Semantics of Spatial Prepositions in the Grammar *NorSource*

Lars Hellan and Dorothee Beermann May 2009

Introduction

This note presents how the semantics of spatial prepositions is represented in the computational grammar *NorSource*. The platform for this implementation is the LKB system (cf. Copestake 2002), with a grammar of the type supported by the *HPSG Grammar Matrix* (cf. Bender et al. 2002}) and employing the semantic formalism MRS. Through its embedding in MRS, the framework is in the tradition of systems where meanings of complex expressions are explicitly composed from the meanings of their parts and the way in which they are combined, with identities of operator-bound variables representing identities of participants in the situations expressed. The analysis to be presented is also in traditions of cognitive semantics as represented, e.g., in Talmy 2000, Jackendoff 1990, and others, as insights from these can be adequately represented in the present formalism. (Cf. Verspoor 1997, Davis 2000, Davis and Koenig 1999, and Trujillo 1995, to mention some of the predecessors of the present system.).

Much of current research in formal and computational grammar assumes that prepositions have a binary relational logical structure. In MRS representations in the tradition referred to, the two arguments are standardly labelled ARG1 and ARG2, where ARG2 represents the NP governed by the preposition, and ARG1 represents the 'external' argument, which can be, for instance, the index of a nominal or verbal head if the prepositional phrase (PP) is a modifier, and the 'logical subject' of the preposition if the PP is used in a predicative function (like in *he is in China* or *he put the book on the desk*) (see section 4 for a general discussion of the 'ARG' attributes). In both types of cases, if the preposition has a spatial meaning, the ARG1 represents the item 'situated', and ARG2 the 'site'; these notions will be referred to also by the terms *FIG* and *GRND* (see below). (For PPs functioning as oblique arguments (like in *rely on John*), it is a more open question what to assign as value of ARG1; these PPs are not addressed in this note.)

To illustrate, for a sentence like *the boy sits behind the house*, a standard MRS representation will provide a 'bag' (cf. \citeA{Cop:Fli:Sag:05}) of elementary predications (EPs); EPs are packages of semantic information pertaining to individual words, and a 'bag' of EPs is essentially a list of EPs semantically exposing all the words used in a sentence. In Figure 1 (which is formally incomplete, in omitting, e.g., information having to do with quantification), in the EP representing the preposition *behind*, the 'house' argument is entered as ARG2 of the relation *behind_rel*, and the sitting event as such as its ARG1, identified by the event index *e2*. (Each EP also has an *ARG0*, whose value is the index of that predicate - referential indices are written with *x*, event indices with *e*).

INDEX e2 [E [TENSE : PRES]]

RELS : ⟨ [def_q_rel, LBL h5, ARG0 x4, RSTR h6, BODY h7], [boy_n_rel, LBL h3, ARG0 x4], [sit_v_rel, LBL h8, ARG0 e2, ARG1 x4], [behind_p_rel, LBL h8, ARG0 e6, ARG1 e2, ARG2 x9], [def_q_rel, LBL h11, ARG0 x9, RSTR h12, BODY h13], [house_n_rel, LBL h10, ARG0 x9] ⟩

Figure 1: Standard MRS for *The boy sits behind the house*.

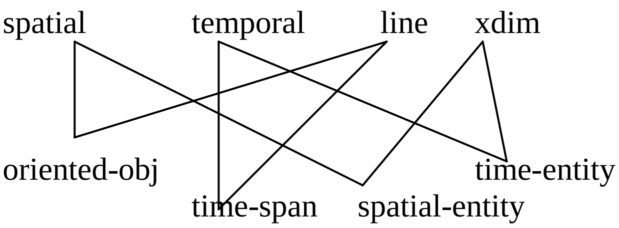
Departing from an MRS structure like this one, NorSource deploys somewhat more differentiated representations of prepositions and their arguments, as we now outline.

1. Basic distinctions

1a. *Line* vs *xdim*

Following Jackendoff 1996, we assume a basic distinction among prepositions residing in whether their arguments represent *one-dimensional* entities or not; one-dimensional entities we refer to as *line* items, as opposed to *xdim* items (subsuming both 'individuals' and events'). The distinction applies to both the spatial and the temporal domain, and is represented in the following top level classification:

(1) A type hierarchy related to arguments of prepositions:



Here our main focus is on the spatial domain. The *line* vs. *xdim* distinction applies to both arguments of a preposition, and we recognize three possible combinations:

(a) ARG1 = *xdim*, ARG2 = *xdim*, called *xdim2xdim* ('xdim-to-xdim', i.e., a relation holding between an *xdim* item and an *xdim* item - cf. section 4 for a closer description of this kind of label);
(b) ARG1 = *line*, ARG2 = *xdim*, called *line2xdim*;
(c) ARG1 = *line*, ARG2 = *line*, called *line2line*.

One may think of these types as representing *situation types*, labelled according to the types of participant roles constituting them (- thus, they are parallel to types like *act-und* - a situation type constituteed by an actor and an undergoer - in the system of Davis 2000; see further section 4).

1b. Subcases of *line*

For spatial *line*-items (i.e., *oriented-objects* in the terms of (1)), there are in principle four ways in which they can be involved as an ARG1 of a preposition, regardless of whether the ARG2 is *xdim* or *line* (the examples in (2a-d) exemplify the type *line2xdim* - i.e., with 'Hamburg' as an *xdim* item – and the examples (2e-h) the type *line2line*, with *along the Rhine* as the *line* item):

(2)

Line2xdim:

a. - as a 'mover' along a path, to be called a *locomotor* (ex: *Ernst ran to Hamburg*)

b. - as an event along a path, to be called a *locomo-event* (ex: *the tour went to Hamburg*)

c. - as an extended object, to be called a *path-obj* (ex: *the road went to Hamburg*)

d. - as a purely oriented object, to be called a *purely-oriented-obj* (ex: *the sign points to Hamburg*)

Line2line:

e. - as a 'mover' along a path, to be called a *locomotor* (ex: *Ernst ran along the Rhine*)

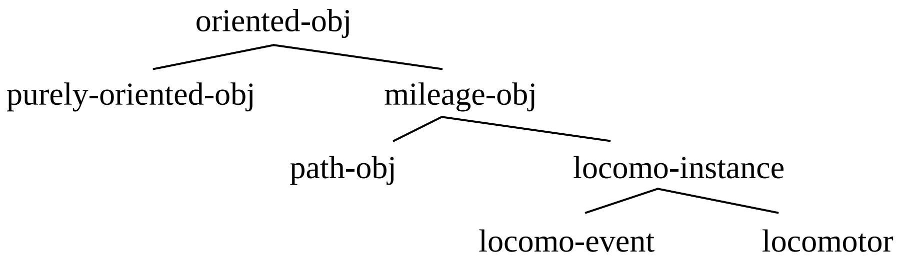
f. - as an event along a path, to be called a *locomo-event* (ex: *the tour went along the Rhine*)

g. - as an extended object, to be called a *path-obj* (ex: *the road went along the Rhine*)

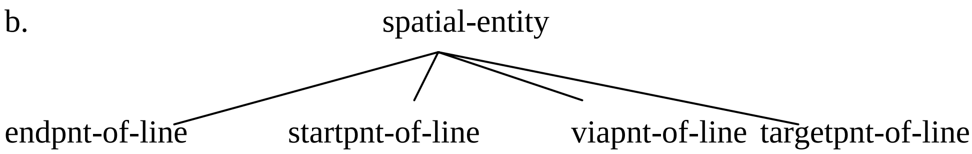
h. - as a purely oriented object, to be called a *purely-oriented-obj* (ex: *the sign points along the Rhine*)

The ARG1 in these examples represent the leaf types in the hierarchy (3a), expanding from *oriented-obj* in (1), with the ARG2 of (2a-d) involving subtypes of *spatial-entity* expanded as in (3b):

(3) a.



b.



The examples in (2a/e) involve a *locomotor* as ARG1 of *to/along*, the examples in (2b/f) a *locomo-event*, the examples in (2c/g) a *path-obj*, and the examples in (2d/g) a *purely-oriented-obj*. As ARG2 of *to*, (2d) has a *targetpnt-of-line*, and (2a-c) an *endpnt-of-line*. In these examples, the exact nature of the ARG1 and ARG2 of *to* and *along* is induced by the accompanying verb and its arguments, however, the general type of the ARG1 is always within the range of *line*, and that of the ARG2 within *xdim* or *line*. Some prepositions may do double duty, as exemplified for *under* in *the dog runs under the table*, where on one reading, *under* encodes an *xdim2xdim* function (the running happens under the table), and on another, it instantiates *line2xdim*, namely when the dog ends up under the table by running there. Prepositions involving *line* arguments are those often referred to as *directional*, but as the example with *under* shows, a prepositional item per se can be underspecified as concerns whether its ARG1 is *line* or *xdim*.

1c. Topological distinctions

Topological distinctions are mainly displayed in *xdim2xdim* type of preposition uses, but to some extent also in *line2xdim* and *line2line* type. The main topological distinctions presently drawn are the following:

Topological features (all boolean):	Definitions:
xdim2xdim:	
FRONT	FIG is in front of GRND
BACK	FIG is behind GRND
EMBEDDED	FIG is embedded in GRND
CONTAINED	FIG is contained in GRND
SCALAR	Relation between FIG and GRND can be quantified
TRANSITIVE	If R(A,B) and R(B,C), then R(A,C)
UPSIDE-OF	FIG is upside in a vertical relation to GRND
DOWNSIDE-OF	FIG is downside in a vertical relation to GRND
INTEGRATED	FIG is integrated into GRND
line2line:	
COEXTENSIVE	The lines are equally long
PARALLEL	Constant distance between the lines
ORIENTATION-SET	Constant orientation of the lines

Table 1: Main topological features

1d. Illustrations

We illustrate the use of the specifications now introduced. First we recast the relevant part of Figure 1 using the new notions (along with still aspectual verb features):

$$\text{RELS}\left\langle \ldots \begin{bmatrix} \text{\_sit\_v\_rel} \\ \text{ARG0 e2} \begin{bmatrix} \text{E.ASPECT} \begin{bmatrix} \text{DYNAMIC} - \\ \text{PROTRACTED} + \\ \text{TELIC} - \end{bmatrix} \end{bmatrix} \\ \text{ARG1 x4} \end{bmatrix}, \begin{bmatrix} \text{\_behind\_p\_rel} \\ \text{ARG0 e6} \begin{bmatrix} \text{ROLE xdim2xdim} \end{bmatrix} \\ \text{ARG1 e2} \begin{bmatrix} \text{ROLE xdim} \end{bmatrix} \\ \text{ARG2 x9} \begin{bmatrix} \text{ROLE xdim} \end{bmatrix} \\ \text{IARG} \begin{bmatrix} \text{CLASS} \begin{bmatrix} \text{FRONT} - \\ \text{BACK} + \\ \text{TRANSITIVE} + \\ \text{SCALAR} + \\ \text{INTEGRATED} - \end{bmatrix} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle$$

Figure 2: Part of enriched MRS for *The boy sits behind the house.*

Here, roles relative to the hierarchy in (1) and the further distinctions exemplified in (2) are specified under the attribute ROLE, an attribute defined inside INDEX. The topological features in Table 1 are introduced under the attribute CLASS, an attribute inside IARG. The role of IARG (for 'inner argument') is a bit like that of ARG0 in specifying properties pertaining to the head or total PP rather than to a specific argument, and distinct from ARG0 in displaying part of the lexical semantics of the item represented (see section 4). For the verb we have represented the main aspectual features with their values for the verb *sit*; aspect is one of the verbal factors interacting with the semantics of prepositions, as will be illustrated shortly.

Figure 3 below illustrates relevant parts of an MRS for *The boy runs to the house.* The aspectual features here represent the telicity of a 'running to' event reaching its end-point. The type *locomotor* is a subtype of *oriented-object*, and thus represents the *line* factor in *to* (as a *line2xdim* preposition). That the role of *oriented-object* is associated with the individual-entity 'the boy', rather than with the running event as such, constitutes one more contrast to the representation in Figure 1 and 2. This point is adressed immediately below.[1]

$$\text{RELS}\left\langle \ldots \begin{bmatrix} \text{\_run\_v\_rel} \\ \text{ARG0 e2}\begin{bmatrix}\text{E.ASPECT}\begin{bmatrix}\text{DYNAMIC} + \\ \text{PROTRACTED} + \\ \text{TELIC} +\end{bmatrix}\end{bmatrix} \\ \text{ARG1 x4}\left[\text{ROLE locomotor}\right]\end{bmatrix}, \begin{bmatrix} \text{\_to\_p\_rel} \\ \text{ARG0 e6}\left[\text{ROLE line2xdim}\right] \\ \text{ARG1 x4} \\ \text{ARG2 x9}\left[\text{ROLE endpnt-of-path}\right]\end{bmatrix}, \ldots \right\rangle$$

Figure 3: Part of enriched MRS for *The boy runs to the house*.

When assigning the ARG1 of the preposition the status of *line*, a mechanism must be supplied which identifies the correct participant in the verbal event as inducer of the line. Thus, in the representation of a directional expression, such as *runs to* in *the boy runs to the house* or *throws... to* in *the boy throws the ball through the window*, it has to be indicated which entity performs the *line* function. Clearly, in *the boy throws the ball through the window*, what induces the path is 'the ball', that is, the direct object, and in *the boy runs to the house*, it is the subject, i.e., 'the boy'. The presumably most straightforward way of identifying the line inducing entity is to equate the ARG1 of the preposition with the relevant participant of the verbal event. This means that the ARG1 of the preposition in these cases is an individual, represented by an *x*-variable, rather than an event.[2]

2. Interaction of *line* specification and topology specification

For *line* expressions with *to*, *from* and *via*, there is no crucial topological information expressed. However, as mentioned, for a sentence like *the boy runs behind the house*, there is one interpretation where as a result of the running, the boy ends up behind the house. Here, *behind* imports into an endpoint directional reading the same topological

---

[1] To the extent that MRSes like those in question are produced by unification grammars, both instances of *x4* in Figure 3 will have the same ROLE specification. Although such an environment is not a prerequisite for the design we are describing, we presuppose it in the specifications given.

[2] See Kracht 2002 for an alternative analysis.

specifications as when it introduces a static site. This situation might be represented through distinct EPs for the activity and its end point (see, e.g., Trujillo 1995), where the end state EP would be the one where the topology of 'behind' would be introduced. Yet, further cases of directional expressions indicate the necessity of topological specification combined with a directionality specification also where no end point is involved. These are cases like *the boy runs through the forest*, and *the boy runs around the forest*. The former involves a *via-point* reading, where the *via-point* is topologically related to the NP governed by the preposition in that 'through' suggests a partial 'in' reading. In the case of 'around', which is an instance of a *line2line* preposition, a relationship of closeness between the circumference of the forest and the path of the running is understood.

In view of this, we want to be able to combine directional and topological specifications in like fashion for the various types of ARG2. For the sentence *the boy runs behind the house*, in the sense where as a result of the running, the boy ends up behind the house, the relevant part of the MRS representation will be as illustrated in Figure 4, where the specifications of ARG1 and ARG2 represent *line* and *xdim*, whereas under IARG, we find what may be seen as the 'constant' or topological contribution of *behind*.

$$
\text{RELS}\left\langle \ldots \begin{bmatrix} \text{\_run\_v\_rel} \\ \text{ARG0 e2} \begin{bmatrix} \text{E.ASPECT} \begin{bmatrix} \text{DYNAMIC} + \\ \text{PROTRACTED} + \\ \text{TELIC} + \end{bmatrix} \end{bmatrix} \\ \text{ARG1 x4} \begin{bmatrix} \text{ROLE locomotor} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{\_behind\_p\_rel} \\ \text{ARG0 e6} \begin{bmatrix} \text{ROLE line2xdim} \end{bmatrix} \\ \text{ARG1 x4} \\ \text{ARG2 x9} \begin{bmatrix} \text{ROLE endpnt-of-path} \end{bmatrix} \\ \text{IARG} \begin{bmatrix} \text{CLASS} \begin{bmatrix} \text{FRONT} - \\ \text{BACK} + \\ \text{TRANSITIVE} + \\ \text{SCALAR} + \\ \text{INTEGRATED} - \end{bmatrix} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle
$$

Figure 4: Part of enriched MRS for *The boy runs behind the house*, in the sense of the boy ending up behind the house.

In *the boy runs through the forest*, the IARG qualifies the topology of the (idealized) mover relative to the ARG2 via-point as one resembling 'in', as displayed in Figure 5:

$$\text{RELS}\left\langle \ldots \begin{bmatrix} \text{\_ run \_ v \_ rel} \\ \\ \text{ARG0 e2} \begin{bmatrix} \text{E.ASPECT} \begin{bmatrix} \text{DYNAMIC +} \\ \text{PROTRACTED +} \\ \text{TELIC +} \end{bmatrix} \end{bmatrix} \\ \text{ARG1 x4} \left[\text{ROLE locomotor} \right] \end{bmatrix}, \begin{bmatrix} \text{\_ through \_ p \_ rel} \\ \\ \text{ARG0 e6} \left[\text{ROLE line2x dim} \right] \\ \text{ARG1 x4} \left[\text{ROLE locomotor} \right] \\ \text{ARG2 x9} \left[\text{ROLE viapnt} - \text{of} - \text{path} \right] \\ \text{IARG} \begin{bmatrix} \text{CLASS} \begin{bmatrix} \text{TRANSITIVE +} \\ \text{SCALAR -} \\ \text{EMBEDDED +} \\ \text{CONTAINED +} \\ \text{INTEGRATED -} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle$$

Figure 5: Part of enriched MRS for *The boy runs through the forest*.

Comparing *The boy runs through the forest* with *The boy runs into the forest*, both will have the CLASS specification given in Figure 5, and the same aspectual specification of the verb. What differentiates them will be the ARG2|ROLE values *endpnt-of-path* for *The boy runs into the forest,* and *viapnt-of-path* for *The boy runs through the forest*.

For the case with 'around' mentioned, the structure will again be analogous, however with the ARG0 of the preposition being *line2line,* and the attribute CLASS will introduce the specifications 'ORIENTATION-SET -, PARALLEL +'.

3. Further features of expressibility

The ARG0 of a preposition can be further used for at least two purposes. One is for encoding quantification of spatial relations, like in *two meters behind the wagon*. Here we want to express that the extent of 'behind' equals two meters, and in MRS terms, this can be stated through a pair of EPs, schematically indicated as in Figure 6 (where a precise rendering of *two meters* is omitted):

$$\text{RELS}\left\langle \ldots \begin{bmatrix} \text{\_ measure} - \text{out \_ rel} \\ \text{ARG1 "two meters"} \\ \text{ARG2 e6} \end{bmatrix}, \begin{bmatrix} \text{\_ behind \_ p \_ rel} \\ \text{ARG0 e6} \left[\text{ROLE line2x dim} \right] \\ \text{ARG1 e2} \\ \text{ARG2 x9} \end{bmatrix}, \ldots \right\rangle$$

Figure 6. Partial MRS-structure of *two meters behind*, e.g., for the sentence *The boy sits two meters behind the wagon.*

Here, the 'behind' relation carries the event index *e6*, and the first EP states that this index is 'measured out' by 'two meters', that is to say, 'behindness' is obtaining to the extent of two meters. We will not go into the mechanisms by which a grammatical parser can produce such an MRS, except for noting that such a representation will be assigned only when the preposition in question is characterized as 'SCALAR +'.

Another situation where coindexation involving ARG0 values carries an expressive function is for expressing whether, in a sequence of two directional PPs like in *He* *walked along the river towards the mountain*, the two PPs represent (temporally) *consecutive* actions, or represent aspects of the *same* action: in the latter case, they will be assigned identical ARG0 values, in the former case, distinct ARG0 values.

4. Roles, arguments and rich semantics

The 'roles' being presently assigned to the ARG1 and ARG2 of prepositions do not constitute a universe delimited to prepositions. As we have seen above for directional constructions, the entity serving as referent of the ARG1 of a preposition like *to* can be a *locomotor* argument relative to the verb, and this is exactly the role we give it also relative to the preposition. Thus, the roles exposed in (1) and (3) should be related to more standardly recognized role inventories including roles like 'agent', 'patient', 'theme', etc.

One design for such an interrelation is that of distinct 'tiers' of semantic specification as proposed in Jackendoff (1987, 1990). Presently, we rather exploit possibilities offered by the type hierarchy design, combining role types from the type universe illustrated in (1)-(3) with roles from the 'agent'-'patient' type universe. Below in Figure 7 is an illustration of the possibilities offered in such a design. Alongside the (1)-(3) types, here instantiated only by the type *locomotor* (with a hierarchical position as defined in (1) and (3a)), two further types *intitiator* and *non-initiator* are introduced, each with familiar subtypes, and cross-decendencies between these new types and the previous ones are then defined, involving, e.g., types such as *init-mover, ag-mover, noninit-mover, affected-mover,* and *theme-mover*, all subtypes of *locomotor* but differentiated according to whether this locomotor is *initiating* the locomotion or not, and in the latter case, whether the locomotor is in addition physically *affected* (as with the object of the verb *throw*) or not. Likewise, if 'agent' is thought as a subtype of 'initiator', one may want to specify an instance of initiated locomotion as having a participant with the role *ag-mover*. The present outline of the system is only partial, as this is not a place for motivating such a system in detail, and Figure 7 is to be seen only as an illustration of how the spatial types defined in (1)-(3) can be defined in a universe shared with more traditional role notions:

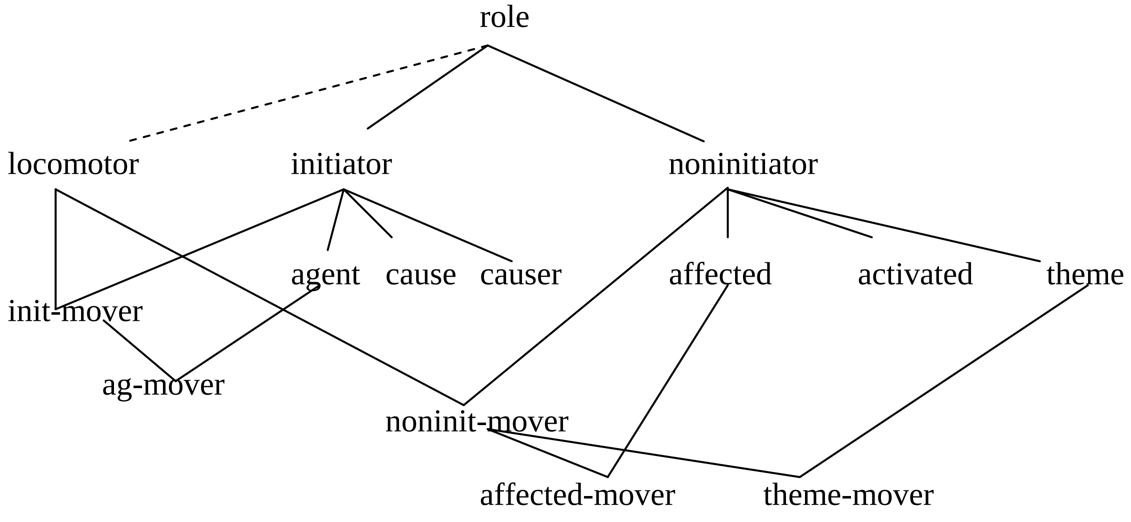


Figure 7. A partial type hierarchy combining spatial/directional types with types reflecting an 'initiator'-'non-initiator' dimension of specification.

In principle, any of the types in the (1)-(3) domain can intersect with types from the 'initiator' system.

Topological properties have been entered on a feature path inside of the attribute IARG, suggesting that these represent more of the 'core' of the meaning of the preposition than values provided for 'ROLE'. In general, it is not always easy to tell what is a 'core' meaning specification and what is a 'role' specification. To consider an example from the verbal domain, suppose we characterize the meaning of *kick* as an instance of *contact-ejection*, meaning "ejection of an item which retains contact with the launching initiator and attains contact with a target". When defining *roles* relative to *kick*, it would in principle be possible to exchange all of this information into role specifications - for instance, one role could be 'entrained launcher', one could be 'entrained launchee', and one could be 'contact target'. However, in much of the literature, there is a tendency to limit 'role' specifications to only the roles that are grammatically realized (which would leave out the 'entrained launchee' - the foot), and for the role labels to be much less specific in their characterizations - ending up with, e.g., 'agent' and 'affected' in the present case. For practical-analytic purposes, the latter format is probably necessary, but it would be useful to have also a locus for the more detailed specification - i.e., for the 'lexical semantics'. The attribute IARG|CLASS is designed to provide such a locus, and for a verb like *kick*, a specification like the one suggested can be encoded there. For *xdim2xdim* prepositions, the topological specifications illustrated above can be regarded as their 'lexical semantics'. The ROLE attributes with the ARG1 and ARG2 of these prepositions, in contrast, are left very much open - thus, the ARG1|ROLE specification of *behind* has to be compatible with the value *xdim* in Figure 2 and *line* in Figure 4, while *to* has an ARG1|ROLE necessarily of type *line*, in both cases with the possibility of added specifications induced by the verb.

There are two more points where the current formalism provides a semantic classification. One is in the distinction between 'referential' and 'event' indices, i.e., between *individuals* on the one hand and *propositions* or *situations* on the other. This distinction is directly encoded in the value of the attribute ARG0 (= INDEX), and is independent of both the ROLE specification and the IARG specification. Thus, the *line* vs *xdim* distinction is orthogonal to the *individual* vs *situation* distinction. In this respect, one can view the ARG0 specification as the exposition of the more general category of the item in question, entered at a locus through which a unification parser will bring it in interaction with similar information provided for other items (more on this in section 5), whereas the IARG contains more specific information, and to a much lesser extent exposing this information to other items. (For ordinary common nouns, IARG will provide information of a type proposed by Magnus and Ross 1988, later proposed by Pustejovsky 1993 under the label 'Qualia'; we do not enter into these here.)

The other, last, point on semantic classification resides in the use of the attributes ARG1, ARG2, ARG3, etc. To some extent, these may seem like replicae of syntactic grammatical functions, as noted, largely representing only those participants which are realized (or localized) grammatically (that includes the ARG1 of a preposition when this is a VP). However, when extended to the verbal system, there are clear contrasts to the realized syntactic patterns: the agent in a passive construction remains ARG1 even when no more realized syntactically, and in 'raising' constructions (like *John seems ill*), the

ARG-encoding will reflect the assumed logical structure (e.g., "seem_v_rel, ARG1 'John ill' "), not the 'surface' grammatical functions. Thus, the 'ARG' distribution encoded for lexical items is indeed a semantic view of its arguments, although a grammatically close one (for instance, one that closely corresponds to 'deep structure' grammatical functions in a classical transformational framework). The order-assignment (of '1', '2', etc.) among the ARGs may in turn seem to follow a dependency hierarchy rather closely, each ARGn+1 being one that would not be instantiated unless ARGn were instantiated. For the representation of prepositions, since there is no 'skewing' between deep vs surface constellations involving those, their ARG1 and ARG2 will seem to have a foot equally much in logical structure and surface syntax. In their case, the choice of '1' and '2' may again be seen as dependency motivated, with *adverbs* being the alternating category employing only ARG1.

In the formal system supported by the Matrix framework and LKB, each attribute (spelled with capital letters) is introduced (or 'declared') by a unique type (and inherited by its subtypes). The ARGn attributes are declared by subtypes of *relation*, the subtype carrying ARG1 and ARG2 belonging to type *arg12-relation*. Further subtypes of this type can be distinguished according to which ROLE values ARG1 and ARG2 have, so that relevant subtypes in our connection are *line-xdim-relation*, *xdim-xdim-relation*, and so forth, and even more finegrained, *locomotor-endpnt-relation*, *pathobj-viapnt-relation*, and so forth. Concerning the attribute entered inside of IARG, for the technical reason of each attribute having one unique declaring type, the type introducing the attribute CLASS has to be in a different hierarchy from those introducing ARG1, ARG2,..., IARG, and this will be the type *spatial* (cf. (1)). Still another type of relevance is the one carrying the label *line2xdim*, which is used as a ROLE value inside ARG0 of a preposition, and likewise for all the other types of the '*...2...*' form. These values are for exposition only.

As a summary of the above, with the types we have now discussed all exposed in an MRS, the more explicit counterpart of Figure 5, for instance, will look as follows:

$$\text{RELS}\left\langle \ldots \begin{bmatrix} \text{arg1-relation} \\ \text{PRED "\_run\_v\_rel"} \\ \text{ARG0 e2} \begin{bmatrix} \text{E.ASPECT} \begin{bmatrix} \text{DYNAMIC +} \\ \text{PROTRACTED +} \\ \text{TELIC +} \end{bmatrix} \end{bmatrix} \\ \text{ARG1 x4} \begin{bmatrix} \text{ROLE locomotor} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{line-xdim-relation} \\ \text{PRED "\_through\_p\_rel"} \\ \text{ARG0 e6} \begin{bmatrix} \text{ROLE line2xdim} \end{bmatrix} \\ \text{ARG1 x4} \begin{bmatrix} \text{ROLE locomotor} \end{bmatrix} \\ \text{ARG2 x9} \begin{bmatrix} \text{ROLE viapnt-of-path} \end{bmatrix} \\ \text{IARG} \begin{bmatrix} \text{spatial} \\ \text{CLASS} \begin{bmatrix} \text{TRANSITIVE +} \\ \text{SCALAR -} \\ \text{EMBEDDED +} \\ \text{CONTAINED +} \\ \text{INTEGRATED -} \end{bmatrix} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle$$

Figure 5': Part of enriched MRS for *The boy runs through the forest*, with *relation*-types and other attribute-inducing types explicitly entered.

5. Semantics in and without a parsing grammar

The present exposition presupposes the scenario of the semantics as being part of a computational grammar. There are in principle two ways in which this could be done: (a) that the semantics is produced simultaneously with the other aspects of the grammatical analysis: (b) that the semantics is only partially produced together with the grammatical analysis, and, instead, later augmented through a 'post-processing' module. We will comment shortly on this difference, and then also comment on how a semantics along the lines presently developed could be formalized in a non-parsing grammar, i.e., in a 'normal' typed feature-structure based grammar.

One of the challenges coming with producing the semantics outlined above directly with a unification based grammar can be illustrated by a sentence such as (5) (from Hellan and Beermann 2005):

(5) The road behind the house goes to the forest.

The PP headed by *behind* receives a locative reading, and is, relative to the present system, of type *xdim2xdim*, thus its ARG1.ROLE is *xdim.* Under noun modification, this is identical to the ARG0 of the noun modified. In the sentence in question, *goes* in its static use can apply only to NPs understood as representing a *line* type object. This will again come out as a ROLE restriction of the ARG0 of the noun. It follows that in the analysis of (5), the NP-internally induced specification of the noun’s ROLE will conflict with its ROLE specification relative to the verbal domain, thus, the standard machinery of index resolution will exclude a perfectly well formed sentence due to colliding ROLE values.

A more practically annoying type of problem with the present design is that even when specifications are in principle compatible, every compatibility has to be explicitly declared in the type system. For instance, although in the present system (cf. Figure 7) *affected* is a subtype of *noninitiator*, and *noninit-mover* is a declared subtype of *noninitiator*, a case where the types *affected* and *locomotor* have to combine is ruled out unless that combination is also explicitly defined, as here by the type *affected-mover*. It takes not so many items in a type inventory before a full overview over such desired combinations becomes difficult to maintain, and grammar maintenance hampered by the detection and amendment of such missing declarations.

Still another concern, even when all combinatorics runs as desired, is that of over-all processing complexity: when the core grammatical information of a large scale grammar is combined with a complex semantic taxonomy, that affects processing speed, and poses further challenges for reducing parsing ambiguity and parse proliferation.

If instead one uses a post-processing design, then, by assumption, incompatible values could be introduced only *after* the unification mechanisms have applied, and hence a reading of (5) could be displayed where *behind* as its ARG1 takes an *xdim* object and *goes* as its ARG1 takes a *line* object, despite these objects being at the same time represented as one and the same individual. Likewise, over-all processing load would then presumably be reduced for the parsing module. Thus, a post-processing design may seem worth exploring; however, so far no attempt has been undertaken (and problems like that posed by (5) are handled by other means, such as stipulating *road* inherently as *line* and making *behind* accept *line* as its ARG1).

We then consider in what respects the present semantic system reflects being a component in a parsing system. Essential to the 'minimal' aspect of MRS is avoiding nesting of complex information, rather having each word, as processed, import its semantic specification on a par with the specifications imported by all other words, in a list-like structure of EPs, where 'holes' and 'handles' situated inside the EPs are the means of encoding embeddedness of one EP inside another. The structure of each EP by itself has the design of an attribute-value matrix, with typed objects as values. If considering a possible non-implemented equivalent of the present design, it would thus retain most of the EP-internal design, but allow much more nesting of AV-structure. This would come out most clearly in causative structures and clausal complementation structures, but can be indicated also for modification structures and directional structures. For instance, for the constellation in Figure 2, we could have that of Figure 8, and for the constellation in Figure 3 that of Figure 9 (using the conceivable option of representing a directional specification by an attribute 'DIR', making this look like a type of argument, although not one of the core participants):

$$
\ldots,\left[\begin{array}{ll}
\multicolumn{2}{l}{\text{\_behind\_p\_rel}} \\
\text{ARG0 e6}\left[\text{ROLE xdim2xdim}\right] & \\
\text{ARG1} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\text{\_sit\_v\_rel}} \\
\text{e2} & \left[\text{E.ASPECT}\left[\begin{array}{l}\text{DYNAMIC} - \\ \text{PROTRACTED} + \\ \text{TELIC} - \end{array}\right]\right] \\
\multicolumn{2}{l}{\text{ARG1 x4}}
\end{array}\right] \\
\text{ARG2 x9}\left[\text{ROLE xdim}\right] & \\
\text{IARG} & \left[\text{CLASS}\left[\begin{array}{l}\text{FRONT} - \\ \text{BACK} + \\ \text{TRANSITIVE} + \\ \text{SCALAR} + \\ \text{INTEGRATED} - \end{array}\right]\right]
\end{array}\right],\ldots
$$

Figure 8: 'Non-MRS' semantic AVM for *The boy sits behind the house.*

$$
\ldots\left[\begin{array}{ll}
\multicolumn{2}{l}{\text{\_run\_v\_rel}} \\
\text{ARG0 e2} & \left[\text{E.ASPECT}\left[\begin{array}{l}\text{DYNAMIC} + \\ \text{PROTRACTED} + \\ \text{TELIC} + \end{array}\right]\right] \\
\multicolumn{2}{l}{\text{ARG1 x4}\left[\text{ROLE locomotor}\right]} \\
\text{DIR} & \left[\begin{array}{l}\text{\_to\_p\_rel} \\ \text{ARG0 e6}\left[\text{ROLE line2xdim}\right] \\ \text{ARG1 x4} \\ \text{ARG2 x9}\left[\text{ROLE endpnt}-\text{of}-\text{path}\right]\end{array}\right]
\end{array}\right],\ldots
$$

Figure 9: 'Non-MRS' semantic AVM for *The boy runs to the house*.

In Figure 8, the head verb event is directly represented as the ARG1 of the preposition, which is indirectly also what the MRS in Figure 3 does; and in Figure 9, the PP is represented in the DIR argument of the verb. In these structures, recursion is no longer minimal, however, all other aspects of the present semantic analysis are carried over. We conclude that for concerns of general linguistic exposition, essential parts of the present analysis can be retained in a non-implementational setting.

6. Final remarks

The present semantics of prepositions has been used in the application *TrailFinder* (Beermann et. al. 2004). Within the group of DELPH-IN grammars, it goes much beyond what is standardly found – indeed, in most such grammars, all spatial prepositions with be represented in the style of Figure 1. The code needed for the richer version can in most respects be simply be added to a standard Matrix grammar (through a file 'Preposition Semantics' available from the authors); the only aspects not available through this add-on are the diversification of ARG1 for locational vs. directional prepositions, and of course the actual lexical types for prepositions, and lexical entries, employed by possible target grammars.

Appendix. Aligning the system with an existing database of prepositions.

It would be desirable if one can make explicit links between the specificational slots offered in the present system, and specification types offered in other existing and developing resources. As an illustration of the point, in the project 'The Preposition Project' (TPP) (cf Litkowski and Hargraves 2005, and http://www.clres.com/prepositions.html), the specification of the English preposition *through* provides as some of its senses the following information (informally rendered):

Sense 1:
Relation Name: ThingTransited
Complement Properties: opening, channel, or location
Attachment Properties: verbs of motion
Sense 2:
Relation Name: ThingBored
Complement Properties: permeable or breakable physical object
Attachment Properties: verbs denoting penetration
Sense 3:
Relation Name: ThingTransited
Complement Properties: sth regarded as homogenous
Attachment Properties: verbs of motion
Sense 4:
Relation Name: ThingPenetrated
Complement Properties: a permeable obstacle
Attachment Properties: a perceived object; sometimes complement of a verb of perception
Sense 5:
Relation Name: ChannelTransited
Complement Properties: an opening or obstacle
Attachment Properties: copula or verb of location

In these specifications, the category 'Relation Name' seems to play a role similar as the attribute path ARG2|ROLE in our system; 'Complement Properties' would provide a closer qualification of ARG2 (a feature already available in the Matrix format for this purpose is SORT), whereas 'Attachment Properties' relate to ARG1. Given this, the prospect of how the actual inventories of descriptive terms could be worked together, in turn, meets with questions of matching taxonomies, which is beyond consideration here. With the present system holding defined links both into actual classifications like the TPP, to computational grammars, and to general semantic analysis, it ought to be a resource for creating explicit connections between such systems.

References

Dorothee Beermann, Jon Atle Gulla, Lars Hellan and Atle Prange. 2004. Trailfinder: a case study in extracting spatial information using deep language processing. In Ton van der Wouden, Michaela Poss, Hilke Reckman, and Crit Cremers (eds) *Computational Linguistics in the Netherlands 2004:* Selected papers from the fifteenth CLIN meeting, pp. 121-131, Leiden, Netherlands, 2004.

Dorothee Beermann and Lars Hellan. 2004. A treatment of directionals in two implemented HPSG grammars. In Stefan Müller (ed) *Proceedings of the HPSG04 Conference*, Katholieke Universiteit Leuven. CSLI Publications /http://csli-publications.stanford.edu/

Emily M. Bender, Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starterkit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In Emily Bender, Dan Flickinger, Frederik Fouvry, and Melanie Siegel (eds) *Proceedings of the Workshop on Grammar Engineering and Evaluation*, Coling 2002, Taipei.

Ann Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.

Ann Copestake, Dan Flickinger, Ivan Sag and Carl Pollard. 2005. Minimal Recursion Semantics: an Introduction. *Journal of Research on Language and Computation*. 281-332..

Anthony Davis. 2000. *The Hierarchical Lexicon*. CSLI Publications, Stanford.

Anthony Davis and Jean-Pierre Koenig. 1999. Linking as Constraints on Word Classes in a Hierarchical Lexicon. *Language* 76. 56—91.

Lars Hellan and Dorothee Beermann. 2005. Classification of Prepositional Senses for Deep Grammar Applications. In: Valia Kordoni and Aline Villavicencio (eds.).

Ray Jackendoff. 1987. The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry. 18.* 369-411.

Ray Jackendoff. 1990. *Semantic Structures* MIT Press.

Ray Jackendoff. 1996. The proper treatment of measuring out, telicity, and possibly even quantification in English. *Natural Language and LinguisticTheory*. 14. 305-354.

Valia Kordoni and Aline Villavicencio (eds.): *Proceedings of the 2nd ACL-Sigsem Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications, Colchester, United Kingdom, ACL-Sigsem, 2005*

Marcus Kracht. 2002. On the Semantics of Locatives. *Linguistics and Philisophy 25: 157-232.*

Kenneth C. Litkowski and Orin Hargraves. 2005. The Preposition Project. In: Valia Kordoni and Aline Villavicencio (eds.).

James Pustejovsky 1993. *Semantics and the Lexicon,* vol 49 of *Studies in Linguistics and Philosophy.* Kluwer, Dordrecht, The Netherlands

Leonard Talmy. 2000. *Towards a cognitive semanitcs.* MIT Press.

Arthuro Trujillo. 1995. *Lexicalist Machine Translation of Spatial Prepositions.* Doctoral thesis, Cambridge University.

Cornelia Verspoor. 1997. *Contextually Dependent Lexical Semantics*. PhD thesis, University of Edinburgh..