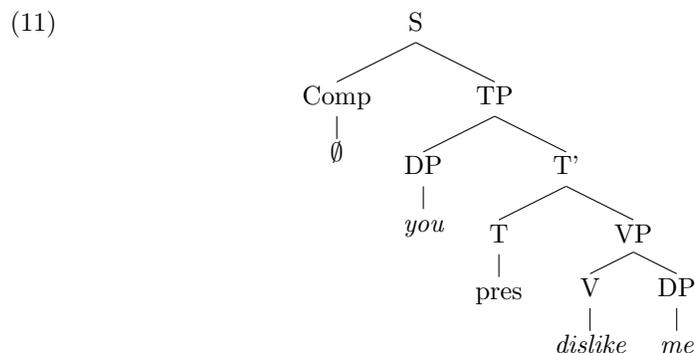


Here K is the structure:



4.1 Construction of some of the representations above

We now go through the construction of some of the representations shown above. We start with the construction of the participant-neutral representation of the content of the sentence (9.A1'). We assume for this sentence the following syntactic structure:



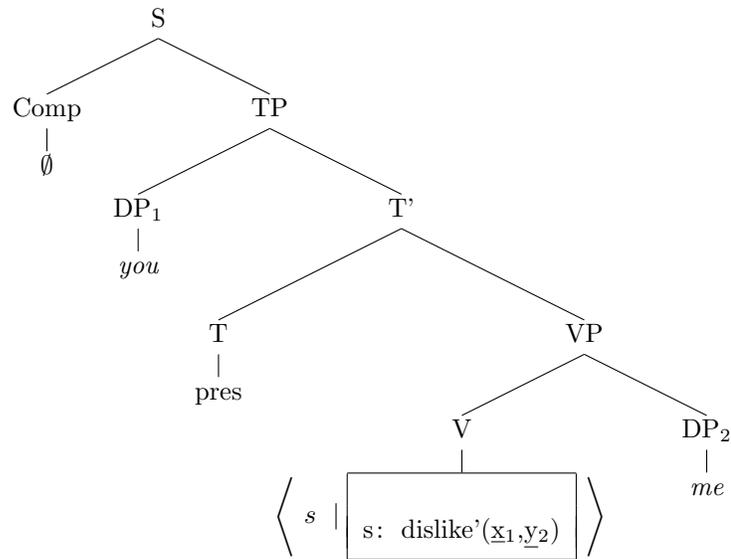
We build our representations bottom up. The first operation we perform is to insert the semantic representation of the verb *dislike*, which is given by the lexical entry (12). The insertion operation involves (i) choosing a ‘fresh’ discourse referent that is to take the place of ‘s’ as referential argument of the verb (since this is the very first operation we may as well use ‘s’ for this purpose),

(ii) putting that discourse referent in the store of the representation (where it will remain until it gets bound through being transferred to some DRS universe) and (iii) turning the other two arguments ‘x’ and ‘y’ into *argument slot* symbols. (These are linked via co-indexing with the argument phrases of (11), which will supply the discourse referents that will fill those slots. This results in the partially interpreted syntactic tree in (13).

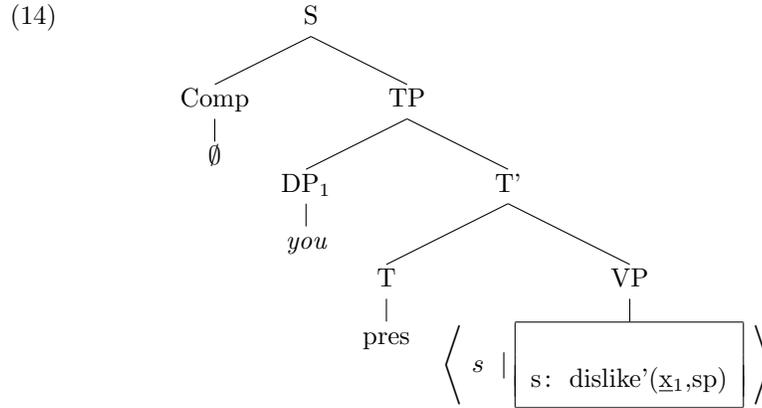
(12)

	<i>dislike</i> (verb)	nom	acc
	s	x	y
Sel. Restr:	state	being endowed with consciousness	
Sem. Repr:	$s : dislike'(x, y)$		

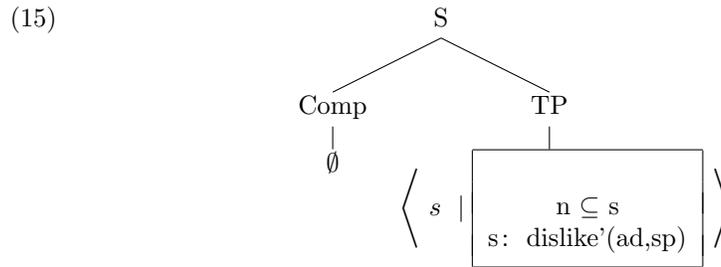
(13)



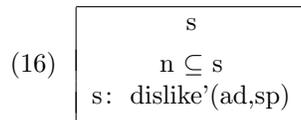
In the construction of participant-neutral representations the first person pronoun is interpreted as the indexical discourse referent *sp*, which represents the speaker of the represented utterance. Combining this interpretation of *me* with the V-representation we get the VP-representation shown in (14).



The next step integrates the information of tense, which temporally locates the state s as surrounding the utterance time n . The second person pronoun *you* is interpreted as the indexical discourse referent *ad*, representing the addressee, which gets inserted into the argument slot of the predicate *dislike'* that is linked to the subject DP. This step leads to the TP-representation shown in (15).



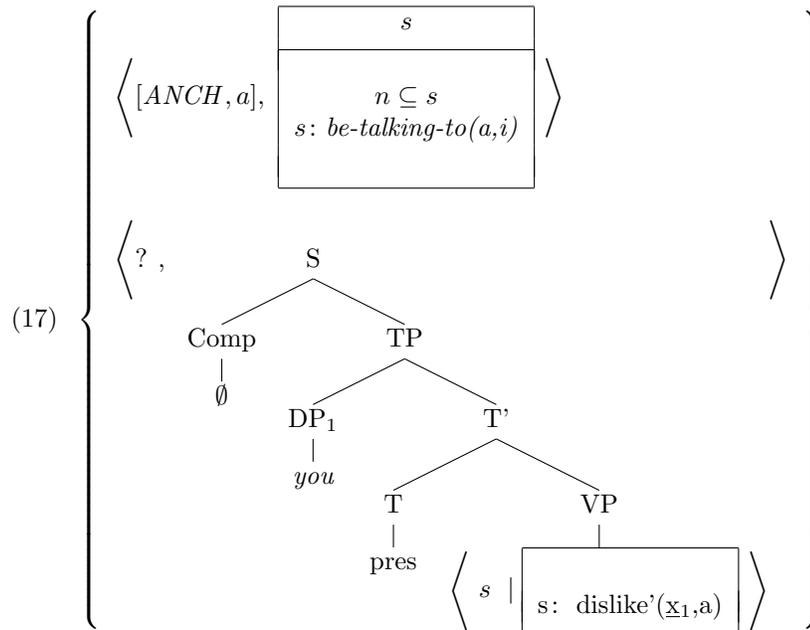
The transition from TP to S effects the transfer of the discourse referent s from the store to the universe of the DRS following it. This leads to the DRS in (16).



We now show the construction of the interpretation that the addressee B of the words in (9.A1') can be expected to construct from these words.

What attitude B will adopt towards this propositional content is necessarily indeterminate at this point, as no propositional content has yet been constructed, towards which B could adopt an attitude. We indicate this by placing a question mark in the slot reserved for the mode indicator.

The interpretation of the first person pronoun *me* by B leads either to the use of an anchored entity representation which B already has of the speaker, the discourse referent of which is inserted in the argument position to which the pronoun occurrence is linked – here: the direct object position of the verb *dislike* – or, in the unlikely event B doesn't have an entity representation for the speaker yet, to the creation of one, which will then be put to the same use. Either way B will have at this point of processing (9.A1') a structure like that in (17).



On the other hand the addressee B will interpret the pronoun *you* as an instruction to interpret the words he is processing as a self-ascription (with respect to the argument position to which *you* is linked). This way of interpreting *you* leads (in conjunction with the various other construction steps involved in getting the semantic representation for the S-node, which are the same as in the construction shown earlier and which therefore we do not display) to the S-node representation shown in (18).

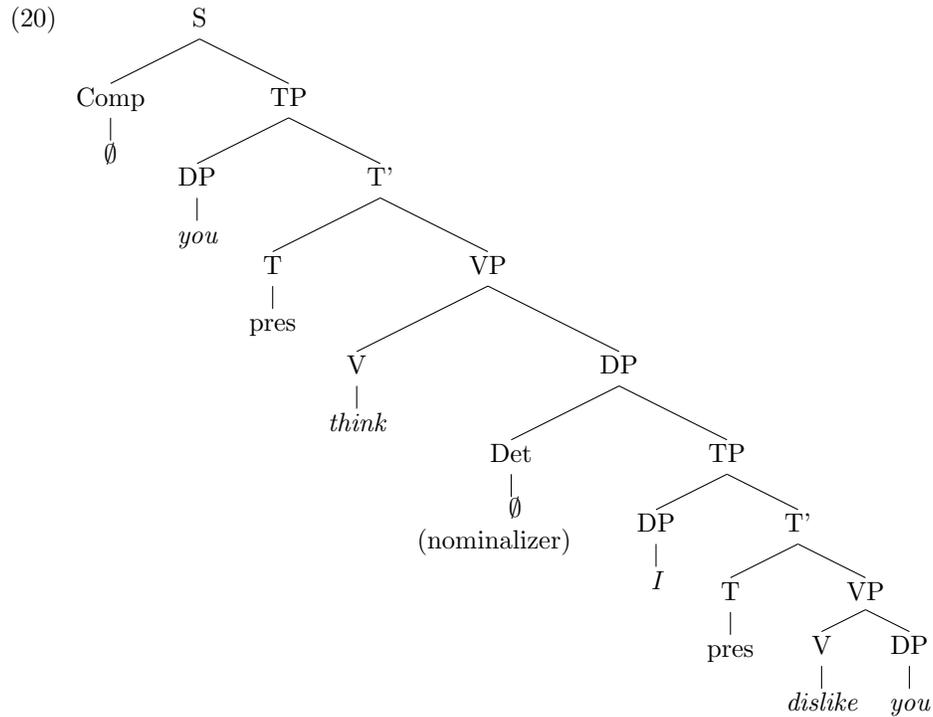
$$(18) \left\{ \begin{array}{l} \left\langle [ANCH, a], \begin{array}{|c|} \hline s \\ \hline n \subseteq s \\ s: be\text{-talking-to}(a,i) \\ \hline \end{array} \right\rangle \\ \left\langle ? , \begin{array}{c} S \\ | \\ \left\langle s \mid \begin{array}{|c|} \hline n \subseteq s \\ s: dislike'(i,a) \\ \hline \end{array} \right\rangle \end{array} \right\rangle \end{array} \right\}$$

$$(19) \left\{ \begin{array}{l} \left\langle [ANCH, a], \begin{array}{|c|} \hline s \\ \hline n \subseteq s \\ s: be - talking - to(a, i) \\ \hline \end{array} \right\rangle \\ \left\langle MOD, \begin{array}{|c|} \hline s \\ \hline n \subseteq s \\ s: dislike'(i, a) \\ \hline \end{array} \right\rangle \end{array} \right\}$$

More of a challenge is the construction of the representations in (10.e - h). We will just show the construction of the participant-neutral representation of (10.B1): ‘You think I dislike you’.

We will make two attempts at this construction. The first attempt will not give us quite what we want. But it will get a number of problems out of the way that won’t have to worry about in our second attempt, which will deal with the remaining problems, concerning the interpretation of the first and second person pronoun occurrences.

Syntactic structure of (B1')



In the present account the semantic contribution made by an attitudinal verb such as *think* involves operations at two different processing levels. The first level is that of the construction of the preliminary representation. The second level is that of presupposition resolution.

At the first level the semantic representation of the attitude verb is inserted at the place(s) where the verb occurs in the syntactic tree and its argument positions are then filled by the semantic contributions of the argument phrases with which those positions are linked; this is the same as it is for any other kind of main verb.

It is only at the second level that the representation is integrated into such a predication. This is the effect of resolving the presuppositional requirement that a suitable *Att*-predication be recovered from the discourse context – *Att*-predications typically play the part of secondary contexts in the sense of (Stalnaker (1988)) – into which the given representation *K* can be incorporated, and which may then be used to resolve any presuppositions that are part of *K*.

In the present instance, where we are constructing the representation for (B1') without the benefit of any discourse context, introducing a new *Att*-predication

is the only option.

We assume the following entry for the verb *think*:

	<i>think</i> (attitude verb)	nom	acc
	s	x	p
(21) Sel. Restr:	state	being endowed with consciousness	proposition
Sem. Repr:	$s : think'(x, p)$		

N.B. The classification of *think* as an attitude verb carries the information that semantic processing of the conditions in which it is the main predicate comes with the requirement to retrieve or introduce a suitable *Att*-predication. We represent this requirement as a presupposition. (It is a general feature of the construction algorithm used here that the representations of the presuppositions generated within a clause are collected in a set that is left-adjoined to the semantic representation of the clause.)

Participant-neutral processing of the complement clause leads (as before) to the following representation of its TP node.

(22) Semantic representation of the TP-node of (20).

$$\left\langle s \mid \boxed{\begin{array}{c} n \subseteq s \\ s: dislike'(ad, sp) \end{array}} \right\rangle$$

The complementizer forms out of this representation a new representation in which the referential argument is now a proposition-representing discourse referent *p*, while the previous referential argument *s* is bound by transfer to the DRS universe. The content of *p* is given by the semantic representation of the complement sentence. We express this by means of an equation of the form ' $p = \wedge K$ ', where *K* is the DRS which specifies the propositional content of the complement clause.

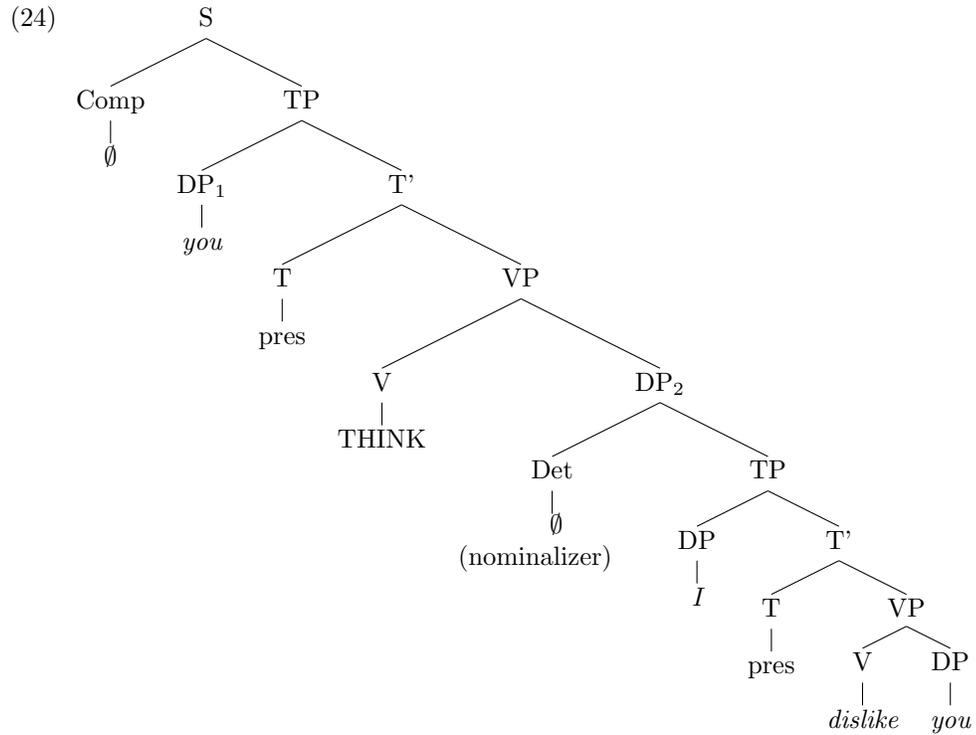
For the case at hand, in which the sister to the nominalizing complementizer of the complement to *think* has the semantic representation given in (22), this leads to the representation in (23) for the ‘direct object DP’ of the complement clause; p is the referential argument of the DP semantics.

$$(23) \left\langle p \mid \begin{array}{c} p = \wedge \\ \begin{array}{c} s \\ n \subseteq s \\ s: \text{dislike}'(\text{sp,ad}) \end{array} \end{array} \right\rangle$$

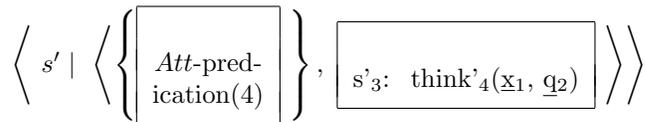
As noted above, lexical insertion for an attitude verb comes with a specification of the requirement to find a suitable *Att*-predication in the discourse context, given in the form of a presupposition. (Introducing a new *Att*-predication in case none can be found in the discourse context, should be seen as an instance of *presupposition accommodation*.) However, since none of the established ways of representing presuppositions within the formalism we are using provides a ready-made representation of this requirement, we encode the requirement simply as ‘Att-predication(j)’, where j is an index that is introduced on the semantic representation of the attitude verb when this representation is introduced for an occurrence of that verb. The index is realized as a subscript on the occurrence of the lexical predicate that is introduced by the attitudinal verb – here the predicate ‘think’.

N.B. In general, the content representation for a presupposition fails to make explicit the options and constraints that govern the resolution of presuppositions of the particular kind to which the given presupposition belongs. (As a rule the particular kind of a presupposition, in the sense of ‘kind’ intended here, is determined by the presupposition’s trigger.) However, these options and constraints must be specified somewhere in the grammar, and if this specification is not part of the presupposition representation as it figures in the preliminary representation of the clause within which the presupposition is triggered, then the specification must be given in some separate module of the grammar. (For us this would be a separate component of the construction algorithm for semantic representations.)

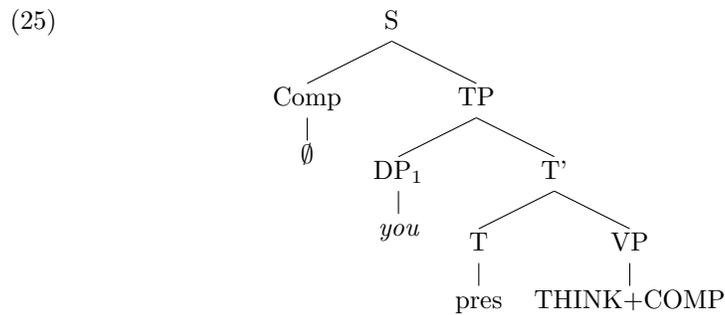
In view of what has just been said about lexical insertion of attitude verbs, the lexical insertion for the occurrence of *think* in (20) leads to the structure in (24). (‘4’ plays the role of the index ‘j’ mentioned in the one but last paragraph.)



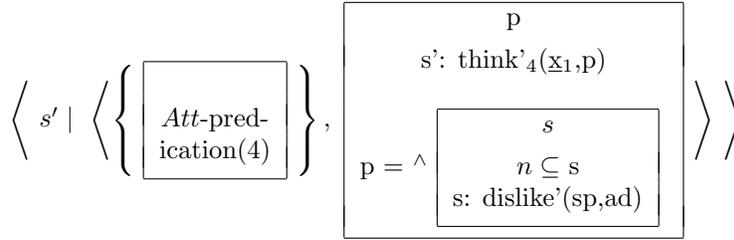
Here THINK is the structure given below:



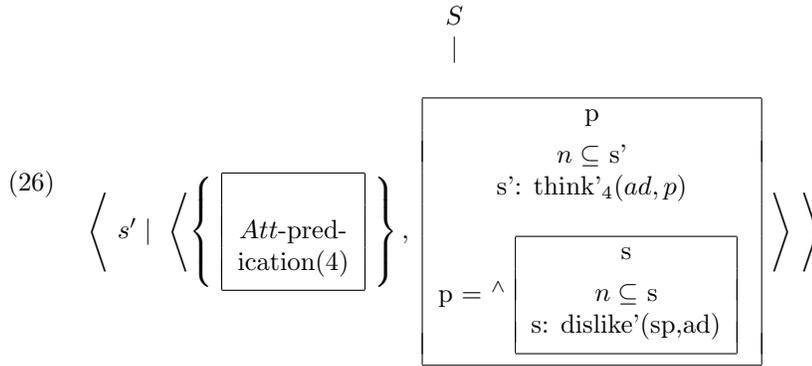
Argument insertion into the second argument slot of the predicate 'think' ', using the representation of the sister to V that is given in (23), leads to (25).



THINK+COMP is short for:



The steps leading from (25) to the representation of the S-node (and thus to the preliminary representation of the sentence) are as in the preceding constructions. In particular, *you* is once again interpreted via the indexical discourse referent *ad*. These steps lead to the S-representation in (26).

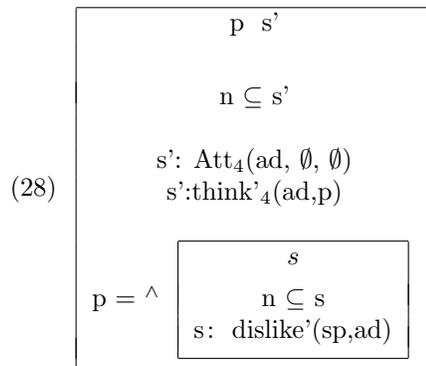


At this point, the *Att*-presupposition is in need of an *Att*-predication into which we can incorporate the attitude ascribing condition with which the presupposition is co-indexed. When a sentence like this is part of a longer discourse, which has already given rise to an *Att*-condition (in an accessible position) with the same agent and temporal location as the attitudinal predication whose predicate is coindexed with this presupposition, then the presupposition could be resolved via this *Att*-condition and the attitudinal predication condition incorporated into it. But when – as in the case before us – there is no such *Att*-condition to be retrieved, the presupposition must be accommodated in full: an *Att*-condition for the given agent and time must be added to the non-presuppositional DRS to which the presupposition was left-adjoined, while the presupposition is eliminated (since it has now been resolved), and the coindexed attitudinal predication condition is then to be ‘transferred’ into the *Att*-condition. The accommodated *Att*-condition should be as non-committal as possible; that is, it should specify the agent and location time(s) of the coindexed predication condition, but contain no information otherwise; see (27).

$$(27) \quad s' : \text{Att}_4(\text{ad}, \emptyset, \emptyset)$$

Once a suitable *Att*-condition has been retrieved or accommodated, the presupposition that is thereby resolved it can be dropped. However, it is still necessary to incorporate the attitudinal predication condition that triggered the presupposition into this *Att*-condition. So at this point a witness of the connection between *Att*-condition and attitudinal predication is still needed. We implement this by transferring the index of the presupposition to the resolving *Att*-condition. (See (28), where the index in question is ‘₄’.

(28) shows the result of accommodating the new *Att*-condition and eliminating the presupposition, but in which transfer of the attitudinal predication into the *Att*-condition is still outstanding.



To turn the transitional structure in (28) into a legitimate representation the *think*'-condition must be incorporated into the second argument of the *Att*-condition with which it is co-indexed. But this incorporation cannot be just plain transfer. The attitudinal predication must also be put into the right form. The attitudinal predicate of the attitudinal predication condition (here, the predicate ‘think’) must be recast in the form of an attitudinal mode.

The general problem here is to define a meaning preserving mapping from attitudinal predicates such as ‘think’ to mode indicators of our formalism for the representation of mental states. Using mode indicators like ‘BEL’, ‘DES’, ‘INT’ and a few others is just a first step in this direction. But for the case at hand it seems reasonable to assume that ‘think’ can be mapped to ‘BEL’.

Once that identification has been made, the remaining transformations that incorporation necessitates are straightforward. When attributions like (B1') are represented as in (26), with a discourse referent p in the object position of the verb, which at the same time is specified as denoting the content of a DRS K via the equation ‘ $p = \wedge K$ ’, we get an attitude characterization of the kind we are after by forming the pair $\langle \text{BEL}, K \rangle$. (Obviously this procedure can be generalized to attributions involving attitudinal verbs and to other predicates, which map onto other mode indicators than BEL.) In the present case this pro-

cedure leads to the condition in (29). We get our final representation by adding this condition to the second argument of the *Att*-predication. This gives us (30).

$$(29) \left\langle BEL, \begin{array}{c} s \\ n \subseteq s \\ s: dislike'(sp,ad) \end{array} \right\rangle$$

$$(30) \begin{array}{c} s' \\ n \subseteq s' \\ s': Att(ad, \left\langle \left\langle BEL, \begin{array}{c} s \\ n \subseteq s \\ s: dislike'(sp,ad) \end{array} \right\rangle \right\rangle, \emptyset) \end{array}$$

However, (30) is not the structure (10.e), repeated below, that was the target of our reconstruction of the interpretation process. In fact, it doesn't seem to come very close.

(10.e)

$$s: Att(ad, \left\{ \left\langle [ANCH, b] \begin{array}{c} s \\ n \subseteq s \\ b \ s' \\ n \subseteq s' \\ s: be-talking-to(b,i) \end{array} \right\rangle, \left\langle BEL, \begin{array}{c} s'' \\ n \subseteq s'' \\ s'': dislike'(b,i) \end{array} \right\rangle \right\}, \{\langle b, sp \rangle\})$$

When comparing (10.e) with the construction that led to (30) we see that the construction steps which produced the divergence occurred early on, when *sp* and *ad* were inserted into the argument slots of the predicate 'dislike'.

A preliminary step towards an interpretation procedure that yields a representation like (10.e) is to treat first and second person pronouns (like other definite

noun phrases) as triggers of ‘reference’ presuppositions – presuppositions to the effect that the interpreter has independent resources for identifying the referent (or semantic value) of the noun phrase.

The reference presuppositions of first and second person pronouns are always resolvable, to the speaker and the addressee of the utterance in which they occur. (This is on the assumption that there is a speaker and, in the case of second person pronouns, that there is an addressee.)

However, when such pronouns occur within the complements of attitudinal predicates, then the question arises *how* they must be seen as referring to speaker or addressee: in a *de re* or a *de se* manner. In the present set-up this means: should the presupposition triggered by such a pronoun occurrence be resolved to a *de re* or a *de se* contribution? It is because it failed to deal with this question that our construction of (30) and the result it produced are inadequate.

Here is my first shot at general principles governing the interpretation of first and second person pronouns in simple and nested attitude attributions. (For simplicity I limit attention to participant-neutral representations, but it should be clear how the principles can be extended to cover utterer and interpreter representations as well.)

1. Tokens of *I* and *you* that do not occur within the scope of an attitudinal predicate are always resolved to *sp* and *ad*. (In utterer representations *I* would resolve to *i*, while *you* would have to be interpreted via some anchored entity representation whose external anchor is the addressee; and in a representation constructed by the addressee *you* would resolve to *i*, whereas *I* would have to be resolved via some anchored entity representation for the speaker. Similar adjustments for speaker- and recipient-representations are required at some places in what follows, but I will not comment on these any further.)

2. Embedded occurrences. Let γ be an occurrence of a singular first or second person pronoun in some sentence *S* and let $\langle AP_1, \dots, AP_n \rangle$ be the maximal sequence of attitude predicates that have γ in its scope. (For pronoun occurrences that do not occur within the scope of an attitude predicate this sequence will be the empty sequence, but that case we have dealt with already.) Suppose that the referential argument which γ contributes to the semantic representation of *S* is the discourse referent β . In other words, β will be the discourse referent that identifies γ 's referent in the reference presupposition that is contributed by γ , and ‘identifying’ the referent of γ takes the form of finding an antecedent for β .

Given the way in which we construct representations of sentences with non-embedded, singly embedded and multiply embedded attitude predicates, the preliminary semantic representation of *S* will contain a sequence $\langle P_1, \dots, P_n \rangle$ of predicates that ‘translate’ the natural language predicates AP_1, \dots, AP_n , and this sequence will eventually be converted into a sequence $\langle Att_1, \dots, Att_n \rangle$ of

nested occurrences of the predicate *Att* through applications of the procedure described in connection with (25) - (30). (We haven't actually presented the representation construction for a sentence with embedded occurrences of attitude predicates, but from the representation construction for **(B1')** it should be clear what happens when the construction principles we used are applied to sentences with attitude embeddings.) Let $\alpha_1, \dots, \alpha_n$ be the discourse referents occupying the first argument positions of the predicate occurrences Att_1, \dots, Att_n . Thus, α_n represents the bearer of the attitudinal state reported by the most deeply embedded attitude predicate that contains γ within its scope, α_{n-1} the bearer of the next occurrence upwards of *Att* that contains Att_n in its scope (and thus also α_n and γ), and so on. It is α_n that plays a crucial role in determining the possible resolutions of the reference presupposition contributed by γ .

The principles that follow apply to the singular first person pronoun *I* (including its morphological variants *me*, *my* and *mine*). Similar principles can be formulated for the second person pronoun *you*.

(31)

- If α_n is itself the referential argument of a first person pronoun, then the strongly preferred resolution of β is to *i*; that is, γ is given a *de se* interpretation. (There are also cases in which it does seem possible to interpret first person pronouns occurring within the immediate attitudinal scope of a first person pronoun in a non-*de se* way. On the whole this appears to be easier for the possessive pronouns *my* and *mine*, but for the other first person pronoun forms such interpretations seem possible as well when the utterance context favours such an interpretation. There are various factors that make it easier to get such non-*de se* interpretations. Among them are: (i) the setting in which the utterance is made: non-*de se* interpretations are easier to come by when the utterance is not part of a face-to-face exchange; (ii) the tense of the uttered sentence: past tense sentences more easily allow for non-*de se* interpretations than present tense ones. But this is a matter that needs to be looked at more closely than I have.)
- If α_n is not the referential argument of a first person pronoun, then the only possible interpretation of γ is *de re*. That is, β must be resolved to the discourse referent δ of an anchored entity representation Δ of the agent represented as α_n , but with the requirement that δ be *directly or indirectly anchored* to the speaker.

Here by 'directly or indirectly anchored to the speaker' I mean the following. If $n = 1$ (i.e. if γ is only in the scope of one attitudinal verb), then the anchoring of δ must be 'direct', i.e. the external anchor of δ must be the discourse referent *sp* (which represents the utterer from a participant-neutral point of view). If $n > 1$, then there are two possibilities. (i) If

α_{n-1} is the referential argument of an occurrence of a first person pronoun, then d can be (and typically will be, but see the cautionary remark two paragraphs above) i ; in this case the referential presupposition triggered by γ has been fully resolved; the buck is passed, so to speak, to the first person pronoun occurrence that has given rise to α_{n-1} . (ii) If α_{n-1} is not the referential argument of a first person pronoun, then the anchoring of δ must be ‘indirect’: the external anchor of δ will now in its turn be the discourse referent δ' of an anchored entity representation, which occurs as a member of the second argument of the *Att* occurrence Att_{n-1} ; and the requirement of direct or indirect anchoring to the speaker is passed on to δ' .

Eventually this procedure will lead to a chain of anchored entity representations $\Delta_n, \Delta_{n-1} \dots, \Delta_k$ ($0 < k < n+1$) with representing discourse referents $\delta_n, \delta_{n-1} \dots, \delta_k$ such that

- (i) β is resolved to δ_n ;
- (ii) for $j = n, \dots, k-1$, δ_j is resolved to δ_{j-1} ;
- (iii) δ_k is either i or sp
- (iv) when $n > k$, then none of the discourse referents δ_j for $j > k$ is either i or sp .

The interpretation principles for singular 2nd person pronouns are the exact mirror images of those given above for pronouns in the 1st person. Once again, an occurrence of a 2nd person pronoun within an attitudinal context can be interpreted as i only when the agent argument of the attitudinal predication is also an occurrence of a 2nd person pronoun. Otherwise the interpretation must involve an anchored representation of the pronoun’s referent (i.e. of the addressee) with the requirement that the external anchor for that representation be directly or indirectly anchored to the addressee.

It is not difficult to see that this algorithm interprets all occurrences of first person pronouns as referring to the speaker (and, by parity, all occurrences of second person pronouns as referring to the addressee). For the cases where $n = 0$ this follows directly from the description. And for the cases where $n > 0$ it is clear as well. When δ_k is sp , then all δ_j with $j > k$, and eventually β , are anchored to sp and thus to the speaker of S . And when δ_k is i , then this means that δ_{k-1} is the discourse referent contributed by a first person pronoun occurrence (occurring as agent argument to some attitudinal predicate in S); whether β is (indirectly) anchored to the speaker now depends on the direct or indirect anchoring to the speaker of δ_{k-1} . There is an element of recursion here; but it should be obvious that this recursion is well-founded. (The chain of nested embedding within attitude contexts of the pronoun occurrence that is responsible for the introduction of δ_{k-1} is clearly shorter than the chain of nested embeddings of our pronoun occurrence γ .)

Two illustrations of how the algorithm works are given below. The first is a detailed construction of the participant-neutral representation of (**B1'**) and the second a construction of the participant-neutral representation of (**B2'**). The presentations of these constructions differ in that the first presents more of the details. The presentation of the construction for (**B2'**) focusses exclusively on the interpretation of the 1st and 2nd person pronoun occurrences. Because of the extra level of attitudinal embedding in (**B2'**) this construction involves an additional complication, and it is that complication to which the presentation will mostly pay attention.

First, then, the construction of the semantic representation of (**B1'**) in which the first and second person pronouns are handled according to the guidelines mentioned above. The final result is now a representation that is for practical purposes equivalent to (10.e).

Following the practice of left adjoining the set of presuppositions generated within a given clause to the DRS representing its non-presuppositional content we now get as representation for the embedded VP the structure in (32).

$$(32) \left\langle s, v \mid \left\langle \left\{ \boxed{v_{s2pr}} \right\}, \boxed{s : dislike'(x_1, v)} \right\rangle \right\rangle$$

Interpretation of the present tense information under T is as before. The subject DP *I* of the complement clause adds a further referential presupposition, so we get as representation for the embedded TP the one in (33).

$$(33) \left\langle s, v, u \mid \left\langle \left\{ \boxed{v_{s2pr}}, \boxed{u_{s1pr}} \right\}, \boxed{s : dislike'(u, v)} \right\rangle \right\rangle$$

The complementizer of the direct object DP of *think* now operates on the non-presuppositional DRS of (33), after transfer of the discourse referent *s* from the store to the universe of that DRS:

$$(34) \left\langle p, v, u \mid \left\langle \left\{ \boxed{v_{s2pr}}, \boxed{u_{s1pr}} \right\}, \boxed{p = \wedge \begin{array}{c} s \\ s : dislike'(u, v) \end{array}} \right\rangle \right\rangle$$

The next steps are as before, leading to the representation in (35). To simplify matters the subject *you* of *think* has been resolved to *ad* 'on the fly', instead of first introducing a reference presupposition for it. (35) just shows the result of the resolution (in the form of the equation 'w = ad').

$$(35)$$

$$\begin{array}{c}
S \\
| \\
\left\langle s' \mid \left\langle \left\{ \begin{array}{c} \text{Att-pred-} \\ \text{ication(4)} \end{array} \right\} \right\rangle, \left\langle v, u \mid \left\langle \left\{ \begin{array}{c} v_{s2pr} \\ u_{s1pr} \end{array} \right\}, p = \wedge \begin{array}{c} s \\ n \subseteq s \\ s : \text{dislike}'(u, v) \end{array} \right\rangle \right\rangle \right\rangle
\end{array}$$

By creating a new *Att*-condition and incorporating the think'-condition into it, in the way described in the course of our first construction attempt, (35) gets transformed into (36).

$$(36) \quad \left\langle s' : \text{Att}(w, \left\langle \left\langle \text{BEL}, \left\langle v, u \mid \left\langle \left\{ \begin{array}{c} v_{s2pr} \\ u_{s1pr} \end{array} \right\}, \begin{array}{c} s \\ n \subseteq s \\ s : \text{dislike}'(u, v) \end{array} \right\rangle \right\rangle \right\rangle, \emptyset \right) \right\rangle$$

What remains is the resolution of the pronoun presuppositions in (36).

Since the first argument of *Att* stems from a 2-nd person pronoun, the default resolution of *v* is to *i*.

The resolution of *u* is different. Since the first argument of *Att* does not represent a first person pronoun, *u* cannot resolve to *i*. Instead, it must be assumed that the bearer *w* of the attitude has an anchored representation for the referent of *u*. Since the given occurrence of *Att* is also an outermost occurrence ($n = 1$), the external anchor of this representation must be *sp*.

Resolving the two presuppositions along these lines leads to the representation in (37).

$$(37) \quad \left. \begin{array}{c} s', w \\ n \subseteq s' \quad w = ad \\ \left\langle [ANCH, u], \boxed{u} \right\rangle \\ \left\langle BEL, \begin{array}{c} s \quad v \\ n \subseteq s \quad v = i \\ s: dislike'(u, v) \end{array} \right\rangle \end{array} \right\}, \{ \langle u, sp \rangle \}$$

We can simplify (37) by replacing the argument-filling occurrences of the discourse referents w and v by their resolutions ad and i :

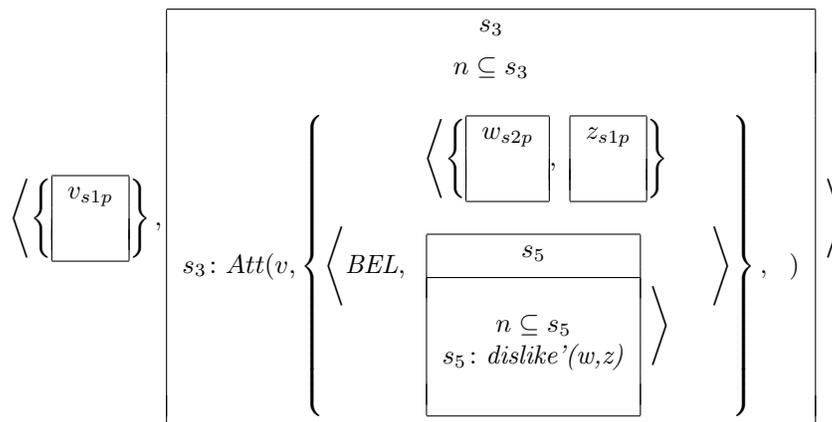
$$(38) \quad \left. \begin{array}{c} s' \\ n \subseteq s' \\ \left\langle [ANCH, u], \boxed{u} \right\rangle \\ \left\langle BEL, \begin{array}{c} s \\ n \subseteq s \\ s: dislike'(u, i) \end{array} \right\rangle \end{array} \right\}, \{ \langle u, sp \rangle \}$$

But for the choice of discourse referents and the missing specification of the internal anchor in (38) this is the same representation as (10.e).

We now come to the representation construction for **(B2')**. This time we assume that the preliminary representation for the sentence has been constructed already, with all the presuppositions triggered by the pronoun occurrences explicitly displayed. This representation is similar to the one in (10.h), except for the pronoun presuppositions. Again we split the representation in two for reasons of space.

$$(39) \quad \text{a.} \quad \left\langle \left\{ \boxed{u_{s2p}} \right\}, \begin{array}{c} s \\ n \subseteq s \\ s: Att(u, \{ \langle BEL, K' \rangle \},) \end{array} \right\rangle$$

b. The structure K' :



Resolution of the pronoun presuppositions is inside out. So we start with those for w and z . Decisive is that w was introduced by a 2nd person pronoun and z by a 1st person pronoun. Since the agent argument of the *Att*-predication that contains w and z in its immediate scope (the one in K') stems from a 1st person pronoun, the default resolution for the presupposition with z is to i . On the other hand, lack of agreement between the agent argument and the pronoun that gave rise to w entails that the w -presupposition must be resolved via an anchored entity representation for the referent of w , which has to be accommodated as part of the inner *Att* attribution. This accommodation consists of two parts. First an internally anchored entity representation has to be added to the second argument of *Att*. (We reuse w as discourse referent of this entity representation.) And second, an external anchor $\langle w, w' \rangle$ has to be added to the third argument of *Att*, with the constraint that w' be directly or indirectly linked to the addressee. It is natural to encode this constraint by means of a kind of presupposition that is introduced by pronouns of the given kind, viz. those triggered by singular 2nd person pronouns. Since w' ‘intrudes’ so to speak into the attitude predication form the outside, its presupposition is to be resolved at a higher level. This is captured by left adjoining the presupposition to the DRS containing the given *Att*-predication condition, so that it comes to form a presupposition set with the v -presupposition that is already present at this level.

Resolution of the presuppositions for w and z thus leads to the following modification K'' of K' .

$$\left\langle \left\langle \boxed{v_{s1p}}, \boxed{w'_{s2p}} \right\rangle, \right. \\
(40) \quad \left. \left. \left. \begin{array}{c} s_3 \\ n \subseteq s_3 \\ \left\langle [ANCH, w], \boxed{w} \right\rangle \\ \left\langle BEL, \begin{array}{c} s_5 \\ n \subseteq s_5 \\ s_5: dislike'(w, i) \end{array} \right\rangle \end{array} \right. \right. \left. \right\rangle, \{ \langle w, w' \rangle \} \rangle$$

N.B. We could also adopt the additional principle according to which the internal anchor of an anchored representation for an occurrence of *you* that is part of an *Att* predication whose agent represents the speaker always takes the form we assumed earlier, viz. as containing a condition to the effect that entity represented is the person who stands to the speaker in the ‘addressee relation’ – that relation which holds between a speaker and the person she has chosen to address and for whom the speaker has a representation which reflects her perceiving him *as* the person she is addressing. And we can represent this relation in the same way as before, in the representations displayed in (10). In that case the anchored representation for *w* will take the form in (41).

$$(41) \quad \left\langle [ANCH, w], \begin{array}{c} \boxed{w \ s_4} \\ n \subseteq s_4 \\ s_4: be-talking-to(i, w) \end{array} \right\rangle$$

The next resolution steps are those of the presuppositions for *v*” and *w*’. The next higher *Att* predicate, which has these discourse referents in its immediate scope, is the outer one of the two in (39), displayed in (39.a). The agent argument of this *Att* predication stems from a 2nd person pronoun. So now it is *w*’ for which the default resolution is to *i*, whereas *v* requires accommodation of an anchored entity representation in the second argument slot of the outer occurrence of *Att*, together with an external anchor $\langle v, v' \rangle$ in its third slot. *v*’ now gives rise to a 1st person pronoun presupposition, which is left-adjoined to the non-presuppositional DRS of (39.a), where it forms a presupposition set jointly with the presupposition on *u*. The resulting representation is shown in (43.a,b).

$$(42) \text{ a. } \left\langle \left\langle \left[\begin{array}{c} u_{s2p} \\ \hline \end{array} \right], \left[\begin{array}{c} v'_{s1p} \\ \hline \end{array} \right] \right\rangle, \left. \begin{array}{c} s \\ n \subseteq s \\ s: \text{Att}(u, \left\langle \left[\text{ANCH}, v \right], \left[\begin{array}{c} v \\ \hline \end{array} \right] \right\rangle \right\rangle, \langle v, v' \rangle \\ \left\langle \text{BEL}, K''' \right\rangle \end{array} \right\rangle$$

b. The structure K''' :

$$\begin{array}{c} s_3 \\ n \subseteq s_3 \\ s_3: \text{Att}(v, \left\langle \left[\text{ANCH}, w \right], \left[\begin{array}{c} w \\ \hline \end{array} \right] \right\rangle, \left. \left\langle \text{BEL}, \left[\begin{array}{c} s_5 \\ n \subseteq s_5 \\ s_5: \text{dislike}'(w, i) \end{array} \right] \right\rangle \right\rangle, \{ \langle w, i \rangle \} \end{array}$$

Neither of the remaining presuppositions is embedded within an attitudinal predication. So they are resolved to the indexical discourse referents sp and ad of participant-neutral representations – u to ad and v to sp . With these resolutions the final version of (43.a) is as in (43). K''' remains as it was.

$$(43) \left. \begin{array}{c} s \\ n \subseteq s \\ s: \text{Att}(ad, \left\langle \left[\text{ANCH}, v \right], \left[\begin{array}{c} v \\ \hline \end{array} \right] \right\rangle \right\rangle, \langle v, sp \rangle \\ \left\langle \text{BEL}, K''' \right\rangle \end{array} \right\rangle$$

I hope that these illustrations help to see how the algorithm I have assumed here for the processing of 1st and 2nd person pronouns works for singly and multiply embedded occurrences of such pronouns and that the reader will be able to extrapolate these lessons to sentences in which the chains of nested attitude attributions are even longer, as for instance in ‘You think that I think that

you think that I dislike you.’ and so on. I should add, though, that as far as I am concerned, intuitions about the possible interpretations of such sentences become quite shaky as soon as we reach embedding depths of three and more. To the extent that the results I obtain by slavishly applying the algorithm proposed here to the interpretation of such sentences look right to me, that is probably more a reflection of my conviction that the algorithm does the right thing than on the basis of an independent understanding of the meanings of such sentences as a natural language speaker.

It should be noted that the algorithm presented here applies also to occurrences of 1st and 2nd person pronouns in attitude attributions whose agent argument is a 3rd person DP. With such agent phrases there is failure of agreement between the agent phrase and pronoun irrespective of whether the pronoun is 1st or 2nd person, so in either case we get an interpretation involving an anchored entity representation accommodated into the representation of the attitude attribution.

A different question is what should be said about the occurrences of 3rd person pronouns in attitude attributions. One would expect that our algorithm could be extended to such pronouns too, following in the wake of Castaneda’s treatment of them as ‘quasi-indexicals’. Some preliminary results on the possibility of extending the algorithm in this direction are presented in the next section.

One moral of this analysis is that it confirms the status of first person singular pronouns as ‘indexicals’: whether such a pronoun occurs within the scope of one or more attitudinal predicates or not, it will always end up as referring, in one way or another, to the producer of the utterance of which it is part. It will either do this in a direct way, by being itself interpreted as *sp* – in the semantic representation of the sentence *sp* is inserted into the argument slot occupied by the pronoun in the represented sentential utterance. Or it will do so indirectly, and that it can do in one of two ways: either (i) it is interpreted as *i* while being in the immediate scope of an attitude attribution whose agent is another occurrence of a first person pronoun, and thus as involving a self-attribution of the referent of that other first person pronoun occurrence, or (ii) its interpretation is connected with such another first person pronoun occurrence via a chain of anchored entity representations where each next one provides the external anchor for its predecessor; in this second case our first pronoun occurrence still plays the part of conveying a self-attribution by the referent of the second occurrence, though of the indirect kind, as when a person A attributes to himself the property that some other person B has such or such an opinion about her. That this contribution of our first pronoun occurrence still is a case of referring to the speaker of S then depends, as we have seen on the fact that this is so for the second occurrence. But, as argued, our algorithm guarantees that this will be so.

5 An example with third person agent phrases

Maier (2009) discusses the examples in (44).

- (44) a. John believes that Mary thinks Peter is cool.
b. John believes that Mary thinks I'm cool.

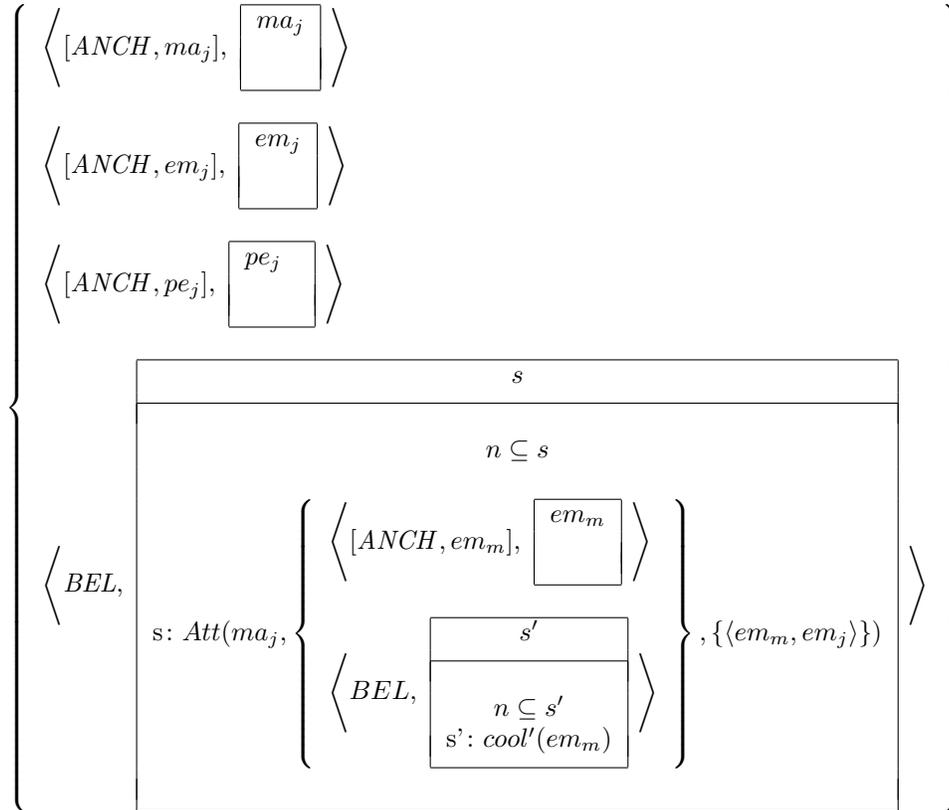
Maier considers a situation in which Mary has just said to John: ‘That guy is cool.’, pointing in the direction where the speaker (of either (44.a) or (44.b)) and some other person called ‘Peter’ are standing, and in which John takes Mary to be pointing at the speaker as she is saying this, whereas she is in fact pointing at Peter. Thus John gets the wrong message, and forms the belief that it is the speaker that Mary thinks is cool, a belief that the speaker could then express by means of (44.b). But at the same time there is a sense in which the speaker’s utterance of (44.b) could be regarded as making a true belief attribution to John: For he does have the belief that Mary thinks that the guy is cool at whom she has just pointed, and that guy *is* Peter (and not the speaker).

Maier’s discussion of these examples is focussed on the question how an intuitively correct semantics for the utterances in (44) can be obtained in a strictly compositional manner from plausible syntactic structures for the sentences in (44.a,b), taking up an insightful discussion of the compositional semantics of attitude reports by Von Stechow and Zimmermann (Stechow & Zimmermann (2004)). Von Stechow and Zimmermann provide a formal reconstruction of a proposal for the semantics of *de re* attitude reports that they extract from (Kaplan (1989)). One of the observations they make is that a tenable version of this proposal requires a syntactically poorly motivated form of movement – so-called ‘*res*-movement’: the DP that is part of the complement of the attitudinal verb and that contributes the referent with respect to which the report is interpreted as *de re* must be moved at the level of If (‘logical form’ in the syntactic sense of Chomsky) to a higher position from which it has scope over the verb. Maier’s discussion of (44) leads him to the conclusion that this problem becomes even more pronounced with iterated attitude reports like those in (44), where multiply embedded DPs (such as *Peter* in (44.a) and *I* in (44.b)) may have to be moved across the scope of two or more attitudinal verbs.

The framework we are using here makes it difficult to address the compositionality issue in the way it is raised in (Stechow & Zimmermann (2004)) and (Maier (2009)). I shall have something to say about this below. But before we can do that – i.e. before we can discuss the principles according to which the representations of the utterances in (44) can be constructed – we first need to address the question what these semantic representations are like. As could be seen from our discussion of the examples in (9), this matter is more involved in our framework than it is in others, since we are distinguishing between the semantics of report sentences and the mental states of conversation participants that use those sentences in communication.

(45) is a first shot at a representation in the formalism we have been using of the relevant part of the mental state of John's that results from his (mis-) interpretation of what Mary has just tried to convey to him. It consists of three anchored entity representations and a belief, and this belief attributes to Mary the belief/opinion that Emar is cool. More precisely, John believes Mary to be in a mental state which contains an anchored representation for the individual that John himself has represented by means of the discourse referent ma_j and a belief to the effect that the referent of that anchored representation is cool. (What (45) does not (and cannot) express is whether the anchored entity representations it attributes to John are properly externally anchored. But we may assume that that is so: the representation with discourse referent ma_j is externally anchored to Mary, the one with em_j is externally anchored to Emar and the one with discourse referent pe_j is externally anchored to Peter.

(45)



In (refn34) the internal anchors of the entity representations for Mary, Emar and Peter have been left blank. This option has been chosen as a way of stay-

ing neutral on the question exactly at what point John’s mental state is being considered. Immediately after Mary has made her utterance his anchors for her, Emar and Peter the internal anchors of these representations might well be perceptual, since at that point John can see each of the three; or at least his anchors might have a perceptual component; but if John’s state is considered some time after the speech event that Maier describes, the four may have parted company and John may no longer be in direct perceptual contact with some or all of Mary, Emar and Peter. He may still be expected to have anchored representations for them; after all he has seen them before and the memory of those perceptual experiences would be enough to support his anchored representations of them. But such details about the nature of the anchors of these three entity representations are immaterial to what I will have to say about the case and so it is best not to make any commitments on this point.

(refnn34) can also be used in semantic representations of the statement (refnn33.b) and in a representation of the relevant parts of the mental state of Emar that would motivate him to make this statement and which he presumably wanted his statement to express. These representations are displayed in (refnn35.a,b) below. (The representation of Emar’s utterance (refnn33.a) and of the mental state that would presumably have been responsible for making it are not displayed here. They are almost completely analogous to those in (refnn35), but are based on a representation of the relevant part of John’s mental state according to which he attributes Mary’s opinion to Peter.)

require at the di thatcase of (refnn33.a) and the state that might have been responsible is analogous.)

(46)

a.

$s'' \quad j \quad m \quad p$
$n \subseteq s'' \quad \text{John}(j) \quad \text{Mary}(m) \quad \text{Peter}(p)$
$s'': \text{Att}(j, (45), \{\langle ma_j, m \rangle, \langle em_j, sp \rangle, \langle pe_j, p \rangle\})$

b.

$$\left\{ \begin{array}{l}
 \left\langle [ANCH, ma_{em}], \boxed{ma_{em}} \right\rangle \\
 \left\langle [ANCH, jo_{em}], \boxed{jo_{em}} \right\rangle \\
 \left\langle [ANCH, pe_{em}], \boxed{pe_{em}} \right\rangle \\
 \left\langle BEL, \boxed{\begin{array}{c} s'' \\ n \subseteq s''' \\ s''': \text{Att}(jo_{em}, (45), \{\langle ma_j, ma_{em} \rangle, \langle em_j, i \rangle, \langle pe_j, pe_{em} \rangle\}) \end{array}} \right\rangle
 \end{array} \right\}$$

(In (46.a) and (46.b) ‘(45)’ abbreviates the representation given under (45).)

The problem with these representations is that they fail to give any indication of what went wrong in the situation Maier describes. As it stands, the formalism we have been using is not equipped for that. To come closer to a representation of what went wrong in the communication between Mary and John we need two further devices, which have not been used in this note so far. The first is that of a *vicarious anchor* (Kamp & Bende-Farkas (2006)) and the second that of a *multiply anchored* entity representation. Vicarious anchors are the internal anchors that utterance recipients form when they interpret a noun phrase in the utterance they are interpreting as used by the speaker in order to refer to some particular entity, for which the speaker herself has an anchored representation. The interpreter can then represent this entity by an entity representation that is ‘designed’ to stand for whatever it is that the speaker used the noun phrase in question to refer to. The interpreter’s entity representation is designed to refer to that entity in that its internal anchor describes the entity as the one that the speaker used the noun phrase to refer to (and for which she has an anchored representation which was instrumental in her use of the noun phrase). Internal anchors of this type are called ‘vicarious’. A vicarious anchor presupposes that the speaker did in fact use the given noun phrase in the way it indicates; and when this presupposition is true, it renders the entity thus referred the external anchor of the entity representation of which the given vicarious anchor is the internal anchor. An example, taken from (Kamp & Bende-Farkas (2006)), is the second anchored representation of (47), the representation that is formed by a university teacher B, who, when he finally makes it to the Department by 1.45,

is told by his colleague A ‘A student was looking for you this morning’ and who takes her use of *a student* as a way of referring to a student of whom she has an anchored entity representation (most likely on the basis of having encountered the student that morning as the student was looking for B).

$$(47) \left\{ \left\langle \left[\text{ANCH}, x_B \right], \left\langle \left[\text{ANCH}, a_B \right], \left[\begin{array}{c} a_B \\ \text{talk-to}(a_B, i) \end{array} \right] \right\rangle \right\rangle, \right. \\ \left. \left\langle \left[\text{ANCH}, x_B \right], \left. \begin{array}{c} x_B \quad x \\ \text{Att}(a_B, \left\{ \left\langle \left[\text{ANCH}, x_A \right], \left[\begin{array}{c} x_A \end{array} \right] \right\rangle \right\}, \{ \langle x_A, x \rangle \}) \\ \text{Realise}(a_B, x_A, \text{'a student'}) \\ x_B = x \end{array} \right. \right\rangle \right. \\ \left. \left\langle \left[\text{BEL}, \left[\begin{array}{c} t \quad e \\ \text{'this morning'}(t) \quad t < n \quad e \subseteq t \quad \text{student}'(x_B) \\ e: \text{look-for}'(x_B, i) \end{array} \right] \right\rangle \right. \right\}$$

As this example shows, vicarious anchors attribute to the speaker (or author) an anchored representation for the entity that the speaker/author chose to refer to by means of the noun phrase she selected for this purpose, and they express the 3-place relation between speaker, entity representation and the chosen NP. We treat this relation as a primitive and denote it with the predicate ‘realise’. In (47) the noun phrase in question is an indefinite NP; but on the whole it is more often definite NPs that give rise to vicariously anchored representations. (Arguably the most prominent type of NP that can lead to the formation of vicariously anchored representations are proper names. This is what leads to the causal chains familiar from the work of Kripke and Chastain (Kripke (1972), Chastain (1975)).)

In the case we are discussing it is plausible to assume that John’s observing Mary’s pointing in the direction of Peter and Emar while saying ‘That guy is cool.’ leads him to form a vicariously anchored representation for the individual to which Mary is referring through her use of *that guy*. But the present situation differs from the one of example (47) in that John also has a direct perceptual access to the individual that he takes Mary to be referring to by means of *that guy*. In other words, John has two modes of access to this individual, one via Mary’s utterance and one via his own visual perception. Or, rather, that would have been the natural way to describe the case if there had been no conflict between the individual that John perceptually identifies as the person that Mary

points at, and that she thus must be referring to, and the person that she does in fact point at and refer to and which thereby is the person that John can be construed as taking Mary to refer to. In such a situation – in which Mary points at Peter and John identifies Peter as the person she is pointing at, or in which Mary would have pointed at Emar and in which John would have taken her to point at Emar – it would be natural to describe John’s reaction as involving the formation of a *doubly anchored* entity representation, one with two internal anchors, one vicarious and one based on visual perception, but both anchoring the representation to the same external anchor (Peter in the first case, Emar in the second). Such multiply anchored entity representations are very common. In fact, most things that we come into contact with repeatedly are things for which we have multiply anchored representations. What the different kinds of forms are that multiply anchored representations can take is a question to which I have no general answer as things stand. But for certain cases, in particular most cases of doubly anchored representations, which involve just two lines of access of the agent to the entity that is represented, a natural answer is suggested by what has been said so far: A doubly anchored entity representation is an entity representation with a pair of internal anchors.

For a (singly) anchored representation to be proper, it must have an external anchor and its internal anchor should articulate the way in which its bearer is causally related to this external anchor (i.e. to the entity that the representation is intended to represent). When this condition is not fulfilled, the entity representation is improper and the content of any propositional representation that uses the entity representation as constituent is undefined. But when an anchored entity representation is proper, then its internal anchor can be seen as determining its external anchor uniquely.

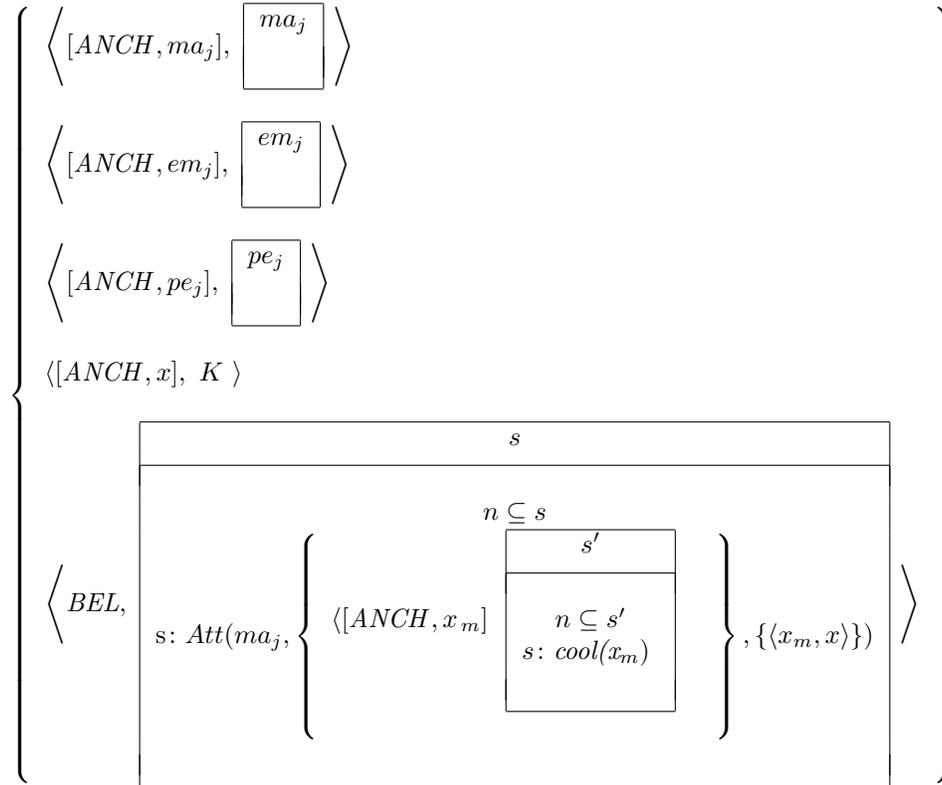
The same possibility of failure that arises for singly anchored entity representations can arise for doubly anchored entity representations as well. But with doubly anchored representations there is another source of impropriety, viz. when the two internal anchors of a doubly anchored entity representation determine *distinct* external anchors. Doubly anchored entity representations for which there is this kind of divergence may be called *incoherent*. It is such an incoherent doubly anchored entity representation that John ends up with on account of his interpretation of Mary’s utterance and the pointing that accompanies it in the situation as Maier describes it: his vicarious anchor for the individual that Mary refers to points to Emar, but his perceptual anchor links him to Peter.

Can an incoherent entity representation be a proper basis for a propositional attitude based on it, such as, in the case at hand, the beliefs that (44.a,b) attribute to John? Strictly speaking, I am inclined to say, no. But that judgement seems to fly in the face of what I have been saying so far, following Maier, about (44.b). The representation (45) of John’s state of mind that I used to give formal expression to that assumption involves a singly anchored representation

for the subject of Mary's attribution, with Peter as its external anchor. And it was on the basis of that way of representing John's state of mind that we could then justify Emar's belief attribution to John as true. But can (45) still be considered adequate in the light of what we have been saying about John's two distinct ways of access to the individual Mary pointed at and the doubly anchored representation that his observation of what Mary did – what she said and at whom she pointed – should have led to?

If our pretheoretical intuitions about (44.b) are correct, then it should still be possible to justify this attribution. But what form should this justification take? The best answer I can suggest is that the description of the mental state of John's that results from his observing Mary need not be described as involving a single doubly anchored representation for the person Mary is pointing at and speaking about, but that this state may also be described as involving two singly anchored entity representations, one with a perceptually based internal anchor and Emar as external anchor and a second one with a vicarious anchor and Peter as external anchor. With these two anchored representations John associates the (false) belief that they represent the same individual; and it is of 'that individual' that John believes that Mary thinks 'he' is cool. (48) is a representation of John's mental state drawn up in this spirit, with (i) anchored representations for Mary, Emar and Peter (as in (45)), (ii) a vicarious anchor for the person to whom Mary has attributed coolness, (iii) the erroneous belief that the referent of this vicariously anchored representation is identical with the referent of his perceptually or memory-based anchored representation of Emar, and (iv) a belief to the effect that Mary attributes coolness to the individual she referred to with *that guy*, which is the entity represented by John's vicariously anchored representation.

(48)



What can (nn37) tell us about who it is that we should say John believe that Mary thinks *s* cool? That isn't immediately clear. When we consider just the part of John's mental state depicted in (nn37) that consists of his belief attribution to Mary and the vicariously anchored representation on which it is based, the conclusion would seem to be that his belief is correctly described by (44.a). It is only when we take the false belief that ' $x_j = em_j$ ' into account as well that we can see how, by combining the belief directly derived from Mary's words with this false belief John could have arrived at a belief that can be characterised by (44.b).

There is something about this assessment, however, that seems to go against the grain. I personally feel a strong inclination to see (44.b) as the more correct (and perhaps even as the only correct) characterisation of John's belief and not (44.a) - and it is my impression that Maier is inclined this way too. I think the main reason why I have this inclination is an observation Maier makes about the case: When you would ask John about who it is Mary thinks is cool, he would no doubt answer: 'Emar'. And he might also go up to Emar spontaneously to tell him how cool Mary thinks he is. (Or, if he himself happens to have a crush

on Mary, it will be Emar that he will now be jealous of.) Peter won't come into it in any way. It will be only when we make John aware of his mistake that we can hope to elicit a grudging admission that, yes, in some sense it *was* Peter that he thought (or should have thought?) Mary thinks is cool. Thus, in so far as John's own verbal and non-verbal behaviour is our most telling evidence for what he believes, things don't look good for (44.a).

Of course, when it is pointed out to John that he has misinterpreted Mary's pointing – that she wasn't pointing at Emar but at Peter – he will revise his belief and now attribute to Mary the opinion that Peter is cool. This revision will take the form of replacing the false belief that ' $x_j = em_j$ ' by the true belief that ' $x_j = pe_j$ '. It restores coherence to John's belief state, which thereby has lost its interest for the problems Maier puts before us.

So let us return to (48). What does it enable us to determine about what John can be said to believe? Is it really true after all that both (44.a) and (44.b) can be considered correct belief attributions to John? And in what sense could they be so considered? As a matter of fact, the form of (48) suggests that there is another question which we should address first: Is there perhaps any sense in which John might be regarded as holding two beliefs in the situation we are considering, one to the effect that Mary thinks Peter is cool and another that she thinks Emar is cool. Or does he have just one belief that, perhaps, could be characterised in either of the two ways?

If (48) suggests this question, it also suggests the answer to it: There is only one belief. For there is only one belief state about who Mary thinks is cool – one component of John's attitudinal state over-all that deals with this issue – that he has formed on the basis of his perception of Mary's speech act. Since he is convinced that Mary did refer to Emar in using the phrase (*that guy* he will, you might say, *interpret* his belief as attributing to Mary the opinion that Emar is cool. If there are also grounds for saying that John has a belief to the effect that Mary thinks Peter is cool, that claim can only be understood as the claim that this same belief state can also be characterised in these alternative terms.¹

How could it be that the same belief state can be characterised in terms as different (and on the face of it incompatible) as these? To the extent that a justification of this can be given at all, it should, I think, go something like this. (44.a) is a *de jure* characterisation of John's belief state, while (44.b) is a kind of *de facto* characterisation. The *de jure* characterisation is based on the principle that it characterises the belief that John should have obtained purely on the strength of Mary's utterance, and that he would have obtained if there

¹This view of the individuation conditions of the attitude constituents of mental states are akin, if I read them correctly, to the *originalist* position defended by Sainsbury and Tye in their forthcoming 'Seven Puzzles' [reference]

hadn't been the concomitant confusion about Mary's pointing, and the belief resulting from it that ' $x_j = em_j$ '. That belief muddies the waters in that it seduces John into taking his belief to be about Mary's opinion of Emar. But erroneous though this way in which John interprets his own belief may be, it is this interpretation, and with it the belief attribution to John made in (44b), it is this interpretation, we have already noted, which determines John's behaviour. In fact, on a functional role account of attitudes such as belief, desire and intention it would clearly be (44.b) that gives the correct description of John's belief.

Our discussion has reached a curious point. On the one hand it now looks as if the 'causally correct characterisation of John's belief is the one given by (44.a). But it is (44.b) that accounts for the effect that John's belief has on his actions. How can that be? (48) suggests the following diagnosis. At the heart of the conflicting belief attributions (44.a,b) are two anchored representations for the person John thinks Mary thinks is cool, the vicariously anchored representation derived from her use of *that guy* and the perception-based representation that John has for Emar. (44.b) rests on John's identification of these two, (44.a) does not. That might suggest a greater legitimacy for (44.a) than for (44.b), if it weren't for a factor that we have not so far mentioned, viz. the difference between these two anchored entity representations in what might be called 'cognitive strength'. Perception typically gives access to rich complexes of information about what is perceived, information that enables the perceiver to recognise the perceived person or object on future encounters. Such information gets associated with the perceptually anchored representations to which perception gives rise; it is this kind of information that endows such representations with their distinctive cognitive strength.

Vicariously anchored representations are not like that. They ride piggy-back on the presumed anchored representations of someone else. In and of itself that gives the possessor of such a representation very little to go by. In fact, vicariously anchored representations tend to act as invitations to get more information about the referents they represent; they are enigmas clamouring for a solution. That is why speakers will so often try to make it easier for their interlocutors to resolve those enigmas. One way to achieve this is to point at the object or person one is referring to in what one is saying: the pointing enables the addressee to form a perception-based anchored representation of the entity he sees the speaker point at and to identify its referent with that of his vicarious anchor for the person or thing that the speaker is referring to in speech.

However, if that is the point of pointing to the thing you are referring to in what you say, then what John does in the scenario that Maier describes, is pretty much what Mary wants him to do: To give more informational substance to his anchor for the person that she is referring to by means of *that guy*. So the anchored representation he ends up with for the person of whom he now believes she thinks is cool is much like the kind of representation she intended

him to end up with – except of course that he connected his vicariously anchored representation with the wrong perceptually anchored representation.

To conclude: The reason why we feel that (44b) captures John’s belief correctly (or at any rate more correctly than (44a)) is not only that (44.b) makes sense as part of an explanation of John’s behaviour, while (44.a) does not. It is also because we feel that in identifying the vicariously anchored representation John derives from processing Mary’s words with an entity representation that John derives from Mary’s pointing gesture, which has the informational richness that the first representation lacks, he is complying with what she intends him to do. The identification leads to a compound representation with the informational richness that she means him to get. In the case at hand, the perception based representation that the vicariously anchored representation gets identified with turns out to be the wrong one. But the result is nevertheless of the sort that she intended. And that is a further reason for thinking that the belief he ended up with is the one that involves this identification of anchored representations.

So far these discussions have only been concerned with the semantic representations of (44.b) and (44.a). But the focus of Maier’s discussion was not so much on that as on the question how an adequate semantic representation (or an adequate semantic value, identified by means of a canonical description in the semantic metalanguage, typically some version of Higher Order Intensional Logic) can be compositionally derived from plausible, independently motivated syntactic structures for these sentences. This syntax-semantics interface question, as Von Stechow and Zimmermann, and Maier following them, pose it, is not easily rephrased in terms of the framework I am using. In the framework used by Von Stechow, Zimmermann and Maier the determination of the semantic value of a given expression E from a given syntactic structure $Syn(E)$ for E is compositional if it can be obtained by a recursive procedure that computes the semantic values (or their representations) of the leaves of the tree on the basis of the theory’s lexicon and the values or representations of intermediate nodes from those of their daughters according to a very limited repertoire of compositional operations. (In the most hard-headed versions of this approach these operations consist of just two: functional application and λ abstraction. But in any case (i) the set of admissible operations should be specified in advance and well-motivated, and (ii) it should be recognisable from the syntactic configuration consisting of the given mother node and its daughters which of the operations should be applied to obtain the semantic value or representation of the mother node from those of its daughters.)

DRT-based frameworks have remained aloof from a commitment to compositionality of such a specific and explicit sort. There is a commitment compositionality in this sense: (i) to a recursively specified syntax (for the natural language under study) which assigns to each expression of the fragment a syn-

tactic structure that takes the form of a tree (which may be decorated in various ways with ‘feature information’), and (ii) the requirement that there be a finite set of ‘composition rules’ which can be applied to local tree configurations and which determine how the semantic representations of the leaves of those local sub-trees can be combined into the semantic representation of its root.

There is a further difference between the strict compositionality concept that underlies the discussions in much of the literature (including those of Von Stechow and Zimmermann and Maier) and the more liberal compositionality commitment of DRT. This difference has to do with what counts as the ‘semantics’ that has to be compositionally computed from syntactic input and lexicon. One of the compositionality requirements of formal semantics in the spirit of Montague is that the items that are provided by lexical entries and then combined into larger items by the semantic composition rules can be defined (by the fundamental operations of the semantic meta-language, typically HOIL) from the constituents of the models that assign expressions their extensions (truth value to sentences individuals to names, and so on). DRT has replaced this commitment by the one that composition rules must operate on the semantic representations (DRSs and DRS conditions) that the theory admits; this requirement is weaker insofar as there may be more than one representation for the same model-theoretic object. Rules operating on semantic representations are thus allowed a greater flexibility than rules operating on model-theoretic objects directly.

This having been said, what remains of the compositionality concerns of Von Stechow and Zimmermann and Maier within the present framework? As noted at the outset of this section, the issue that is Maier’s concern (and before him that of Von Stechow and Zimmermann) is the apparent need their accounts have to assume the ‘*res*-movements’ that would have to precede the compositional computation of the semantics from the syntax. Such *res*-movements of the *res*-denoting DPs make it possible to interpret those DPs by the same ‘local’ interpretation rules that are also applicable to DPs occurring outside attitudinal contexts. We have seen that the ‘wide scope’ effects that result via *res*-movement are obtained in a quite different way within the framework we have been using here. The interpretation rules for occurrences of first and second person pronouns within attitudinal contexts that were given in Section 4.1 indicate how ‘wide scope’ interpretations of such noun phrases in the complements of attitudinal verbs can come about: When an embedded occurrence of a 1st or 2nd person pronoun agrees in person (and number, though this is a factor we have not considered here) with the agent argument of the nearest attitude predication in whose scope it stands, then the pronoun can be interpreted by means of the indexical discourse referent *i*. Such interpretations establish ‘local’ binding which in a framework like that of Maier et al wouldn’t require long distance *res*-binding. (What is needed to obtain such a pronoun interpretation is for the feature information carried by the pronoun to be brought together with the feature information of the agent phrase. This requires some kind of

feature percolation from one of the two DPs to the other; but the percolation is essentially local.)

The more challenging cases are those where there is no person (and number) agreement between pronoun and agent phrase. In such cases, it was argued, the pronoun must be interpreted as making a *de re* contribution, which takes the form of an anchored entity representation for the discourse referent introduced by the pronoun, together with a newly generated constraint on the external anchor of this representation. to the effect that it in its turn must be anchored in the right way: to the speaker if the pronoun is a 1st person pronoun and to the addressee if the pronoun is 2nd person. It is this constraint on the external anchor of the ('locally constructed') anchored representation for the pronoun which ensures that the pronoun will end up correctly bound eventually. As we have seen, this can take one of three forms: (i) there is no higher attitude predication and the external anchor is identified as *sp*; (ii) there is a higher attitude predication and the agent phrase of the attitude predication one level up agrees in person (and number) with the pronoun, in which case the external anchor is (typically) identified as *i*, or (iii) the agent phrase of the attitude predication one level up does not agree with the pronoun; in this case the external anchor becomes the discourse referent of an anchored representation that gets accommodated within the scope of the attitude predicate one level up, while the interpretation constraint gets passed on to the external anchor of this new anchored representation. It is the mechanism that permits the construction of the chains of anchored representations that can come about this way, and which can become as long as you like, that does the work of (cyclic) *res*-movement in a set-up like that of Von Stechow and Zimmermann and Maier.

It might be added that mechanisms of this sort are not limited to 1st and 2nd person pronouns. Mechanisms of the same kind are responsible for the interpretation for the interpretation of occurrences of other types of DPs in nested attitude attributions. In particular, these mechanisms can assign chains of anchored representations to 3rd person pronouns and also to directly referential noun phrases such as proper names, demonstratives and referential definite descriptions. To discuss the extension of the mechanism we have given to these third person noun phrases would exceed the scope of this note, but let us assume that such an extension is in place. A construction algorithm for semantic representation into which this extension has been incorporated will be able to assign to attitude-embedded occurrences of such third person DPs the same sorts of chains of anchored representations that the algorithm of Section 4 assigns to certain occurrences of 1st and 2nd person pronouns, and once again without an appeal to *res*-movement preparatory to the semantic representation construction.

For someone like me, who believes that anchored representations should be part of the analysis of any kind of *de re* interpretation of DPs the chaining mechanisms for anchored representations are a natural and attractive way of

dealing with 'long distance *de re*' interpretations. If they obviate the need for *res*-movement as a syntactic operation, then so much the better. But of course, the trade off between *res*-movement and the chaining mechanisms that can do its job involves a large scale change in syntax-semantics architecture over all.

6 Comparison with some other work.

[What follows are some very sketchy remarks that compare the present approach with two of the many others that can be found in the literature. I am adding these remarks only as an acknowledgement of the two papers on which I will comment. This should be neither understood as implying lack of interest or any other kind of negative attitude towards the rich and growing literature on attitude attribution and self-ascription. The choice of the two papers I will comment on briefly is motivated by my conviction that the issues they raise are most directly relevant to the issues raised in this note. But there are no doubt others, which I would have included here if I had read them more carefully, or that I would have included had I read them at all, but which, so far, I haven't. I hope to replace this section with a more responsible job shortly.

1. The nature of *de re* and *de se* interpretations and the difference between them is the central theme of some recent work of Ninan (Ninan (2011*b*), Ninan (2011*a*)). Ninan adopts (like Von Stechow and Zimmermann and Maier) a Lewisian framework in which all attitude attributions are analysed as properties that agents can self-attribute in different modes of self-attribution: If I say, in world *w* at time *t*, 'I am hungry', then that is an expression of my attributing to myself (in that world and at that time) the property of being hungry in the doxastic mode, and if I say that I want to see 'The King's Speech', then that is an expression of my attributing to myself the property of seeing 'The King's Speech' in a bouletic mode; and so on. Lewis adopted this way of analysing *de se* attitudes and *de se* attitude attributions, because it is unequivocally description-free, a desideratum that goes back to the work of Perry and others from the second half of the sixties (Perry (2000)). Lewis extends these cases of 'genuine' self-attribution to cases where the self-attribution is purely formal, i.e. where the self-attributed property does not distinguish the given self from any other individual at the given time in the given world. For instance my belief at *t* in *w* that Mary is in Paris is now analysed as my doxastically attributing to myself, at *t* in *w*, the property that anyone has at *t* in *w* iff Mary is in Paris at *t* in *w*. (i.e. if that proposition is true at *t* in *w*).

In the version of this theory that Ninan takes as his point of departure, self-attribution also involves a temporal dimension. My attributing to myself in *w* at *t* the property of being hungry is in this version really a matter of attributing to myself and the time at which I do the attributing the relation which holds between *x* and *t* if *x* is hungry at *t*. This temporal dependence, according to which having a propositional attitude is to attribute to oneself and the time at

which the attributing occurs a certain relation between individuals and times is now part of the general form of all attitude attributions.

Formally, this proposal leads to the following revision of the nature of attitudinal modalities and the semantics of attitudinal verbs. To concentrate on the case of belief: in pre-Lewis modal accounts of belief the belief state of an agent a (at a time t and in a world w) is characterised as a set $\text{dox}(w,t,a)$ of possible worlds - those worlds that are compatible with the totality of a 's beliefs at time t in world w ; and a proposition p (also characterised as a set of possible worlds) counts as believed by a at t in w iff $\text{dox}(w,t,a) \subseteq p$. In Lewis account, according to which all propositional attitudes are time-dependent self-ascriptions, a 's doxastic state at t in w is now to be identified as a set $\text{dox}'(a,t,w)$ of *centred worlds*, where a *centred world* is a triple $\langle w',x,t' \rangle$, where w' is a possible world, t' a time and x an agent in w' at t' . For a to have a certain belief p is now for p to be a relation between individuals and times such that $\text{dox}'(a,t,w) \subseteq p$.

A further aspect to Lewis' proposal was to reduce all (alleged) cases of *de re* attitudes and attitude attributions to cases of *de se*: for instance, John's '*de re*' belief about Mary that she is in Paris is reduced to John's doxastic self-attribution of the property ' x stands in acquaintance relation R to an individual y who is Mary and is in Paris'. Ninan argues – decisively, as far as I am concerned – that this reduction of the *de re* to the *de se* cannot work in general. In particular, it fails for what he terms 'counterfactual attitudes concerning a given 'res' '. (An example is that of an agent Ralph who has encountered a certain individual only once on an occasion where it has been impossible to identify this individual in terms other than those involving specifics of the encounter and who imagines what it would have been like to have encountered *this* individual under very different circumstances. On the Lewis account the only way of describing this mental act is as self-ascription-in-imagination of the property of standing in some acquaintance relation R' to an individual to whom one stands in the relation R , where the relations R' and R are mutually exclusive by necessity. In other words, on such a Lewisian account such an act of the imagination would be self-contradictory, which intuitively it is not.

To mend this defect within an essentially Lewisian framework Ninan proposes to generalize the notion of a centred world to that of a *sequenced world*. A sequenced world is a tuple $\langle w',t',x_0,\dots,x_N \rangle$, where x_1,\dots,x_N are, intuitively speaking, the entities about which the agent has *de re* attitudes. These entities are fixed by acquaintance relations in a *base world*, but there is no need for the individuals occupying the same slots in other sequenced worlds to satisfy those acquaintance relations. In this way Ninan's 'counterfactual imagination problem' for the original Lewis approach can be resolved. Finally the sequenced worlds approach is a genuine extension of the centred worlds approach in that the individuals playing the part of the 'self' also figure among those that make up sequenced worlds, viz. the individuals occurring as the elements x_0 . In the base world the filler of this position is identified as standing in the relation of identity to the agent; but again in the sequenced worlds that now make up an

agent a 's doxastic state in w at t the element x_0 need not be identical with a .

Ninan's use of slots of sequenced worlds to represent the entities about which the agent is in a position to entertain *de re* thoughts are reminiscent of DRT – most particularly, perhaps, in reformulations like that of Muskens (1989), where discourse referents are treated as 'registers'. But the approach we have taken in this note suggests another analogy: between the positions in the sequenced worlds that make up the doxastic state of an agent a in a world w at a time t and the entity representations that a has in w at t . The difference, however, is that in the approach adopted here the referents (= external anchors) of the anchored representations which are the harbingers of *de re* thought in our approach are fixed, whereas the occupants of a given slot in the different sequenced worlds that make up an attitudinal state may vary. (I doubt that this variability is ultimately what we want, but the matter is delicate and would require more careful discussion than I am able to give it here.)

On the face of it the difference between *de se* and *de re* attitudes looks less radical on Ninan's account than it does in the account used here. But whether this is a difference of substance or cosmetics is something that can be decided only when the two approaches are applied to more complicated cases. (Relevant in this connection would be how Ninan's theory would deal with examples like the Maier cases discussed in the last section, but I have had no opportunity yet to look into this closely.)

2. In several recent papers Higginbotham discusses a number of aspects of the *de se* as distinct from the *de re* (Higginbotham (2008), Higginbotham (2010)). One of Higginbotham's concerns in these papers is the relation between *de se* and immunity to error. There has been considerable confusion on this issue in the philosophical literature. Higginbotham points out – and there is no question that he is right – that while some self-attributions are not immune to doubt, in the sense that for all the self-attributor knows he could be wrong in attributing a certain property to himself, there is never any room for doubt in such cases that it is *to himself* that the attributor makes the attribution; the one to whom he attributes the property in an act of self-attribution cannot be anyone other than himself. This point is of particular importance in connection with memory. To remember being in a certain situation – and thus to remember having the property that consists in being in such a situation – involves an *ipso facto* identity of the one who does the remembering with the one to whom the property is being attributed *in* the given act of remembering; and here too there is no room for mistake as to who it is that the agent of the act of remembrance is attributing to – that can only be he himself.

Another question that Higginbotham discusses are the linguistic forms that different natural languages make available for describing or reporting self-attributions. In English, combinations of an attitude verb (or verb of saying) and an infinitival complement can only be understood as describing self-attributions by the verb's

subject, whereas finite complements are neutral between such an interpretation and one according to which the sentence attributes to the subject an attitude that is either *de re* or *de dicto*, but not *de se*. These observations are clearly important and they will become essential to the project of which the present note has sketched the first steps, when that project will be developed further and the rules for the interpretation of 1st and 2nd person pronouns in attitudinal complements are generalised to sentences with infinite complements. On the other hand it is not clear to me from what I have seen from Higginbotham's work so far how it might handle the interpretation of 1st and 2nd pronouns (or, for that matter, 3rd person pronouns) in nested attributions.

7 Conclusions

- The first part of this presentation outlines the syntax and use of a formalism for the representation of simple and complex attitudinal states. (The model-theoretic semantics for the formalism was not presented.) The emphasis is on the representation of complex states, consisting of several attitudes, possibly of distinct attitudinal modes.
- The formalism is capable not only of representing attitudes with propositional content but also has the means for representing entities.
- Entity representations play an essential part in thought *de re*. An entity representation of the kind discussed here makes any thought that contains it as a constituent *de re* with respect to the entity it represents. Thus entity representations are purveyors of singular content.
- In addition, a thought can have singular content in virtue of being *de se*. The formalism offers a distinct mode for representing thought *de se* (and '*de nunc*') in the form of the 'internally indexical' discourse referent *i* (and *n* in the case of *de nunc*).
- Both *de re* representations and *de se* representations are relevant to the interpretation of first and second person pronouns when they occur in the complements of attitudinal predicates. When such a pronoun occurrence is to be interpreted as making a *de re* or a *de se* contribution to the reported attitude or attitudes is not nearly as straightforward a matter as might have been thought.
- The second half of the paper deals with this question in some detail, by:
 - (i) showing how attitude reports containing *I* and *you* can be represented within the system;
 - (ii) showing how such representations can be constructed from the natural language attitude reports as inputs, and

(iii) stating principles for the interpretation of first and second person pronouns in the different positions they can occupy in attitude reporting sentences.

- Section 5 of the paper considers a case discussed by Maier, in which an agent misinterprets an utterance by another agent and thereby forms a mistaken belief about the attitudes of this second agent. Cases of this sort are a good test case for the present approach, since they allow entity representations to play a part in accounting for our intuitions that goes well beyond what can be illustrated at the hand of the examples discussed in the preceding sections.

References

- Beaver, D. & Geurts, B. (2005), Discourse representation theory, *in* E. Zalta, ed., ‘Stanford Encyclopedia of Philosophy’, Stanford University, Palo Alto.
- Chastain, C. (1975), Reference and context, *in* K. Gunderson, ed., ‘Language, Mind and Knowledge (Minnesota Studies in the Philosophy of Science, VII)’, University of Minnesota Press.
- Genabith, J. v., Kamp, H. & Reyle, U. (2011), Discourse representation theory, *in* D. Gabbay, ed., ‘Handbook of Philosophical Logic’, Kluwer.
- Geurts, B. (1997), ‘Good news about the description theory of names’, *Journal of Semantics* **14**, 319–348.
- Higginbotham, J. (2008), Expression, truth, predication, and context., *in* R. Schantz, ed., ‘Prospects for Meaning’, de Gruyter, pp. 473–494.
- Higginbotham, J. (2010), On words and thoughts about oneself, *in* I. Stojanovich, F. Recanati & N. Villanueva, eds, ‘Context-Dependence, Perspective and Relativity’, de Gruyter, pp. 253–282.
- Kamp, H. & Bende-Farkas, A. (2006), Epsitemic specificity from a communication-theoretic perspective. submitted to the *Journal of Semantics*.
- Kamp, H. & Reyle, U. (1993), *From Discourse to Logic*, Kluwer.
- Kaplan, D. (1989), Demonstratives, *in* H. W. J. Almog, J Perry, ed., ‘Themes from Kaplan’, Oxford University Press.
- Kripke, S. (1972), Naming and necessity, *in* D. Davidson & G. Harman, eds, ‘Semantics of Natural Language’, Reidel, Dordrecht, pp. 253–355.
- Lewis, D. (1979), ‘Attitudes de dicto and de se’, *The Philosophical Review* **88**(4), 513–543.

- Maier, E. (2009), Iterated *de re*: A new puzzle for the relational report semantics, in A. Rieger & T. Solstad, eds, 'Proceedings of Sinn und Bedeutung 13', Vol. 2, pp. 347–355.
- Muskens, R. (1989), Meaning and Partiality, PhD thesis, University of Amsterdam.
- Ninan, D. (2011*a*), Counterfactual attitudes and assignment-sensitivity. Unpublished Manuscript.
- Ninan, D. (2011*b*), Self-location and other-location. Unpublished Manuscript.
- Perry, J. (2000), *The Problem of the Essential Indexical and Other Essays (Enlarged Edition)*, CSLI, Stanford.
- Stalnaker, R. (1988), Belief attribution and context, in R. Grimm & D. Merrill, eds, 'Contents of Thought', University of Arizona Press.
- Stechow, A. v. & Zimmermann, T. (2004), A problem for a compositional account of *de re* attitudes, in G. Carlson & F. Pelletier, eds, 'Reference and Quantification: the Partee Effect', CSLI Press, pp. 207–228.