A Minimal Recursion Semantic Analysis of Locatives

Fredrik Jørgensen
University of Oslo
ILN, PO Box 1102 Blindern
0317 Oslo, Norway
fredrijo@student.iln.uio.no

Jan Tore Lønning
University of Oslo
ILN, PO Box 1102 Blindern
0317 Oslo, Norway
jtl@iln.uio.no

Abstract

We consider different types of locative PPs from a semantic and computational linguistic point of view. The semantic analysis is based on a proposal by Markus Kracht, and we show how this analysis can be incorporated into Minimal Recursion Semantics. We have made a pilot implementation in LKB. The resulting system may be applied in a transfer based machine translation system.

1 Introduction

Locative prepositional phrases (PPs) is a challenge both to formal semantics and computational linguistics. Locatives is a topic between the strictly logical and the purely lexical where formal semantics can make a contribution by building complex models.

Locative PPs can either be static, as in (1), or directional, as in (2). Some occurrences can be read in both ways, as in (3).

(1) Kim is running in Central Park.
(2) Kim is running to the school.
(3) The mouse ran under the table.

The goal for formal semantics is to describe the different truth-conditions of these two readings correctly and to decide under which conditions the different readings are available. For computational linguistics, and, in particular for machine translation, it is useful to separate between the two uses since they may translate differently. For example, the static in and directional into in English may both correspond to the Norwegian i.

A particular locative prepositional phrase conveys two types of messages. It depicts a particular region (or a class of regions) in space, and it depicts a certain relationship to that region. All the readings of (3) relate to the region which is under the table. What is conveyed about this region varies between the different readings, however. On the static reading\(^1\), the whole event is located inside this region. The most likely directional reading, the cofinal reading, describes an event starting outside the region and ending up inside the region. There is also a less likely directional reading where the event both starts and ends outside the region but passes through it (the transitory reading). A different directional reading is found in (4) where the event starts inside the region and ends outside it (the coinitial reading).

(4) A cat appeared from under the table.

Here it is the preposition from which carries the direction while under depicts the same region as in (3). The goal for formal semantics is to capture this double semantic content and, in particular, show how it is derived compositionally. From a translational point of view, one challenge is how this differs from language to language. What is expressed with one preposition in one language may be expressed by

\(^1\)We adopt Kracht (2002)'s terminology for naming the different readings. Static corresponds to what some other authors call locative. In addition, we find four directional readings: coinital (source), cofinal (goal), approximative (toward) and transitory (via/through).
two prepositions, or even by locative case, in the other.

Our goal is to handle these phenomena within an experimental machine translation system, the LOGON system (Oepen et al., 2004). The system is based on semantic transfer where the transfer representations are based on Minimal Recursion Semantics (MRS) (Copestake et al., 2003). As starting point (section 2) we take the later developments within formal semantic approaches to locatives, in particular (Kracht, 2002). We then (section 3) present minimal recursion semantics and discuss how the insights from Kracht shall best be implemented. In section 4, we discuss how this can be used for machine translation and illustrate with some examples from the translation from Norwegian into English. In section 5 we briefly present an implementation of the analysis from section 3.

2 Semantic analysis of locatives

Syntactically, some occurrences of locative PPs are normally characterized as complements to verbs, e.g. examples (5) and (6), while other occurrences, like (1), are considered adjuncts to VPs.

(5) Kim put the book on the table.

(6) Kim lives in Oslo and Sandy works there.

Traditionally, there have been two different approaches to the semantics of locative PPs. One has been to consider the locative PP to be referential and denote a region. This seems like the best approach to capture the anaphoric use in (6). The other approach has taken the modificational use of the locative PP as basic and let the PP denote a predicate on the location of objects or events. Finally, some approaches have assumed that both types of denotations are needed for the locative PPs.

In (Kracht, 2002), a novel approach to the semantics of locatives is described. On Kracht’s view, locative PPs consist of two layers. One lower layer, a localiser phrase (LP), defining a region or location, and one higher layer, a modaliser phrase (MP), defining events of motion with respect to this region. The analysis is motivated by both the dualistic nature of locative semantics, and by the syntactic and morphological data on how locatives are realized.

The layering is claimed to be apparent in many languages, realized as different morphemes in locative cases, or combinations of prepositions.

In Norwegian and English, this layering is not frequently explicit, but it is found in constructions like (4), corresponding to:

(7)

Kracht assumes an ontology with the following basic types: $e$ (entities), $t$ (truth-values), $i$ (time points), $v$ (events) and $r$ (regions). Regions are actually path connected subsets of the set of space points, but for our purposes, we can simplify and consider $r$ a basic type.\(^2\) A sentence like (1) says something about the relationship between the region where Kim is running and the region within the park. For this there is a primitive function taking objects to their regions, $loc'$. As objects can be moving, this function has to be time dependent, hence of type $e \rightarrow (i \rightarrow r)$. Then $loc'(cp)(t)$ will be an expression (of type $r$) for the region of Central Park (i.e., $cp$) at time $t$. The LP cannot denote such a simple region, however, but has to be a set of regions, what Kracht calls a neighborhood.\(^3\) In the example, this will be the set of all regions that can be claimed to be in the park.\(^4\) Moreover, also such neighborhoods may vary with time, cf. behind the car. Hence the denotation of the LP has to be a parametrized neighborhood, in formal notation $\lambda t \lambda r [i(loc(cp)(t), r)]$. Here $r$ is a variable ranging over regions, while $i$ is a particular relation between two regions locating the second

\(^2\)We will also make some other simplifications compared to Kracht. In particular, we will not consider plurals.

\(^3\)We will freely speak of sets even though they are handled formally as characteristic functions in the lambda calculus. If $\phi$ is a formula (of type $t$) and $x$ a variable of some type $\tau$, we will talk of $\lambda x[\phi]$, which is formally of type $\tau \rightarrow t$, as a set of objects of type $\tau$.

\(^4\)For the prepositions in it could have worked to let in the park denote the maximal region which is in the park, but that does not work for on the table. For the book to be on the table, it does not suffice that the location of the book is contained in the maximal region which is on the table. The region where the book is located has to make contact with, and be supported by, the table. Moreover, parts of the book does not have to be on the table for the sentence to be true.
region within the first. It can be further defined if we go down to the geometrical structure of regions, but will not be of interest here. The relevant point is that this is the localising aspect of the preposition in, while other prepositions contribute a different local content here.

For adverbial modification, Kracht claims the MP denotes a set of events. The nuclear sentence (8) also denotes a set of events (9) and the modification works as an intersection between the two sets.

(8) Kim is running.

(9) $\lambda e[run^t(e,kim')]$

If we write $L$ for the logical representation of the LP, the static MP can be written $st^t(L)$. Here $st^t$ is a function of type $(i \rightarrow (r \rightarrow t)) \rightarrow (v \rightarrow t)$ which can be defined by:

(10) $st^t(L) = \lambda e[\forall t \in C time^t(e) [L(t)(loc^t(e)(t))]]$

For any time point $t$ during the event $e$, $e$ is located within $L$.

Static PPs are said to be in static mode. In addition to the static mode, Kracht identifies four types of modes for directional PPs: coinitial, cofinal, transitory, approximative. They each contribute a different function $ci^t$, $cf^t$, $tr^t$, $ap^t$ (corresponding to M) of the same type as $st^t$ yielding a result of the form

(11) $ci^t(L) = \lambda e[ci^t(\mu(e), L, time^t(e))]$

(12) $cf^t(L) = \lambda e[cf^t(\mu(e), L, time^t(e))]$

(13) $tr^t(L) = \lambda e[tr^t(\mu(e), L, time^t(e))]$

(14) $ap^t(L) = \lambda e[ap^t(\mu(e), L, time^t(e))]$

The function $\mu$ picks out the mover of the event. This might in some cases be the denotation of the subject of the sentence, in other cases the denotation of the object.

The $ci^*$-relation between the three arguments $\mu(e), L$ and $time^t(e)$ is shorthand for a more complex expression of a relationship between the three. In (4), this is the set of events $e$ (related to a location $L$) such that (in prose) “the mover of the event is in $L$ at the start of the event, and outside $L$ at the end”, corresponding to the natural meaning of from. The other relations $cf^*, tr^*$, $ap^*$ are also shorthand for more complex expressions of relationships between $\mu(e), L$, time$^t(e)$.

Kracht claims that LPs occur in restricted contexts. We saw that LPs may be selected by a modaliser (e.g. from), but LPs may also serve as complements of verbs. This adds further potential to the theory, as Kracht’s theory may be applied to verbs like put. Such verbs take both static and directional complements, and express directionality in both cases. The two-layering of Kracht enables us to let the verb select the full PP (MP) as the syntactic argument, but only the denotation of the LP as its semantic argument. The verb can thus express the movement, while the PP contributes the goal of the movement to the semantics.

3 MRS analysis

3.1 Introducing MRS

We now turn to the task of expressing the semantics of Kracht in a meta-level language for logic representations, Minimal Recursion Semantics (MRS) (Copestake et al., 2003). MRS is a framework for computational semantics allowing underspecification. The semantic representations in MRS can be considered a meta-level language for representing an underlying object language which is typically first-order logic with generalized quantifiers. Example (16) is a MRS structure for the sentence in (15). The basic building blocks are the elementary predications (EPs), like $dog(x)$ and $chase(x,y)$, corresponding to atomic formulas.

(15) Every dog chases some cat

(16) $\{h0, h7, \{h1: every(x,h2,hA), h3: dog(x), h4: some(y,h5,hB), h6: cat(y), h7: chase(x,y)\}, \{h0 = q h7, h2 = q h3, h5 = q h6\}$

MRS is often described as flat semantics, as the MRS structures are non-recursive. This is achieved by labeling all EPs with handles. In the example $h3$ is the label on the predication $dog(x)$. In addition, an EP may have handles as arguments, cf. the

---

5 Kracht is not explicit with respect to how the function $loc^t$ extends from individual arguments to events. We have assumed that it also here results in a function from time points to regions. Hence, we have made some modifications to the corresponding expression in (Kracht, 2002, p.193).
argument $h2$ of $\text{every}(x, h2, hA)$. Quantifiers introduce special relations in an MRS, corresponding to generalized quantifiers. The first argument of a quantifier relation is the bound variable, the second is the restriction, and the third is the body. These handles are employed in the mapping from the flat meta-level representation to the object-language expressions with explicit representations of scope.

Scope underspecification is represented in MRS by having handles in argument position of EPSs that are not the labels of other EPSs. The handles are linked (or equated) in the construction of object-language expressions. The unconstrained handle arguments $hA$ and $hB$ in (16) may be linked in two ways to construct the FOL expressions in (17) and (18).

(17) $\text{some}(y, \text{cat}(y), \text{every}(x, \text{dog}(x), \text{chase}(x, y)))$

```
  h0: some(y, h5, h1)
    h5: cat(y)  h1: every(x, h2, h7)
      h2: dog(x)  h7: chase(x, y)
```

(18) $\text{every}(x, \text{dog}(x), \text{some}(y, \text{cat}(y), \text{chase}(x, y)))$

```
  h0: every(x, h2, h4)
    h2: dog(x)  h4: some(y, h5, h7)
      h5: cat(y)  h7: chase(x, y)
```

To constrain the possible linkings of handles and thereby the possible scope relations, MRSs may contain equality modulo quantifiers. The constraint for the handle $h$ and the label $l$ (of some EP), $h = q l$, called a qeq constraint, expresses that either $h = l$, or some non-repeating chain of one or more labels of quantifier EPSs intervene. An object language formula is described by the MRS structure only if it respects all the constraints and all variables are bound.

The general form of an MRS can be considered a four-tuple $\langle GT, LT, R, C \rangle$ where $R$ (relations) is the bag of EPSs, and $C$ (constraints) is a bag of all the qeq constraints. In addition, $GT$ (global top) is the label of the topmost EP of the structure, while $LT$ (local top) is the label of the topmost EP of a substructure. The $LT$ is needed for rules of semantic composition. Furthermore, the $GT$ is always qeq to the $LT$ of the root phrase.

Semantic composition in MRS is performed as follows: The bag of EPSs ($R$) and the bag of qeq constraints ($C$) of the daughter nodes are appended on the mother node. Furthermore, the global top ($GT$) of the daughters are equated on the mother mode. Scopal and intersective combination are treated differently. For intersective combination, the local tops ($LT$) are equated. For scopal combination, the $LT$ of the mother is equal to the $LT$ of the scoping daughter, and an qeq constraint is added between the scopal argument of the scoping daughter and the label of the scoped-over daughter.

For quantifiers, the scopal argument is the restriction of the quantifier, while the body of the quantifier is left unconstrained. This means that the handle variable in restriction position is qeq to the label of the restriction’s $LT$. For example, if every corresponds to the expression (19) and cat corresponds to (20) then every cat corresponds to (21).

(19) $\langle h0, h1, \{h2: \text{every}(x, h3, h4)\}, \{\}\rangle$

(20) $\langle h0, h5, \{h5: \text{cat}(x)\}, \{\}\rangle$

(21) $\langle h0, h1, \{h2: \text{every}(x, h3, h4),
                  h5: \text{cat}(x)\}, \{h3 = q h5\}\rangle$

### 3.2 MRS analysis of modaliser phrases (MPs)

Kracht’s analysis is based on higher order functions and functional application. How can that be ported to a system based on (first order) predication where composition is built by unification? And how shall the phenomena be described in a grammar? We will work our way top down starting with how MPs are combined with other phrases and then proceed to the inner structure of MPs.

First of all, we follow the neo-Davidsonian approach taken in the large-scale English Resource Grammar (ERG)\(^6\), an HPSG grammar using MRS as the format for semantic representation. Verbs introduce events, and adverbialex intersective modification involve unification of the event variables of the verbal relation and the modifier relation. Both models are now event based considering the semantics

\(^6\) [http://www.delph-in.net/erg/](http://www.delph-in.net/erg/)
of a sentence to be a set of events. This facilitates the task. Moreover Kracht (2002) considers all adverbial locative PPs (i.e., MPs) as intersective modifiers. They denote (characteristic functions of) sets of regions. But sets—even higher order ones—are exactly the type of structures that can be represented by EPS. Hence the MP can be represented with an EP containing an event variable. Schematically, (22) and (23) are composed into (24). As we described in the previous section, intersective adverbial modification in MRS is expressed by unification of the labels of the intersected EPS. Shared labels are in MRS treated as implicit conjunction, and unification of labels will thus have the same effect as conjoining the two EPS. The event variables are unified in the grammar.

(22) \[ \text{[A cat appeared]}' = \langle h0, h5, \{ h1: \text{some}(x, h2, h3), h4: \text{cat}(x), h5: \text{appear}(e, x) \}, \{ h2 = q h4 \} \rangle \]

(23) \[ \text{[from under the table...]}' = \langle h6, h7, \{ h7: \text{"from under the table"}(e') \}, \{ \ldots \} \rangle \]

(24) \[ \text{[A cat appeared from under the table]}' = \langle h0, h5, \{ h1: \text{some}(x, h2, h3), h4: \text{cat}(x), h5: \text{appear}(e, x), h5: \text{"from under the table"}(e') \}, \{ h2 = q h4, \ldots \} \rangle \]

### 3.3 MRS analysis of localiser phrases (LPs)

We now turn to the question on the inner structure of the MP; the representation of the LP and how the MP is constructed from the LP. This is more delicate as an L representing an LP is a higher order function—of type \((i \to (r \to t))\)—which in Kracht’s description both occurs as a function (relation) as in example (10) above and as an argument as in example (11). Our solution is based on the fact that in actual representations of sentences we only see such Ls in argument positions. Ls in functor (or predicate) positions have to be considered when inspecting the truth conditions in more detail but does not have to show up in the representations. These occurrences can be calculated from the representations by axioms like (10) and (11). We can therefore consider type \((i \to (r \to t))) a primitive type \(l\) in our representation language.

The function \( e' \) takes two arguments, \( L \) and an event, and yields a truth value. Hence it can be represented in MRS as a relation taking two arguments. Ignoring the analysis of the table, this yields the MRS representation in (26) for the MP in (25).

(25) \[ \text{[MP from [LP under [NP the table ] ]]} \]

(26) \[ \langle h0, h1, \{ h1: \text{ci}'(e, l), h1: \text{under}(l, x), \{ \} \} \rangle \]

(27) \[ \langle h0, h1, \{ h1: \text{ci}'(e, h2), h2: \text{under}(x), \{ \} \} \rangle \]

A different possible analysis in MRS is shown in (27). In (Copestake et al., 2003, p. 27-28), it is argued that there is no need to make the temporal entities explicit in the analyses of temporal locatives, e.g. *on Sunday*. A simple, generic looking grammar is preferable, provided that enough information is included for a more explicit representation to be derived. We agree with this claim in general. But when it comes to locatives, introducing a location variable may prove useful, in particular for providing an antecedent for anaphoric binding, cf. example (6).

The full semantic representation of example (4) is shown below:

(28) \[ \text{[A cat appeared from under the table]}' = \langle h0, h5, \{ h1: \text{some}(x, h2, h3), h4: \text{cat}(x), h5: \text{appear}(e, x), h5: \text{"from under the table"}(e') \}, \{ h2 = q h4, h6 = q h8 \} \rangle \]

#### 3.4 Variants of MPs

Semantically, we assume locatives have an MP structure of the form [ M [ L NP ] ]. Syntactically, this can be realised in several different ways. In the paradigmatic example considered so far, one preposition, *from*, contributes the modaliser M (example 29) while another preposition, *under*, contributes the localiser L (example 30).

(29) The mouse appeared *from* under the table.

(30) The mouse appeared from *under* the table.

(31) The mouse ran *under* the table.

(32) Felix sat *here*.

(33) The cat ran from *here* to *there*. 
Sentence (31) represents the by far most common case, namely that the locative preposition introduces two semantic relations, a modaliser and a localiser relation. In (32) and (33), we find localisers that infer the location from the real-world context, rather than relating the location to the prepositional complement. In our implementation, these localisers are one-place predicates, assuming that they are insensitive to the linguistic context in determining the location. Thus, *here* in (32) constitutes a full MP, whereas *here* and *there* in (33) are complements of modalisers, and therefore analyzed as full LPs.

The locatives (29–33) contribute to the MRS representations as shown in (34–38).

(34) \[ \ldots \text{from under} \ldots \ldots' = \langle \ldots h_5: \text{ci}'(e, l) \ldots \rangle \]

(35) \[ \ldots \text{from under} \ldots \ldots' = \langle \ldots h_5: \text{under}(l, x), \ldots \rangle \]

(36) \[ \ldots \text{ran under the} \ldots \ldots' = \langle \ldots h_5: \text{st}'/\text{cf}'/\text{tr}'(e, l), \]
\[ h_5: \text{under}(l, x) \ldots \rangle \]

(37) \[ \text{Felix sat here}' = \]
\[ \langle \ldots h_5: \text{st}'(e, l), h_5: \text{here}(l) \ldots \rangle \]

(38) \[ \ldots \text{from here to there} \ldots' = \]
\[ \langle \ldots h_5: \text{ci}'(e, l_1), h_5: \text{here}(l_1), \]
\[ h_5: \text{cf}'(e, l_2), h_5: \text{there}(l_2) \ldots \rangle \]

### 4 Translation

#### 4.1 Classifying locative prepositions

The modaliser and localiser functions constitute two natural dimensions to classify locatives along. The modaliser dimension seems to have an upper boundary, as Kracht (2002) claims to have found evidence for five modalisers. The localiser dimension has in principle no upper boundary. Transitive locatives are classified schematically along these dimensions in Table 1. As we see, only the transitive localiser *at* has an approximative mode. And the localiser function *under* lacks the coinitial mode. The coinitial mode of *under* may be expressed by the complex PP *from under*, as we have seen.

Such a classification of locatives is of interest monolingually, as different locatives of the same mode share syntactic and semantic properties. They appear in the same syntactic constructions and have the same impact on telicity. The LP as a semantic level is also interesting in other contexts, such as locative PPs as complements of verbs of putting. These verbs take directional and static PP complements, and express movement into the LP (denotation) independently of the directionality of the PP. But the classification is also of interest bilingually, as it seems to explain interesting differences across languages, which in turn can be exploited for translational purposes.
4.2 Translating transitive locatives

The Norwegian preposition *i* is ambiguous between a static and directional reading, cf. Table 1. This is also true for the English preposition *in*, whereas the English *into* is unambiguously directional. But the Norwegian preposition may be disambiguated by a the context, e.g. by a preceding intransitive locative (*inne/*inn*), or by factors such as the telicity of the event. We now have the repertoire to transfer the locative *i* as unambiguous, when disambiguated by the context. And furthermore, *i* may be translated to the locative *into* in the cases where *i* is unambiguously cofinal, and to *in* in the cases where *i* is in fact ambiguous.

The translation flow in a semantic-based transfer MT system is shown below. The modes are treated as interlingua predicates, whereas the localiser functions are language-dependent, and thus translated in the transfer module.

**Translation flow of the locative *i (in):***

<table>
<thead>
<tr>
<th>Cofinal <em>i</em></th>
<th>Ambiguous <em>i</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>[ i_{cf} ]</td>
<td>[ i_{st/cf} ]</td>
</tr>
<tr>
<td>↓ analysis ↓</td>
<td>↓ analysis ↓</td>
</tr>
<tr>
<td>{cf'(e, l), l_{no}(l, x)}</td>
<td>{st/cf'(e, l), i'_{no}(l, x)}</td>
</tr>
<tr>
<td>↓ transfer ↓</td>
<td>↓ transfer ↓</td>
</tr>
<tr>
<td>{cf'(e, l), in'_{en}(l, x)}</td>
<td>{st/cf'(e, l), in'_{en}(l, x)}</td>
</tr>
<tr>
<td>↓ generation ↓</td>
<td>↓ generation ↓</td>
</tr>
<tr>
<td><em>into</em></td>
<td><em>in</em></td>
</tr>
</tbody>
</table>

4.3 Translating intransitive locatives

Intransitive locatives, or context-dependent locatives, provide even more interesting mismatches between Norwegian and English locative expressions. In Table 2, we see that Norwegian has a richer inventory for expressing context-dependent spatial relations. An example is the Norwegian locative *herfra*, which translates to the English complex *from here*. These kinds of mismatches are now easily explained. In Norwegian, we find a lexicalization for the combination of the coinitial mode and the *here* (‘proximate’ or ‘close-to-speaker’) localiser function. In English, we do not, and the two semantic relations (modaliser and localiser, respectively) are realized as separate lexical items.

(39) Translation flow of the locative *herfra*:

\[
\text{herfra} \downarrow \text{analysis} \downarrow \{ci'(e, l), her'_{no}(l)\} \downarrow \text{transfer} \downarrow \{ci'(e, l), here'_{en}(l)\} \downarrow \text{generation} \downarrow \text{from here}
\]

4.4 Disambiguation of mode

We mentioned that a mode-ambiguous locative may be disambiguated by the context. This disambiguation may take different forms. Consider the following sentences:
The ambiguous locative in (40) is disambiguated in favor of the cofinal mode in sentence (42). When the locative inn occurs alone, as in (41), it is analyzed syntactically as a PP and semantically as an MP, just like i huset in (40).

How shall then the PP complex in (42) be analyzed? There are several possibilities. Either both locatives may be considered adjuncts, or the former may select the latter as a syntactic complement. In either case, the semantic relationship between the two may either be intersective modification or functor and argument. Some data (far from conclusive) suggest that the intransitive locative inn selects the second PP as its argument. In particular, observe that the first locative not only selects the mode of the second PP, but can actually change it. A PP that is interpreted as static when modifying a VP alone, is interpreted as a cofinal when preceded by inn. This is illustrated by the following examples:

(43) Helikopteret fløy over_st/tr byen.
    Helicopter.DEF flew over city.DEF.
    ‘The helicopter flew over the city.’

(44) Helikopteret fløy inn over_cof byen.
    Helicopter.DEF flew in over city.DEF.
    ‘The helicopter flew (to) over the city.’

These data suggest that the second locative is not actually disambiguated, as the cofinal mode is not an option without the first locative present. This has led us to represent the PP complex as the latter PP being a complement of the former locative, and that the former is treated as a modaliser as shown in (45).

With respect to the semantics of this construction, it may seem as inn functions as a marker for directionality, much in the same respect as fra. When the locative inn occurs alone as in (41) it is analyzed as cofinal mode with respect to a location which is the interior of some contextually relevant object (e.g. inside the house). In (42), there is no reference to the interior of any object. Rather, the expression means that there is a movement into the location expressed by the LP complement. Thus, we analyze inn as a cofinal modaliser in these constructions.

5 Implementation

MRS is a general format for representing different formal object languages and may be combined with different syntactic models. It has in particular been combined with HPSG grammars where it has been used in the large English Resource Grammar (ERG). HPSG (Pollard and Sag, 1994) is a unification-based non-transformational theory, using multiple inheritance type hierarchies and unification of typed feature structures as central formal mechanisms. Also the analysis we have presented so far is compatible with different syntactic models. We have made a small fragment of an HPSG grammar to illustrate the analysis. This fragment was implemented in the Linguistic Knowledge Base (LKB) system (Copestake, 2002). The LKB system is an implementation environment for writing HPSG grammars. The grammar was based on the Matrix “starter kit” for building HPSG grammars (Bender et al., 2002).

The Matrix contains types for building MRSs as the semantic content of a sentence compositionally. Principles restricting the syntax-semantics interface, as described by (Copestake et al., 2001), are also encoded in the Matrix. For instance, a sign (phrase or word) only exposes its handle, main internal argument, and external argument to the rest of the grammar, such that e.g. a verb cannot control the arguments of its complement.
Figure 1: AVM representation of signs
To give an idea of the implementation we have included two signs in figure 1 corresponding to the two prepositions from and under in the phrase from under the table. The CAT feature contains the syntactic information while the CONT feature contains the semantic content, in form of a feature/value representation of the MRS. The syntax-semantics interface obeys the principles described in (Copestake et al., 2001), as the features in CAT only access the HOOK of the semantic contents of its syntactic complement. During processing, the RELS of the daughters of a node is collected to a set on the mother.

We see from the representation of the sign for fra that it is a preposition which selects a PP, and more specifically a LP. This is represented here by constraining the value CAT to be prep, and the value INDEX to be of type l. These constraints express that the sign is a locative denoting a location (in contrast to an event), thus a LP. These representations are somewhat simplified, as the actual descriptions of these signs are quite complex, having to deal with many other aspects of the grammar. A detailed description of the implementation can be found in (Jørgensen, 2004).

6 Conclusions

We have shown how the analysis of (Kracht, 2002) can be implemented in a computational linguistic system based on Minimal Recursion Semantics. In particular, we have shown how the higher order functions may be represented in the flat semantics. We have then shown how the resulting system may be applied in a transfer based machine translation system.

References


