

Constructing Compact Representations of Ambiguous Language

Kurt Eberle and Carl Vogel

Institute for Computational Linguistics

University of Stuttgart

Azenbergstr 12

D-70174 Stuttgart

Germany

{kurt,vogel}@ims.uni-stuttgart.de

Abstract

An application of DRT to the development of a computational semantics for a broad coverage fragment of German is described. The architecture of the system is influenced by human use of natural language in that initial processing is a fast but shallow process. For any sentence in the fragment, a single representation is constructed from which exactly the appropriate disambiguations can be enumerated or reasoned with directly if full analysis isn't essential. The representational assumptions are outlined, as is the scope of the current coverage in the semantic fragment and its implementation. We detail the semantic analysis assumed for questions.

Keywords: DRT, underspecified semantics

1 Introduction

This paper gives an overview of semantic analysis within a broad-coverage fragment of German.¹ We adopt an extension of underspecified discourse representation theory (UDRT) (Reyle 93; Reyle 95) in order to provide succinct representations of ambiguous elements. These extensions are motivated at length elsewhere (Eberle 96b). Here we give an overview of the formalism in which representations are constructed and illustrate how this is put to use in composing semantic representations within the fragment. A corresponding implementation of the fragment builds compositional representations of input sentences analyzed

¹The project is called "An Underspecified Discourse Representation Theory Semantics for a Fragment of German" in the DFG funded, "Special Research Project" on "Theoretical Foundations for Computational Linguistics". This subproject, situated in Stuttgart, has a sister in Tübingen whose aim is to develop a corresponding syntax in Head-driven Phrase Structure Grammar (HPSG)(Pollard & Sag 94). On UDRT integration with HPSG, refer also to (Frank & Reyle 95). Another sister project in Stuttgart is developing a reasoner which offers discourse disambiguations on the basis of the semantic representations produced by our project. The projects are finishing up the third of an initial three years of funding. This would have been impossible without Hans Kamp, Julia Hockenmaier, Frank Keller, Peter Krause, Antje Rossdeutscher, Michael Schiehlen and in fact the entire Semantics Group at the Institute for Computational Linguistics in Stuttgart. The work reported here was done while both authors were at IMS in Stuttgart; now Eberle is at IBM, Heidelberg, and Vogel is at Trinity College, Dublin.

in an HPSG syntax, to be disambiguated and to some extent interpreted by a theorem prover developed in a sister project.

Consider an example discourse from (Kamp *et al.* 95). The sentence in (1) is ambiguous; it doesn't specify whether there were two separate traveling events or one joint traveling. Nonetheless, we want a single representation that captures both readings.

(1) Fritz and Sabine travelled to Paris.

- (2) a. There they met.
b. There they separated.

The sentences (2.a) and (2.b) constitute alternative subsequent discourse, each supplying information via the lexical semantics of its verb that disambiguates the first sentence. The semantics of the individual sentences is articulated within an extension of UDRT, and representations of these are constructed in the feature logic assumed by the Comprehensive Unification Formalism (CUF) (Dörre & Dorna 93). An accompanying reasoner based on a tableau theorem prover, and also implemented in CUF (and Prolog), provides disambiguations for such underspecified representations. This paper focuses on the extension of UDRT and implementation of that extension in the AVM language supplied by CUF. We explain how the framework captures quantifier scope and collective/distributive ambiguities. We also show our implementation of an analysis of embedded and unembedded questions which takes advantage of underspecification in the case of argument queries to allow us one representation even in the case of a semantically complex phenomenon.

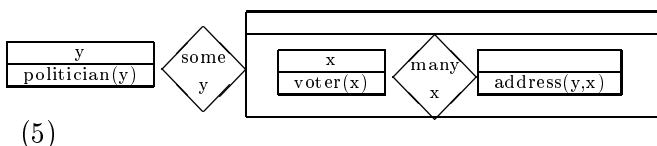
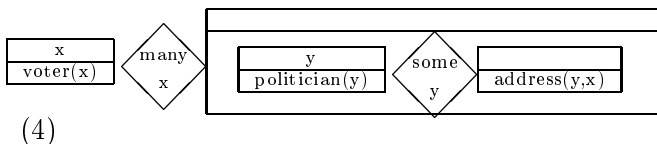
2 Underspecifying DRT

As presented in the literature (Pollard & Sag 87; Pollard & Sag 94), HPSG adopts Cooper storage as a mechanism for accommodating scopal ambiguities, however this mechanism isn't expressive enough to articulate partial knowledge of scope.

Thus, for instance, with respect to a sentence with three quantifiers Q_1 , Q_2 and Q_3 , the storage mechanism allows for obtaining all of the possible permutations, or to leave all or some quantifiers on the stack, but there is no means to express a constraint like Q_3 *has scope over* Q_1 (which might develop from a specific syntactic configuration) that restricts the possible outcomes. Cooper storage makes it possible to obtain different readings, but it cannot really express constraints about the acceptable readings and pack these readings into one underspecified representation. What is required is a general mechanism for stipulating partial orders among scopal elements, and this exactly the move made within UDRT.

This theory offers a mode of representation that provides for representing scopal ambiguities as well as constraints on ambiguities, and it gives a logic that allows for directly reasoning with underspecified structures (König & Reyle 96). Example (3) illustrates the advantage of moving to UDRT representations. Possible readings depending on the order of scoping of the relevant determiner phrases (DPs) in (3) are depicted in the DRSs in (4) and (5). Here we use the duplex conditions familiar for generalized quantifiers. For simplicity, we exclude here tense information predicated of the implicit event argument of the verb, but it should be pointed out that ambiguities arising from focus and temporal adverbials are one of the important phenomena covered within the fragment (Eberle 96a).

- (3) Many voters were addressed by some politicians.



Following the suggestion of (Reyle 93; Reyle 95), we add labels to the salient structures in the representation, and are able to specify a partial ordering over these labels which compactly represents both readings in a single structure, leaving enumeration of the disambiguated readings to a separate module (König & Reyle 96). The UDRS in Figure 1 illustrates how a representation language that takes advantage of partially ordered

substructures can yield a single compact representation of ambiguity. A line connected to an empty box in the ambiguous representation indicates its potential contents in a disambiguated DRS. That is, if l_1 fills the empty box in l_T then we have the reading corresponding to (5), because l_B is subordinated to l_2 and there is no other place for l_2 to then go except l_{12} ; similarly, if l_2 fills the empty box in l_T , then l_1 must fill in the empty box of l_2 with l_B giving the content of l_{12} , yielding the reading in (4). In a linear notation relative to the same labeling, we have the following constraints on scope: $l_1 \leq l_T, l_2 \leq l_T, l_B \leq l_{12}, l_B \leq l_{22}$. Because it is a partial ordering, intermediate stages of knowledge about relative scope of elements (such as that available from syntax) can be expressed by adding *immediate subordination* ($<$) constraints between labels. Notice that no label is given in this example for the quantifier within the duplex condition. This is not essentially so, and in fact, solutions to the parts of the problem of antecedent contained deletion in ellipsis resolution hinge precisely on flexible scoping of the quantifier itself (Beil 96).

Extensions to UDRT include enriching the labels with features (Eberle 96b). Information is associated with labels as appropriate to a *distinguished discourse referent* within the structure labeled at that point. For example, the event variable is the distinguished referent of a VP, and an individual or sum in the case of DPs. Thus, associated with a label will be syntactic information like case or person, as well as semantic information like whether the sum alone is available for subsequent discourse modification or summed individuals as well. Enriching the labeling in this way makes it possible to use the same approach used in relative scope ambiguity also in the case of distributivity/collectivity, as well as to use functional application in a compositional construction process. Exactly those DRFs that decorate a label are available for linking with anaphors in subsequent discourse.

As an example, consider the semantic contribution of a quantifier focusing on potential readings of the sentence in (6). Here it will be essential to make explicit the event argument of the verb.

- (6) Several scientists had an idea.

This sentence has three different readings which follow from particular orderings of scope relations among the referring elements. In one case, there is one idea and one event (perhaps com-

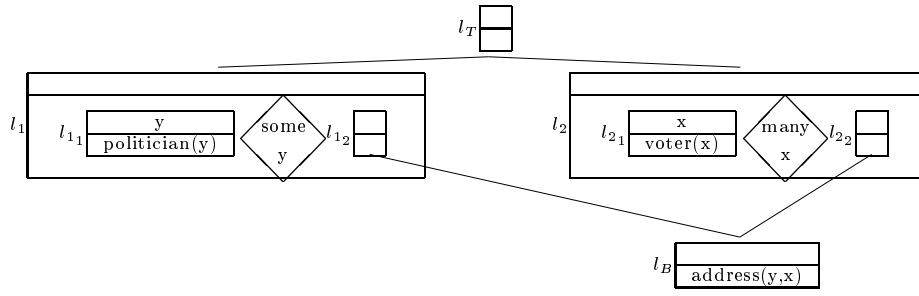


Figure 1: A UDRS for Scopal Ambiguity

plex) in which several scientists collectively had the idea; in another, several scientists each individually had an arbitrary idea (an idea potentially different from the ideas had by the others) in separate thinking events; and in the last, several scientists individually had the same idea in different thinking events.² Formalizing this, as mentioned, requires making explicit the event argument of the verb and additionally formulating summations over the events and individuals involved. The summations will be denoted with capital letters (X , for sums of individuals; E , for event sums), using notation defined as follows:³

$$(7) \quad X, E :: \begin{array}{|c|} \hline x \\ \hline P(x) \\ \hline \end{array} \begin{array}{c} \text{Quant} \\ x \end{array} \begin{array}{|c|} \hline e \\ \hline V(e,x) \\ \hline \end{array} \iff \alpha \wedge \beta \wedge \gamma$$

where the conditions α , β , and γ are defined:

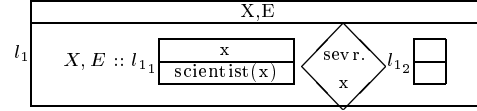
$$\alpha : \begin{array}{|c|} \hline x \\ \hline P(x) \\ \hline \end{array} \begin{array}{c} \text{Quant} \\ x \end{array} \begin{array}{|c|} \hline e \\ \hline V(e,x) \\ \hline \end{array} \quad \beta : X = \left\{ x \mid \begin{array}{|c|} \hline x, e \\ \hline P(x), V(e,x) \\ \hline \end{array} \right\} \quad \gamma : E = \left\{ e \mid \begin{array}{|c|} \hline x, e \\ \hline P(x), V(e,x) \\ \hline \end{array} \right\}$$

Now, a compact representation of the quantified NP (6) can be given as in (8). Assume that l is the label for the verb semantics and that l' is the label of the DRS containing a condition **idea**(y). Also let $<$ express an immediate containment relation between the structure labeled by the element on the left in the structure pointed to by the label on the right, and let $\leq$ be the reflexive transitive closure of that relation. Then the following equations, in conjunction with the structure in (8) enumerate the three readings of (6) mentioned above.

$$(8) \quad \llbracket \text{several scientists} \rrbracket =$$

²The possibility of narrow scope for the idea, and wide scope for the event (several scientists collectively having an idea, each potentially different from the idea the others have, all at the same time and as a single event) is not a reading of this sentence.

³ P and V are metavariables for predicate names.



1. $l < l_1 \wedge l' < l_1$ = collective having of the idea in one event
2. $l \leq l_{12} \wedge l' \leq l_{12}$ = distributive having of ideas in separate events
3. $l \leq l_{12} \wedge l' < l_1$ = distributive having of one idea in separate events.

It was stated above that the labels are decorated with distinguished discourse referents, but these are not indicated in association with l_1 in (8). The real structure of l_1 looks something more like this: $l_1^1 \langle l \leq l_{12} \supset E, l < l_1 \supset e \rangle$. This conveys information about the upper referential index of the DP, that which is available for anaphora in subsequent sentences, namely that it is the sum X (that is, the referent of a DP is the set it constructs). The result index, corresponding to the event variable of the verb is also available as a sum (E) just in case the verb is in the scope of the quantifier; if the verb semantics is merged with the top DRS of the DP (l_1) then there is only one event, and the set formed by abstraction over that event is not available for subsequent adverbial modification nor for anaphora. Finally, the lower referential index is that which is available from the DP as an argument to the verb — this is the sum (X) just the verb is contained directly in l_1 , and the individuals when the verb is in the scope of the quantifier $l \leq l_{12}$.

In addition to decorated labels for structures, we have different sorts of structures that can have labels. That is, we distinguish disambiguated DRSs (which have a universe and conditions), partial DRSs (PDRSs) in which there is a separation of a ‘bottom’ element from a complex condition, and further underspecified DRSs which have

not just one complex condition, but which has a list of functors which apply in turn to the bottom element. The bottom element corresponds, essentially, to the relation carried by the matrix verb of a sentence. Ordering constraints (objects of type *ord* on the SUBS list) on labels associated with a DRS will allow a PDRS to express all of the readings from something like collective/distributive ambiguity all in one representation. Partial DRSs are used in the fragment to represent the semantics of adverbs and adjectives. A more complex structure is similar to the PDRS, but has a functor list (FSET): the set of functors (PDRSs for example) obtained in the composition of constituents are taken as applying to the bottom element (again, the verb semantics). These applications do not take place during the semantic construction, but instead are left for a post-construction disambiguation process to enumerate. The bottom element indicates the case, sortal restrictions, and theta-role restrictions of the arguments, stipulating conformity with the lower referential index of the corresponding DP. The possible identifications between the argument and an instantiation (either the individual discourse referent (DRF) or the summed DRF) are dictated by the alternation conditions in the lower referential index.

An example of the feature-structure encoding of partial and functor DRSs is demonstrated in (9) which shows part of the representation for (10) (*peter loves maria*).⁴

(10) Peter liebt Maria.

The feature structure omits much of the information that is not relevant to discussion at the moment. The empty λ -list indicates that the structure is saturated — it yields the object indicated as the value of the feature RESULT. The RIND feature gives the result index of the structure — it is that corresponding to the referential argument of the verb and has tense features as appropriate. The FSET feature indicates the set of functors (PDRSs in this case) which are to be applied to the value of the BOT feature during disambiguation. It was mentioned earlier that the referential index (always a sum, if it exists at all) and lower referential index (either an individual or a sum) can be distinct, and indeed, this fact is used to capture collective/distributive ambiguities as it is the lower referential index which is passed in as

⁴Unfortunately, space does not permit full explication of the feature-value ontology here.

the argument of the verb. Here, in the case of proper names, the values of these indices are identical and bear information associated with the restrictions on the corresponding role. With proper names, those structures are properly determiner phrase semantics: each is a PDRS rather than a functor set DRS. The value of the PDRS feature on each is not shown but it is essentially just a DRS which introduces α -conditions as special DRS conditions for modeling the presuppositions accompanying the use of proper names (van der Sandt 92). The SUBS feature details the subordination constraints appropriate to each PDRS; in this case the constraints are not interesting — it is just that the conditions introduced by each PDRS are subordinated to the root level of the PDRS (rather than to some structure embedded within it). Within the value of the BOT feature on the functor DRS as a whole we have another PDRS whose conditions stipulate the relations between discourse referents introduced elsewhere in the structure: the index [19] is the discourse referent for the event argument of the verb, and it stands in the designated relations to each of the discourse referents ([4] and [12]) that come from the referential indices of the arguments to the verb. The task of semantic composition is to derive such a structure from the lexical entries associated with words used in the sentence using syntax to guide the composition process.

Admittedly, the meanings corresponding to these underspecified representations are not as available to inspection as the original DRS boxes, however the constraints on labels of semantic substructures, including the logic permitting inferences about scopal relations, does indeed offer a very compact representation for ambiguity. The price paid for efficient semantic construction is deferment of ambiguity enumeration until after the recognition process. However, without underspecified representations, the recognition process cannot terminate until the enumeration of disambiguations does.

2.1 Flat Semantics

The approach to lexical ambiguity taken within this project is again to devolve responsibility to modules outside the basic semantic construction routine accompanying the basic parse. Disambiguations are deferred until as late as possible, unless sortal restrictions can be brought to bear early. In order to avoid intractable multiplications of readings in the wake of lexical ambiguity,

(9)

$$\begin{array}{l}
\text{lambda: } \langle \rangle \\
\text{rind: } \left[\begin{array}{l} \text{cat: } [1] \text{ verb } [] \\ \text{mtv: } [2] \left[\begin{array}{l} \text{diathesis: active} \\ \text{mood: indi} \\ \text{tense: present } [] \end{array} \right] \end{array} \right] \\
\text{key: lieben} \\
\text{fset: } \left\langle \begin{array}{l} \text{rind: } [3] \\ \text{refind: } [7] \left[\begin{array}{l} \text{cat: noun } [\text{case: nom}] \\ \text{ref: } [4] [\text{sort: human}] \\ \text{rolespec: } \end{array} \right] \\ \text{lrefind: } [7] \\ \text{key: peter} \\ \text{num: } [8] \\ \text{pdrs: } \left\langle \begin{array}{l} \text{subs: } \left\langle \begin{array}{l} \text{ord } \left[\begin{array}{l} \text{o_arg1: } [9] \\ \text{o_arg2: } [8] \end{array} \right] \\ \text{o_rel: } \leq \end{array} \right\rangle \end{array} \right\rangle \\ \text{bot: } [10] \left[\begin{array}{l} \text{rind: } [3] \\ \text{num: } [9] \end{array} \right] \end{array} \right\rangle \\
\text{result: } \left[\begin{array}{l} \text{basic_akt: stative} \\ \text{cat: } [1] \\ \text{mtv: } [2] \\ \text{ref: } [19] [\text{sort: loves}] \end{array} \right] \\
\text{key: lieben} \\
\text{bot: } \left[\begin{array}{l} \text{univ: } \langle [19] \rangle \\ \text{rels: } \left\langle \begin{array}{l} \text{relcond } \left[\begin{array}{l} \text{r_rel: } [\text{r_name: lieben}] \\ \text{r_args: } \langle [19] \rangle \end{array} \right] \\ \text{relcond } \left[\begin{array}{l} \text{r_rel: } [\text{r_name: arg1}] \\ \text{r_args: } \langle [19] [4] \rangle \end{array} \right] \\ \text{relcond } \left[\begin{array}{l} \text{r_rel: } [\text{r_name: patient}] \\ \text{r_args: } \langle [19] [12] \rangle \end{array} \right] \end{array} \right\rangle \end{array} \right] \\
\text{rind: } [11] \\
\text{refind: } [15] \left[\begin{array}{l} \text{cat: noun } [\text{case: akk}] \\ \text{ref: } [12] [\text{sort: human}] \\ \text{rolespec: } \end{array} \right] \\
\text{lrefind: } [15] \\
\text{key: maria} \\
\text{num: } [16] \\
\text{pdrs: } \left\langle \begin{array}{l} \text{subs: } \left\langle \begin{array}{l} \text{ord } \left[\begin{array}{l} \text{o_arg1: } [17] \\ \text{o_arg2: } [16] \end{array} \right] \\ \text{o_rel: } \leq \end{array} \right\rangle \end{array} \right\rangle \\
\text{bot: } [18] \left[\begin{array}{l} \text{rind: } [11] \\ \text{num: } [17] \end{array} \right] \end{array} \right\rangle
\end{array}$$

we provide a *flat* semantics for those ambiguous items, and this is elaborated out into an appropriate deep semantics only if there is the necessary disambiguating information available. This involves specifying the semantic lexicon for those entries with a more generic structure, essentially giving unevaluated functional terms that subsume each of the possibilities, and which (later) can be evaluated following the rules of the corresponding functions when sufficient disambiguating information is available or if the task of enumerating disambiguated readings has arrived in the process which follows basic semantic construction.

2.2 Coverage

There is not sufficient room here to give detailed discussion of semantic analyses adopted within the fragment. Let it suffice to say that the intention is to construct semantic representations for the included phenomena which to the greatest extent possible reflect the state of current wisdom on a range of phenomena. The fragment itself includes problems in quantification, modification (that arising from relative clauses, adjectives, prepositional phrases and adverbials), coordination, embedding verbs, questions and presupposi-

tion. We have implemented flat and deep semantic representations for adjectives, some adverbials, argument and adjunct prepositional phrases, quantification, as well as sentence and question embedding verbs. The fragment does not at present encompass auxiliaries. The current fragment assumes a basic quantificational approach to the semantics of questions. The implementation of the fragment, the semantic lexicon and construction rules, is in CUF (Dörre & Dorna 93), a constraint solving system built on top of Prolog and designed precisely for efficient processing of typed feature structures like those assumed by HPSG. We implement the UDRS language in feature structures using CUF to facilitate the integration of syntax and semantics. Additionally, the theorem prover for UDRS representations built with a sister project is also constructed in the same language, although it remains to integrate these two modules. The reasoner, designed by Esther König-Baumer, in its current state can from alternative representations of the discourse produce the salient disambiguations; however, the two representation languages have not yet been integrated. Therefore, there is not yet throughput from text, through HPSG syntactic analysis, and

the extended UDRT semantics, into disambiguations and other reasoning. The current implementation does interface with syntax, obtaining the functor argument structure from the HPSG feature structures. We emphasize (in part because the extent of the coverage might otherwise be difficult to believe) that the task undertaken by this project has been to specify the composition of semantic representations. Underspecified representations are used in order to speed the initial recognition process. The task of enumerating and adjudicating among disambiguations, along with nontrivial interpretation in general, is left to subsequent modules. Nonetheless, the construction of semantic representations is an important task on its own.

3 Some Particular Examples

3.1 Embedded declarative sentences

The semantics of an embedded clause is mediated by the complementizer.

Often, complementizers are assigned the empty semantics in the sense that their semantic contribution only consists of percolating the argument sentence semantics (into a subcategorized argument position of the verb meaning for instance). We deviate from this, because we think that the introduction of descriptors for DRSs simplifies matters in many respects and because we think that in the presence of such descriptors the meaning of a complementizer is to introduce such a descriptor or to choose a descriptor or referent of the argument sentence. The introduction of DRS descriptors, p , with $p : K$, that are interpreted as first order representatives of what they describe allow for simple and uniform modelings of parts of speech that accept classical first order objects as well as propositions as bearer of some thematic role (Eberle 96b, p. 78)

Here we assume that (11) is the feature structure representation of the sort of condition appropriate to ‘reifications’ of propositions.

$$(11) \quad p_descr_cond \left[\begin{array}{l} pd_prop: \boxed{1} \\ pd_descr: \boxed{2} \end{array} \right]_{pdrs} [\quad]$$

For the case of embedded declaratives, we have the implicit assumption that the DRS described by the labeled structure corresponding to $\boxed{2}$ is a true description of the proposition $\boxed{1}$. This means that the representation of the complementizer is such that it introduces a DRS condition like (11); part of the representation appropriate to the complementizer is given in (12).

Here we have a λ -abstraction over the label for a verb semantics. This is structure-shared with the value of BOT in the result, embedded inside the PDRS condition describing the embedded

proposition. The structure of $\boxed{10}$ isn’t shown; it just introduces α -conditions for temporal reference appropriate to the tense of the embedded sentence, as well as the descriptions that conjunctively describe the proposition reified as the DRF $\boxed{8}$. The result of applying the semantics for the complementizer to the semantics of a tensed declarative sentence is a further instantiated feature structure taken from the value of RESULT.

Some further considerations clarify the significance of this representation. First, we do not in general use DRFs for propositions in the semantics of sentences. In particular, this is the representation only of declaratives embedded under an attitude verb, or discourses that require propositional anaphora. However, outside a belief context, other constituents are available whose semantic import is to establish such a referent.

(13) Theo believed that the paper had 100 pages and it frustrated him.

(14) The paper had 100 pages. It frustrated Theo.

(15) The paper had 100 pages. That fact frustrated Theo.

Consider (13–14); in (13) it is quite easy for the pronoun **it** to refer to the proposition that the paper has 100 pages, but not in the case of (14) in which it is the paper that frustrated Theo rather than the proposition. On the other hand, if there is a specific constituent available whose semantics involves the forming of a proposition, like **that fact** in (15), then indeed that proposition can be the argument of **frustrated**. At the outset of this section we pointed out that it is at the point of embedding (under a complementizer, for example) that a propositional DRF is introduced. The second point to make in this connection is that the semantics of questions is slightly different in that for both embedded and unembedded questions we form a corresponding DRF appropriate to that semantic object — more detail about this is given in the following section. Finally, we note that admitting propositions as DRFs does ultimately entail incompleteness in a reasoner, but the paradoxes of self-reference are not part of the fragment of German that we are interested in here.

3.2 Questions

The representation of questions is much like the representation of embedded clauses: a referent is introduced which can be construed as a reification of the proposition expressed by the clause.

$$(12) \left[\begin{array}{l} \text{lambda:} \left\langle \begin{array}{l} \boxed{7} \\ \text{rind:} \boxed{4} \text{ vind} \left[\begin{array}{l} \text{akt:} \boxed{1} \text{ het} \\ \text{basic_akt:} \boxed{1} \\ \text{mtv:} \left[\begin{array}{l} \text{diathesis:} \\ \text{mood: indi} \\ \text{tense:} \end{array} \right] \\ \text{ref:} \boxed{2} \\ \text{rt:} \boxed{3} \end{array} \right] \end{array} \right\rangle \\ \text{result:} \left[\begin{array}{l} \text{rind:} \boxed{4} \text{ eind} \left[\begin{array}{l} \text{ref:} \boxed{8} \end{array} \right] \\ \text{univ:} \left\langle \boxed{8} \right\rangle \\ \text{drs:} \left\langle \begin{array}{l} \text{conds:} \left\langle \text{p_descr_cond} \right\rangle \\ \text{pd_descr:} \text{ pdrs} \left[\begin{array}{l} \text{pd_prop:} \boxed{8} \\ \text{rind:} \boxed{4} \\ \text{key:} \boxed{5} \\ \text{num:} \boxed{9} \\ \text{pdrs:} \left[\begin{array}{l} \boxed{10} \\ \text{subs:} \left\langle \begin{array}{l} \text{ord} \left[\begin{array}{l} \text{o_arg1:} \boxed{6} \\ \text{o_arg2:} \boxed{9} \\ \text{o_rel:} \leq \end{array} \right] \end{array} \right\rangle \end{array} \right] \end{array} \right\rangle \\ \text{bot:} \boxed{7} \end{array} \right] \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

In the case of propositions, there is an implicit predication of truth of the description with respect to the proposition-DRF. The same is also roughly the case with questions, except that questions make explicit that the choice between predication of truth and of falsity is not certain. Additionally, argument roles can be parameterized in questions but not in propositions.

$$(16) \text{ q_descr_cond} \left[\begin{array}{l} \text{qd_prop:} \boxed{1} \\ \text{qd_descr:} \boxed{2} \text{ pdrs} [] \\ \text{qd_roles:} \boxed{3} \langle \rangle \end{array} \right]$$

Thus, (16) indicates the structure of conditions introduced by questions. When the list of queried roles ($\boxed{3}$) is empty then we have a yes-no question — this is just a test of the truth of the embedded sentence. If the list of queried roles is not empty, then we have either an adverbial question or one about fillers of argument roles in the embedded sentence. With adverbial questions, the argument of the query role is the value of QD_PROP, itself, and otherwise the values are determined by disambiguations of the alternations given in the lower referential index of each queried argument.

$$(17) \left[\begin{array}{l} \text{qd_prop:} \boxed{8} \\ \text{qd_descr:} \text{ pdrs} \left[\begin{array}{l} \text{rind:} \boxed{4} \\ \text{key:} \boxed{5} \\ \text{num:} \boxed{9} \\ \text{pdrs:} \boxed{10} \\ \text{subs:} \left\langle \begin{array}{l} \text{ord} \left[\begin{array}{l} \text{o_arg1:} \boxed{6} \\ \text{o_arg2:} \boxed{9} \\ \text{o_rel:} \leq \end{array} \right] \end{array} \right\rangle \\ \text{bot:} \boxed{7} \end{array} \right] \\ \text{qd_roles:} \left\langle \left[\begin{array}{l} \text{r_rel:} \left[\begin{array}{l} \text{r_name: wie} \\ \text{r_type: whq} \end{array} \right] \\ \text{r_args:} \langle \boxed{8} \rangle \end{array} \right] \right\rangle \end{array} \right]$$

A condition that would be introduced into the partial DRS for an adverbial q-word, *wie* (how),

is given in (17). This object of type *q_descr_cond* contrasts with the complementizer condition in (12). It has an additional feature for queried roles, and in the case of an adverbial the argument of the query is the DRF for the state of affairs described by the content of the interrogative.

A query about an argument role will involve quantification; hence, the argument of the query relation on the QD_ROLES list will be the lower referential index of the quantifier for each queried role. We assume that a wh-word like *welcher* (which) is a kind of basic quantifier and needs to combine with an NP to form a DP, and that a wh-word like *wer* (who) is a sort of DP (in both cases, the kind is a specialization for questions — the motivation for this will become clearer in a moment when we discuss composition for embedded interrogatives involving argument queries) with a relatively empty restrictor. In terms of the PDRS notation used earlier, we assume the structure (18) for *welcher mann*. As before, it is open whether the semantics of the verb is merged directly into l_1 or subordinated by l_{12} .

$$(18) \begin{array}{c} \boxed{\begin{array}{c} X, E \\ X, E :: l_{11} \left[\begin{array}{c} x \\ \text{man}(x) \end{array} \right] \text{ welcher } l_{12} \left[\begin{array}{c} x \end{array} \right] \end{array}} \end{array}$$

Composition of these DPs with the embedded verb semantics (or matrix verb in the case of non-constituent questions), because they are a specific sort of DP, percolates the information about queried roles. The point at which composition saturates a verb with queried arguments is the point at which a condition of the form described in (16) is added, and the value of QD_ROLES for that

interrogative meaning will consist of a list of each of the queried roles. This means the final composition is as if with a complementizer, as it is with complementizer-like words that we have associated the introduction of a QD_CONDITION which ‘declares’ the presence of a question. In the case of adverbial or y/n queries, it is easy to see the q-word (**how**, **whether**) as a complementizer. The HPSG syntactic analysis from our sister project (Richter 96) in fact assumes that the initial wh-word (actually, they assume this for phrases like *welcher mann* as well) in an embedded interrogative functions exactly like *dass* (that). Of course, we do not require from the semantics that the leftmost wh-word is a complementizer, since questionhood is declared by saturation of the verb in the presence of queried roles.

4 Discussion

Human acceptance of sentences as well-formed or not is a fast but shallow process even in attending to semantics, while subsequent critical analysis of linguistic inputs for integration into the hearer’s world view or evaluation for potential causes for action is a comparatively slower process. By underspecifying ambiguity using extensions to Reyle’s UDRT we are able to offer compact semantic representations of ambiguous inputs for later modules to disambiguate into classical DRSs, if disambiguation is necessary. The framework offers a compositional version of UDRT and has been used in specifying a broad coverage fragment of German, inclusive of phenomena like quantificational ambiguity (scope, and collective/distributive), event anaphora, modifiers, the attitudes, and questions.

One aspect of the extension is annotation of distinguished DRFs on UDRS labels, exactly the DRFs available for anaphoric reference. The upper referential index of a DP corresponds to the sum formed from the restricter and quantified variable (see (7)), but this is not sufficient.

- (19) a. Every farmer owned a donkey.
 b. They were less expensive than tractors.

The anaphor in (19.b) has no DRF available to refer to under the present formulation. To accommodate such reference we’d need to generalize the notion of an upper-referential index to a vector $\vec{X}$ of sums over each of the DRFs introduced in the scope of the quantifier. The sums are prohibited from occurrence in the scope of the quantifier creating them, as in classical DRT (Kamp & Reyle

93). In the case quantifiers occurring in the scope position, the upper-referential index (also a vector) of the embedded quantifier would be merged with the vector for the embedding quantifier.

The extended UDRT is implemented in CUF’s typed feature structures. The feature structures given here subsume the semantic representations actually built by performing the functional applications during composition. This semantics construction component is available for integration via an appropriate interface to any syntax that delivers information about functor-argument structure to guide composition, as well as other information (e.g. number and gender) from the syntax-semantic interface, or to a reasoner intended to extract information from underspecified representations, even without enumerating disambiguations (e.g. identifying presuppositions).

References

- (Beil 96) Franz Beil. On antecedent contained ellipsis. To appear in the proceedings of *Paris Syntax and Semantics Conference*, October 12-14, 1995., 1996.
- (Dörre & Dorna 93) Jochen Dörre and Michael Dorna. Cuf — a formalism for linguistic knowledge representation. In J. Dörre, editor, *Computational Aspects of Constraint-Based Linguistics Description*. DYANA-2 Report R1.2.A, Volume 1. ILLC/Department of Philosophy, University of Amsterdam, 1993.
- (Eberle 96a) Kurt Eberle. Disambiguation by information structure in drt. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 334–339, 1996. COLING’96, Copenhagen, Denmark.
- (Eberle 96b) Kurt Eberle. Flat underspecified representation and its meaning for a fragment of german. manuscript, Institute for Computational Linguistics, University of Stuttgart, Project:SFB-B9., 1996.
- (Frank & Reyle 95) Anette Frank and Uwe Reyle. Principle based semantics for hpsg. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, pages 9–16. ACL, 1995.
- (Kamp & Reyle 93) Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer, 1993.
- (Kamp et al. 95) Hans Kamp, Uwe Reyle, and Antje Rossdeutscher. Underspecified representation, discourse inference and lexical meaning. Unpublished manuscript, Institute for Computational Linguistics, University of Stuttgart, 1995.
- (König & Reyle 96) Esther König and Uwe Reyle. A general reasoning scheme for underspecified representations. To appear in *Festschrift for Dov Gabbay*. Kluwer., 1996.
- (Pollard & Sag 87) Carl Pollard and Ivan Sag. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Centre for the Study of Language and Information, Stanford University, CA., 1987.
- (Pollard & Sag 94) Carl Pollard and Ivan Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications, 1994.
- (Reyle 93) Uwe Reyle. Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, 10:123–79, 1993.
- (Reyle 95) Uwe Reyle. On reasoning with ambiguities. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, pages 1–8. ACL, 1995.
- (Richter 96) Frank Richter. Unorthodoxe ruminationen. ms. University of Tuebingen., 1996.
- (van derSandt 92) Rob van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377, 1992.