A treatment of directionals in two implemented HPSG grammars

Dorothee Beermann
NTNU Trondheim

Lars Hellan
NTNU Trondheim

Proceedings of the 11th International Conference on
Head-Driven Phrase Structure Grammar

Center for Computational Linguistics, Katholieke Universiteit Leuven

Stefan Müller (Editor)

2004

Stanford, CA: CSLI Publications

pages 357–377

Beermann, Dorothee & Lars Hellan. 2004. A treatment of directionals in two implemented HPSG grammars. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 357–377. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2004.20.

Abstract

This article proposes a semantics of directional expressions in Norwegian and German, regarded as VP modifiers. The analysis uses Minimal Recursion Semantics, as an integrated part of the HPSG Grammar Matrix-backbone. Directional expressions are analyzed as predicating of an individual, the 'mover'. Context dependent directionals like *here* receive a decomposed analysis. Telicity values reflecting the presence of various types of directional and locative expressions are computed.

1 Introduction

We here report on an implementational approach whose main goal is to explore the incorporation of *lexical semantics* as part of the semantic interface of grammars with standard design. The grammars in question cover Norwegian and German, and are based on the *HPSG Grammar Matrix* (henceforth: the Matrix).[1] Both grammars are distinct from existing larger grammars for the same languages, allowing some experimental flexibility not readily available in larger grammars. We will show that with a rather modest supply of resources for the encoding of semantic information, we are able to compute aspectual values for directional and other constructions of some complexity, and to perform some amount of semantic decomposition reflecting the presence of multiple parameters encoded in locative and directional adverbs.[2]

Section 2 gives a background introduction to the formal basis of our proposal, and in particular to the representational format of *Minimal Recursion Semantics* (MRS) and how it is integrated in the grammar formalism.

In section 3 we discuss three domains for which we would like to suggest a more pronounced semantics than so far provided in the most developed HPSG grammars (those for English, Japanese and German): these domains are directionals, locative anaphors such as 'here', 'there', etc., and aspect specification.

Section 4 exposes the formal analysis of the phenomena presented in section 3, and in section 5 we provide some expansions of the analysis and some discussion.

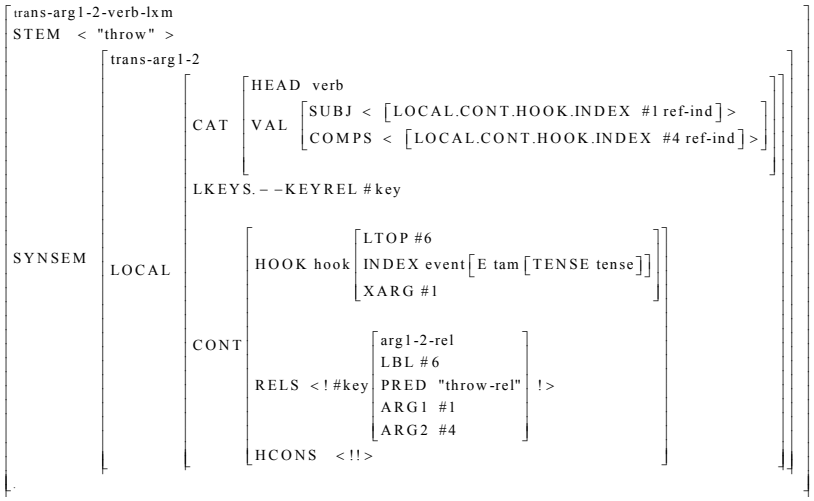
[1] (Bender, Flickinger and Oepen 2002, Bender and Flickinger 2003, Flickinger, Bender and Oepen 2003). Based essentially on the *ERG* (*English Resource Grammar*) (Flickinger 2000), and also JACY (Siegel and Bender, 2002), the Matrix provides types inducing a system of typed feature structures, the essential lexical and syntactic types, together with MRS representations. Grammars of some size developed from the Matrix include Norwegian (Hellan and Haugereid 2003), Italian (Gonella and Mazzini 2003), Spanish, Korean and Greek.

[2] The grammars in question have been built into a small end-to-end application for extracting information from hiking route descriptions for use in a web portal for hiking route queries, called 'Trailfinder' (cf. Beermann, Gulla, Hellan and Prange 2004).

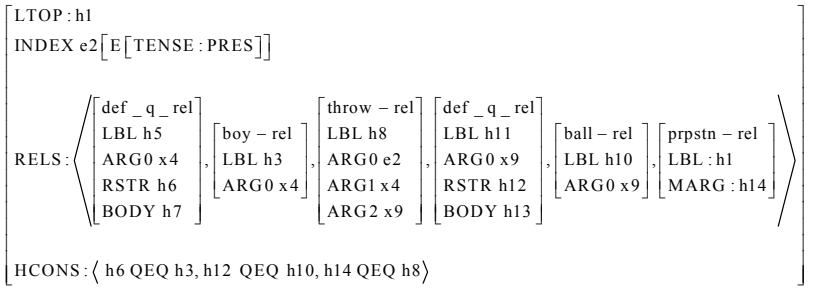
2 Minimal Recursion Semantics in a Matrix grammar

The main format of semantic representation in a Matrix grammar is MRS (cf. Copestake et al. 2003, Flickinger et al. 2003). MRS representations are designed to represent in 'flat' structures the embedding of scopal relations as found in standard logical and semantic formalisms, with an expressive capacity at least that of predicate logic. As currently used, MRS representations accommodate argument structure information, variable binding, scope of quantifiers and other operators. The interaction between grammatical specification and MRS representations (using Matrix.0.5) can be partly seen from (1) and (2) below, where the lexical specification of the verb *throw*, represented in (1), is reflected in the MRS produced for the sentence *the boy throws the ball* ((2)):

(1) Feature structure description for the lexical item *throw*:



(2) MRS for the sentence *The boy throws the ball*:



In accordance with a standard MRS set up as illustrated in (2), for any constituent *C* (of any rank), the RELS list in its CONT specification is a 'bag' of those elementary predications (EPs) which are expressed inside *C*. Lexical specifications are standardly not decomposed and in most cases introduce one EP, labelled according to the stem form. In a lexical specification like that for *throw*, thus, there is one EP, entered as value of LKEYS--KEYREL, and reentered in the RELS list. In the build-up of larger constituents headed by *throw*, all elements on the daughters' RELS lists will be entered on the RELS list of the larger constituent, including the EP from *throws*'s RELS list. An example of such a larger constituent is the sentence *The boy throws the ball*, for which the RELS list displays six EPs, of which one reflects *throw*; cf. (2). Two EPs here reflect the subject argument of the verb, as is seen by the coindexation of the ARG1 of the verb and the ARG0 (corresponding to 'bound variable') of the determiner and the noun, and two EPs reflect the object argument, similarly indicated by variable identity; the remaining EP represents 'message type' (cf. Ginzburg and Sag (2001)). Scope properties are expressed in the HCONS list, 'x QEQ y' meaning essentially that x scopes over y. HCONS thus records the scopal tree of the constituent in question, as outlined in Copestake et al. (to appear).

3 Directional expressions

By 'directional expressions' (or simply 'directionals') for the three languages under consideration (Norwegian, German and English - note that we use English as exemplifying language unless a point specific to one of the other languages is made), we understand preposition-headed and adverb-headed expressions like *to the church*, *from here*, *through the park*, *up*, *up through the chimney*, etc. Such expressions can qualify either the subject of a sentence, as in

John ran from here to the church

or the object, as in

John threw the ball from here to the church.

We treat directionals as V- or VP- modifiers; this position is motivated in section 5.

For subject-oriented directionals, a well-known property is their impact on the *telicity* status of the modified VP: Expressions like *to the park* induce telicity, by the criterion of allowing combination with expressions like *in an*

hour, as opposed to *for an hour*, while expressions like *along the river* allow *for an hour* as a further qualifier, but disallow *in an hour*;[3] thus:

(3)

a. *John ran to the church* **for two hours*
in two hours

b. *John ran along the river* *for two hours*
**in two hours*

Likewise, *place* adverbials align with *along the river*:

c. *John ran in the wood* *for two hours*
**in two hours*

To be noticed is that when both types are represented, it is the telicity-inducing type which prevails:

d. *John ran along the river to the church* **for two hours*
in two hours

The same distinctions apply to directional *adverbs*. These, however, unlike directional prepositions, carry inherent *contextual anchoring* of one sort or another. Consider the following pairs in Norwegian: [4]

(4)

a. 1. *hit* 'to here', as in *Gutten løper hit* ('the boy runs to-here')
boy-DEF run-PRES to-here

2. *dit* 'to-there', as in *Gutten løper dit* ('the boy runs to-there')

b. 1. *opp* '(to) up', as in *Gutten hopper opp* ('the boy jumps up')
boy-DEF jump-PRES to-up

2. *ned* '(to) down', as in *Gutten hopper ned* ('the boy jumps down')

What the pairs in (4) have in common is that they anchor the directionality relative to a contextual correlate given in the discourse. For sentences in isolation, such as (4a), this correlate is by default the *speaker*, the meaning in both cases being that the motion has as its *endpoint* a location related to the speaker - in (a.1) *close* to the speaker, in (a.2) (more) remote from the speaker.

[3] Cf., e.g., Smith (1991, 1997).

[4] We will not attempt to give an exhaustive picture of the usage of the various adverbs and prepositions of Norwegian. Thus, for instance, we ignore those occurrences of 'place' adverbs where they function as predicatives (and then presumably as participant predicators), as in (i):

(i) Jon satte vasen her
Jon put the vase here

In (4b), an endpoint is again expressed, but in these constructions, the correlate can be any given *landmark* (not excluding the speaker), so that in (b.1), the endpoint of the motion is *high* relative to that landmark, in (b.2) *low*.

Exactly these same distinctions turn up in a series of corresponding *place* adverbs:

(5) a.1. *her* 'here', as in *Gutten løper her* ('the boy runs here')
2. *der* 'there', as in *Gutten løper der* ('the boy runs there')
b.1. *oppe* '(at) up(stairs)', as in *Gutten hopper oppe* ('the boy jumps up')
2. *nede* '(at) down(stairs)', as in *Gutten hopper nede* ('the boy jumps down')

In (5), what is contextually correlated is not endpoints of movements, but places of events. Thus, the events in (5a) *take place* in locations related to the speaker, and in (5b) related to some discourse-salient landmark.

As exemplified by (4b) vs. (5b), the contrast between event/place modifying adverbs and participant modifying adverbs is morphologically flagged in most cases by the absence vs. presence of a word final *-e*. This contrast holds systematically in Norwegian, and is further exemplified in (6), with (a) as directional adverbs, (b) as place adverbs:

(6)
a. 1. *bort* '(to) away', as in *Gutten løper bort* ('the boy runs away')
2. *vekk* '(to) away', as in *Gutten løper vekk* ('the boy runs away')
3. *ut* '(to) out', as in *Gutten løper ut* ('the boy runs out')
4. *inn* '(to) in', as in *Gutten løper inn* ('the boy runs in')
5. *hjem* '(to) home', as in *Gutten løper hjem* ('the boy runs home')

b. 1. *borte* 'away', as in *Gutten er borte* ('the boy is away')
2. *vekke* 'away', as in *Gutten er vekke* ('the boy is away')
3. *ute* 'out(side)', as in *Gutten løper ute* ('the boy runs outside')
4. *inne* 'in(side)', as in *Gutten løper inne* ('the boy runs inside')
5. *hjemme* 'at home', as in *Gutten løper hjemme* ('the boy runs at home')

Corresponding to the directionals in (4) and (6a) is furthermore a series of 'along path' directionals, listed in (7):

(7) *hitover* 'here-wards', as in *Gutten løper hitover* ('the boy runs herewards'); *ditover* 'there-wards'; *oppover* 'upwards'; *nedover* 'downwards'; *bortover* 'away'; *utover* 'outwards'; *innover* 'inwards'; *hjemover* 'home-wards'

The items in (7) differ from those in (4) and (6a) in that the 'along-path' concept, expressed by 'over', appears in a compound form together with the contextualized place morpheme. (7) is thus reminiscent of German contextualized directional adverbs such as '*hierher*' 'to here' and '*dorthin'* 'to there'. Different from Norwegian, however, German also contextualizes the path description such that '*hierher'* means towards the speaker seen from the speaker's perspective and '*hierhin*' means towards the speaker seen from the hearer's perspective. Common to all of these adverbs is that the orientation of the movement relates to some contextually understood entity or place. For all cases, we call this the *contextual correlate*.

In the case of the end-point directional adverbs and along-path directionals in this group, there are in effect *two* entities contextually invoked: the endpoint/path as such, and then the correlate relative to which the end-point/path is situated (as higher than, lower than, close to, remote from, etc.). With a place adverb like *her*, in contrast, only *one* contextually invoked entity is relevant – in the default case the speaker, expressed as *close to* the event.

Cutting across the contextual correlate distinctions for adverbs is the value they induce for telicity: the end-of-path directionals induce telicity, the other two types not.

Prepositions, in contrast to the adverbs now illustrated, have no contextually determined inherent participants. In the other respects, they cross-classify like the adverbs (as exemplified above for English): directional prepositions can induce telicity, such as Norwegian *til*, German *zu* ('to'), or they contribute atelicity, such as Norwegian *langs,* German *entlang* ('along'); and in the latter respect, they group along with place prepositions. Intuitively speaking, the circumstance that prepositions are those words which govern an NP that explicitly indicates a correlate of movement or location, while adverbs are words lacking such an NP, will seem to match the contrast with regard to contextual determination: one might hypothesize that in their semantics, both prepositions and adverbs are two place relations, and that, in this domain, what characterizes adverbs is that their 'semantic object' is contextually induced, while for prepositions, it is syntactically induced. The analysis to be presented below implements this view.

Central to the analysis is also a distinction between directional PPs/adverbs and event modifiers in what they are taken as being predicated of. For Norwegian we saw already that a majority of contextualized adverbs reflect this distinction morphologically. For German, further compelling evidence for such a differentiation comes from the group of prepositions that either govern *accusative* or dative case, where the accusative evokes a directional interpretation while the *dative* case forces a 'locative' interpretation. This is illustrated in (8):

(8) Case in German

a.	Der	Junge rennt	in der	Kirche.
	The.NOM	boy runs	in the.DAT	church
	The boy runs in the church.			
b.	Der	Junge rennt	in die	Kirche.
	The.NOM	boy runs	in the.ACC	church
	The boy runs into the church			

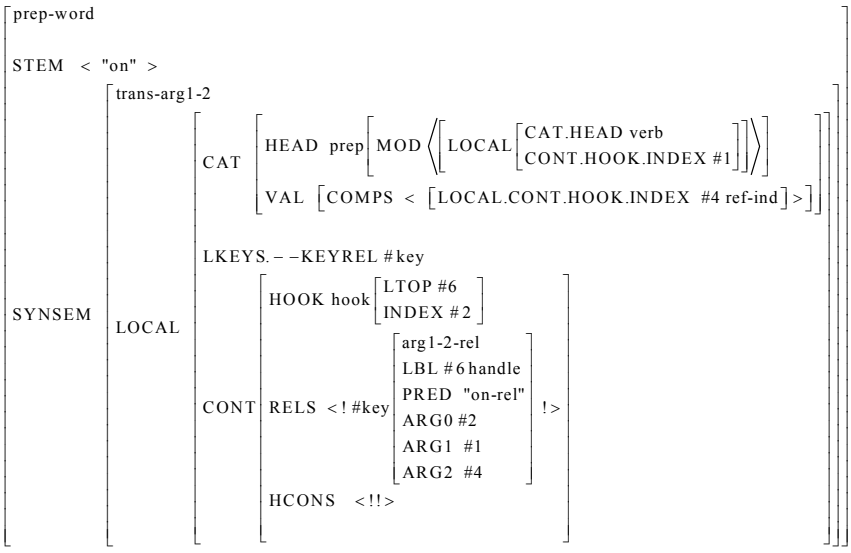
The implementation of this differentiation is laid out in 4.1 below.

4 Analysis

4.1 Modifier predication of a *participant* vs. modifier predication of an *event*

Implementations of VP-modifiers (like that found in the ERG) commonly construe them as *event modifiers*; thus, in the analysis of *John runs in the wood*, the PP *in the wood* is construed as predicated of the index associated with *run*, and, hence, of the index of the event as such. This analysis corresponds to a paraphrase such as 'John's running is in the wood'. Technically, the value of the ARG1 of the preposition is reentered with the event index (INDEX) of the verb, as illustrated in (9) below. The feature MOD here introduces the item *modified by* the preposition, i.e., the verb, and the value of ARG1 of the preposition is re-entered with the INDEX value of the MOD item:

(9) Feature structure description for the event modifier preposition *on*:



While grammars like ERG (as of 2003) extend this approach to directionals as well, an alternative treatment of directionals, which we will advocate here (and which is prevalent in much of the literature, e.g., Jackendoff (1990), is to construe directionals as predicated of *the mover*, i.e., the entity performing the directional motion. For *to* as in *John ran to the church*, this is to say that the ARG1 of the preposition is reentered with the ARG1 of the verb, rather than with its INDEX, and that in *John threw the ball to the church*, the ARG1 of the preposition is identical to the ARG2 of the verb. A similar contrast will be recognized for adverbs. We thereby implement the general contrast between event modifiers and directionals (illustrated, e.g., in (8) for German) as a difference in what the preposition's/adverb's ARG1 is coindexed with - the verb's INDEX for event modifiers, and one of the verb's arguments for directionals.

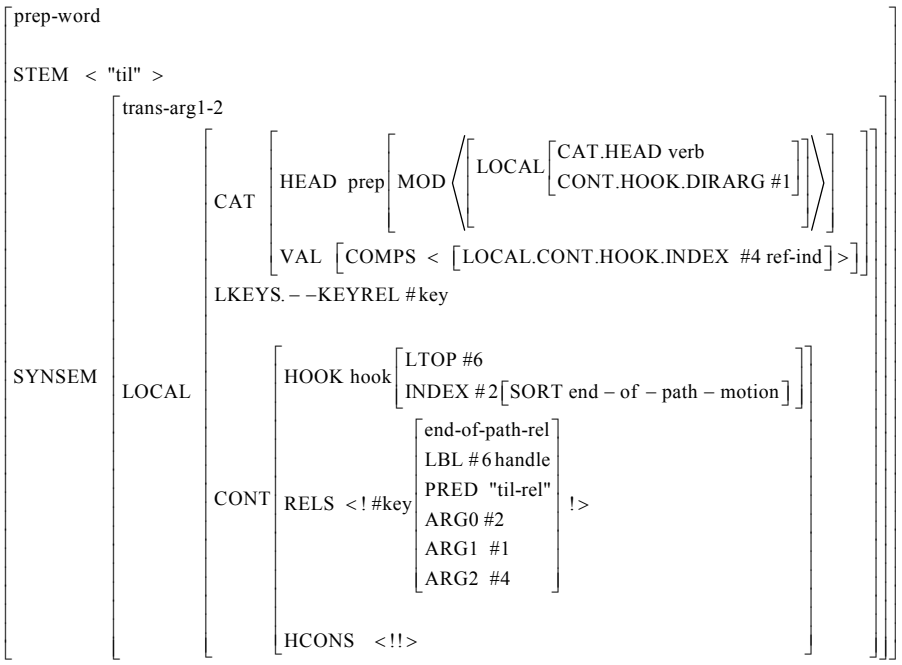
It might be asked if such selection information could alternatively be represented by a more morpho-syntactically flavoured feature, such as 'DIRECTIONAL *bool'*. Although a possibility in principle, we do not see any non-arbitrary way of making such an alternative marking. The proposal that the ARG1 of a preposition like *to* equals the ARG1 of the verb, and thus (in the standard case) is a *referential index*, in contrast, seems intuitively reasonable. It also receives some support from constructions like *The road goes to the church*, where, although the subject is not a mover, the phrase *to the church* clearly qualifies a 'thing'-like entity ('the road') rather than an event. This position will now be explicated and illustrated.

The analysis of *endpoint-of path* prepositions here proposed is illustrated in (10), for the Norwegian preposition *til* ('to'); this illustrates the approach we are taking. Crucial here is the identification of the preposition's ARG1 with the 'Mover' argument of the verb. As we have seen, the latter may be either an ARG1 or an ARG2, according to what type of verb it is - either one whose *subject* is a mover, or one whose *object* is what is set into motion, as for *kaste* ('throw'). The specification of *til* as such should be independent of this choice, i.e., of whether it qualifies a subject or an object. To implement such an independence, we enrich the semantic specification of the verb with a feature which exposes which of its arguments - if any - is a 'Mover', and make this index accessible to the preposition's MOD specification.

In the Matrix inventory of features, there is already one feature, XARG (for 'external argument'), at the path `SYNSEM|LOCAL|CONT|HOOK`, which serves for exposing arguments, e.g., for control specifications. All lexical items with predicative content have an XARG feature, and for verbs, it is typically the argument expressed by the *subject* which is exposed by this feature. Since a verb like *throw* will need XARG to expose its subject like all other verbs, this is not a feature that can be used for exposing a 'Mover' argument in general. We

therefore introduce a new feature for this purpose, called *DIRARG*.[5] This is a feature which will be 'activated' only for verbs one of whose arguments performs a movement, and synsem-subtypes will be distinguished among both intransitive and transitive synsems, to accommodate the presence or absence of such an argument. The presence of this feature thereby serves as a mark of a 'motion-verb', and will be present whether or not a directional modifier actually occurs - it thus reflects an 'inherent' classification of verbs, according to their 'path'-taking potential. In (10), this feature is exposed inside the MOD specification of the preposition, analogously to where an event modifying preposition exposes INDEX as in (9):

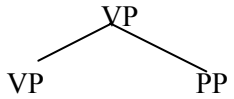
(10) Feature structure description for the preposition *til* ('to') as a participant modifier:



[5] Technically, this is done by specifying a subtype of the type *hook* (the type which introduces the feature XARG - cf. (1)), and let this subtype - *dirhook* - introduce a feature DIRARG in addition to XARG. DIRARG will be activated by verbs expressing motion, in such a way that if it is the *subject* of the verb which performs the motion, then the verb's DIRARG = ARG1, and if it is the the *object* of the verb which performs the motion, the verb's DIRARG = ARG2. These identities are encoded in the appropriate subtypes of intransitive and transitive verb SYNSEMs, respectively, while for *to*, its ARG1 will always be identical to the DIRARG of the verb, enforced by the specification under the MOD attribute of *to*, as seen in (10).

4.2 Inducing telicity values

The AVM in (10), in addition to elements explained above, also contains a specification for the attribute SORT. SORT is a feature inside the path of INDEX, allowing for further semantic specification of an item. In the present case, it introduces a specification *end-of-path-motion* which serves in inducing a *telicity* value for verb phrases composed with a *til*-PP. This is effected through a combinatory rule for the constellation



which specifies the head V projection with

SYNSEM|LOCAL|CONT|HOOK|INDEX|TELIC +

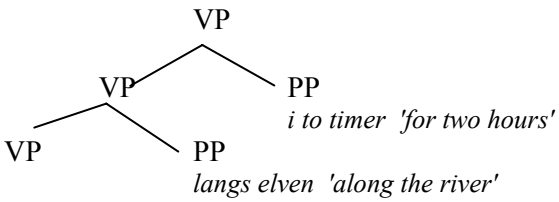
in case the PP is marked as in (10). A preposition like *langs* 'along', in contrast, will carry a corresponding specification *along-path-motion*, which fails to induce this effect, leaving the V projection specified as

SYNSEM|LOCAL|CONT|HOOK|INDEX|TELIC bool

A PP like *i to timer* 'for two hours' requires of its sister VP head that it be marked

SYNSEM|LOCAL|CONT|HOOK|INDEX|TELIC -

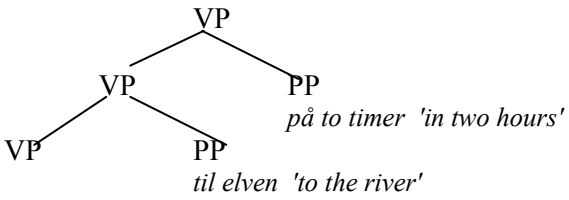
The type *bool* is compatible with '-', hence a PP with *langs* allows for a combination with *i to timer*:



In contrast, a PP like *på to timer* 'in two hours' will combine only with a V projection marked

SYNSEM|LOCAL|CONT|HOOK|INDEX|TELIC +

which means that a PP with *til* will provide a licensing specification for *på to timer*, as in



but prevent *i to timer* from combining. In this way, the 'overriding' effect of prepositions like *to* and *til* illustrated in (3) is captured.[6]

The SORT specification associated with a *place*-preposition (or event modifiers, more generally) is in turn compatible with *bool* as telicity value, and thus combine with *i to timer* 'for two hours', in the same way as *langs* and other non-end-point prepositions.

In this sketch, 'telicity' addresses only this factor in so far as it is affected by VP modifiers; in this respect, all verbs are by themselves unmarked, i.e., characterized by *bool*. A distinct feature accommodates inherent aspectual features of verbs, as well as combinatorial aspect induced by the presence vs. absence of objects. As our present concern is with directionals, we will not go into these aspects of verb semantics, nor how they interrelate with features providing specification of whether the verb expresses movement, reflected in the presence of the feature DIRARG.[7]

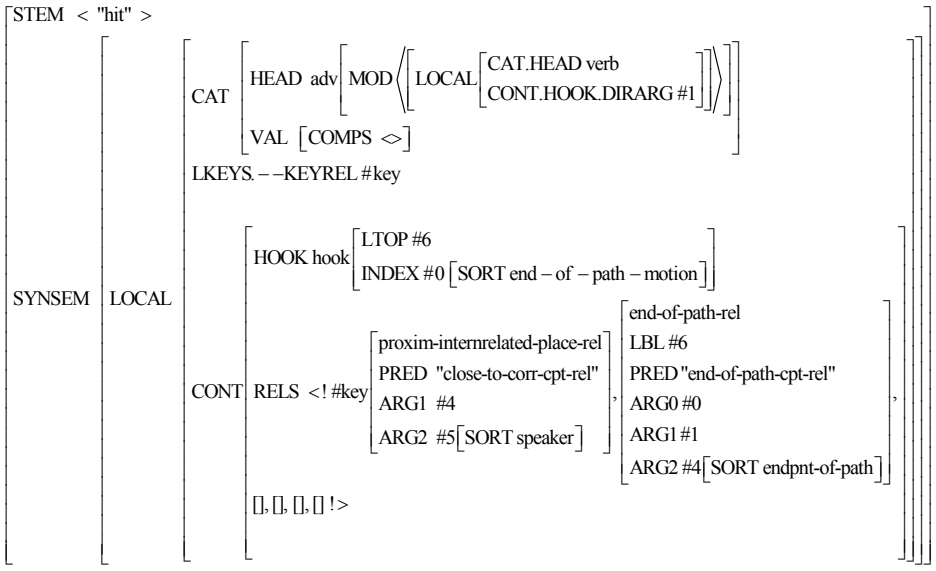
4.3 Inherent participants and decomposition

We now turn to adverbs with inherent participants, as discussed in section 3. We first consider similarities and contrasts between the end-of-path preposition *til* illustrated in (10), and the adverb *hit* 'to-here', represented in (11) below. Both of these are analyzed as predicating of a *participant*, hence, as seen in (11), also in the MOD value for *hit*, the head verb is specified with regard to its DIRARG.

[6] It may be noted that this account will license *John ran in two hours*. Although this is not what one would most obviously want, the construction possibly has an interpretation like 'within two hours, John brought it about that he could run', and we therefore regard this case of apparent overgeneration as possibly harmless.

[7] This is clearly a domain where use of the feature SORT is again relevant. Since the practice in the Matrix is to keep SORT values as atomic (i.e., SORT introduces no features by its own), and verb semantics is well known as requiring some complexity of specification (cf., e.g., Davis 2001, Davis and Koenig 2001, Wechsler 1995), it is an issue for further exploration exactly what features will be necessary in this area.

(11) Feature structure description for the adverb *hit* 'to-here':



A further parallel is that *hit*, just as *til*, has the value *end-of-path-motion* for the path SYNSEM|LOCAL|CONT|HOOK|INDEX|SORT, whereby its capacity of inducing telicity is represented in exactly the same way as it is for *til*. Still a parallel, announced near the end of section 3, is that both have on their RELS list an EP of type *end-of-path-rel* with an ARG1 and an ARG2, representing the semantic parallelism between directional prepositions and adverbs noted in the discussion.

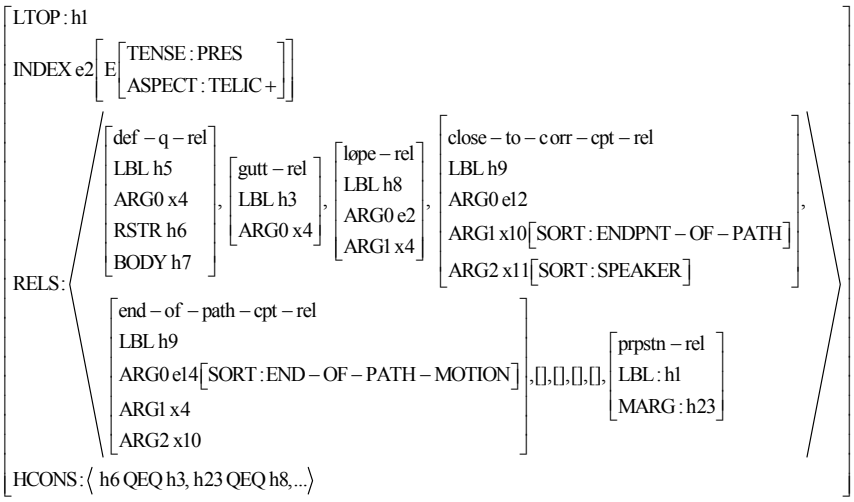
The interesting difference between (10) and (11) resides in their specifications under RELS: while in (10) there is only one item on this list, in (11) there are (essentially) two. These are binary abstract predicates: the *end-of-path-cpt-rel* relates the Mover (its ARG1) to an endpoint (its ARG2), while the *close-to-corr-cpt-rel* relates the endpoint (its ARG1) to the inherently understood speaker (its ARG2).[8] The items tagged #4 and #5 are the inherent participants.[9]

[8] The part *-cpt-* (for 'conceptual') of these labels refers to the circumstance that these predicates are introduced via *decomposition* of lexical items, and thus do not have a spelling matching that of a lexical item.

[9] As all participants carrying an x-type variable need to be quantified to yield wellformed logical representations (i.e., wellformed MRSs), and there are no overt quantifiers doing this, such quantifiers (with PRED value "pronoun-q-rel"), along with 'restriction' values (with PRED value "zero-pron-context-corr-rel"), have to be introduced into the lexical specification of this item as well; the last four items on the RELS list serve for these purposes. For our present purposes, these four EPs are marked by '[]' in (11), as well as (12) below.

(11) represents a lexical item, and its RELS list is its contribution in the semantic composition of any constituent in which it takes part. We show with (12) an example of an MRS composed by contribution of *hit*:

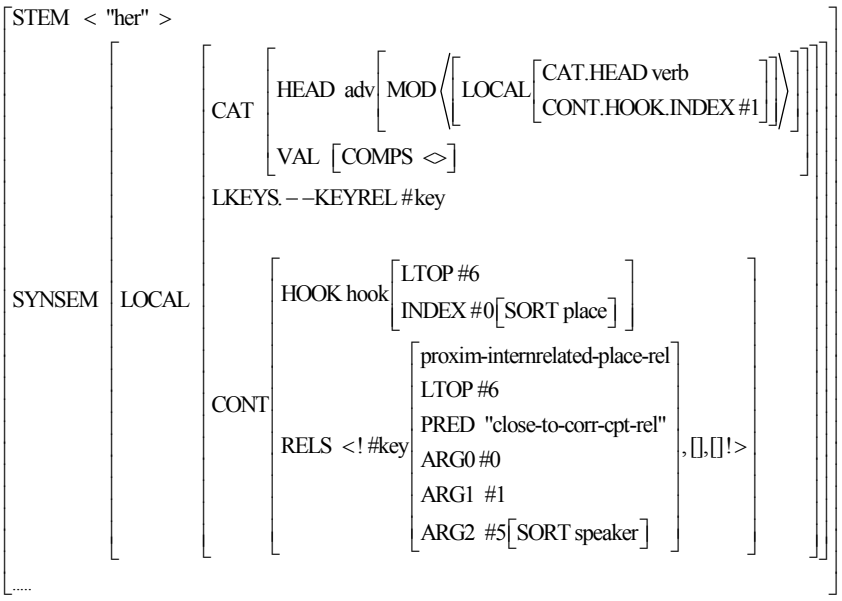
(12) MRS for the sentence *Gutten løper hit* ('the boy runs to-here'):



It will be noted that the whole MRS construct has the specification TELIC +, induced by the SORT specification of *end-of-path-rel*, in pecisely the same way as indicated above for a preposition like *til*.

Having illustrated the contrast between items with (as with *hit*) and without (as with *til*) inherent participants, but like status as to what they are predicated of (viz., the *Mover*), let us next illustrate the contrast between a predication of a Mover (*hit*) and a predication of an event (*her*), along with a contrast between having two and having one inherent participant. At the same time, we illustrate the representation of two distinct words having a common 'semantic feature', here 'closeness to speaker', a property in common between *hit* and *her*:

(13) Feature structure description for the adverb *her* 'here':[10]



Like the preposition *on* illustrated in (9) above, this is a modifier qualifying the *event* index of the verb. The circumstance that both *her* and *hit* have the PRED value *close-to-corr-rel* in their RELS specification, with ARG2 as 'speaker', accounts for their common feature of referring to a place 'close-to-speaker'. The SORT specification *place* being one which does not induce telicity, *her* is correctly predicted to be combineable with *i to timer* 'for two hours'.

On the basis of the feature structures shown, we briefly indicate how some of the other contrasts illustrated in section 3 are encoded, focussing on endpoint-adverbs, with (4a2, 4b) and (6a) repeated:

(4a) 2. *dit* 'to-there', as in *Gutten løper dit* ('the boy runs to-there')
boy-DEF run-PRES to-there

(4b) 1. *opp* '(to) up', as in *Gutten hopper opp* ('the boy jumps up')
boy-DEF jump-PRES to-up

2. *ned* '(to) down', as in *Gutten hopp ned* ('the boy jumps down')
boy-DEF jump-PRES to-down

(6a) 1. *bort* '(to) away', as in *Gutten løper bort* ('the boy runs away')
2. *vekk* '(to) away', as in *Gutten løper vekk* ('the boy runs away')
3. *ut* '(to) out', as in *Gutten løper ut* ('the boy runs out')
4.*inn* '(to) in', as in *Gutten løper inn* ('the boy runs in')

[10] Since this item has only one inherent participant, the abbreviatory place holders on the RELS list are now only two, as opposed to four in (11)/(12).

In relation to the lexical representation (11) of *hit*, and the example (12) of its projection into a sentential MRS, the representation of *dit* will differ only in having *remote-from-corr-cpt-rel* rather than *close-to-corr-cpt-rel* as value at the path SYNSEM.LOCAL.LKEYS.--KEYREL.PRED. The representation of *opp* will differ from that of *hit* in having *high-relto-corr-cpt-rel* rather than *close-to-corr-cpt-rel* as PRED-value, and having ARG2| SORT specified as *landmark* rather than *speaker*. The representation of *ned* will differ from that of *opp* only in having *low-relto-corr-cpt-rel* rather than *high-relto-corr-cpt-rel* as PRED -value. The representation of *bort* will differ from that of *opp* only in having *remote-from-corr-cpt-rel*; the representation of *ut* will differ from that of *opp* only in having *outside-relto-corr-cpt-rel* as PRED -value, and the representation of *inn* will differ from that of *opp* only in having *inside-relto-corr-cpt-rel* rather than *high-relto-corr-cpt-rel* as PRED -value.

These specifications are induced through a small hierarchy of *relation* types, of *semsort* types, and of *word* types, where the latter regulate the *relation* types and the specifications in the first EP in the RELS lists (just illustrated above), and the *relation* types induce the *semsort* types. As indicated by this survey, the specifications serve very much like specifications in a componential analysis table, however, being restricted to features of clear grammatical relevance in the languages.[11]

This subsection has demonstrated a mechanism of semantic componential analysis which preserves important characteristics shared between componentially decomposable and non-decomposable words; examples of the latter are the parameters of participant vs. event modification, and telicity- vs. non-telicity inducing function. We now discuss some of the assumptions made further.

5 Directionals as adjuncts

5.1 The status of iteration

In constructions like

John ran from here to the church

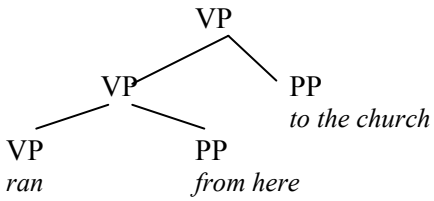
[11] Thus, in German semantic parameters of motion such as the telic/atelic distinctions as well as contextual anchoring are, e.g., introduced in the morphological form of particles/ adverbs, as for example (cf. Müller 2002):

(i)	Er	springt	auf /	rauf /	hierauf
	he	jumps	up	on top	on top of this

To demonstrate the application of semantic decomposition for machine translation, we aim in later work at showing an MRS-based set of transfer rules for a fragment of directionals for an MT component between the two languages.

we treat both *from here* and *to the church* as adjuncts, with an assumed structure roughly of the form (14):

(14)



An indication seemingly immediately in favor of such an analysis is the circumstance that such directionals can occur without any upward bound on their number, as in (15):

(15) *John ran from here down via the park along the creek up to the church*

If these directionals were to be analyzed as arguments, it would suggest that standard assumptions about fixedness of valence were to be abandoned, *run* apparently having an indefinite number of possible arguments.

However, we need to contrast constructions like (15) with expansions of constructions like (16a) as in (16b) (from Norwegian - analogous cases would be possible in the other languages): the use of *inn* in (16a) one would treat as an argument, since *sette* 'put' most reasonably should be treated as having two complements in its valence, an object and a locative argument. The multiple occurrences of adverbs and PPs in (16b) will seem to constitute just another instantiation of the pattern in (16a), and hence a case where multiple adverbs/PPs are 'packed' into one argument slot:

(16) a. Jon satte den inn
John put it in
b. Jon satte den inn i boksen i hjørnet bort fra sollyset
Jon put it into box-DEF in corner-DEF away from sunlight-DEF

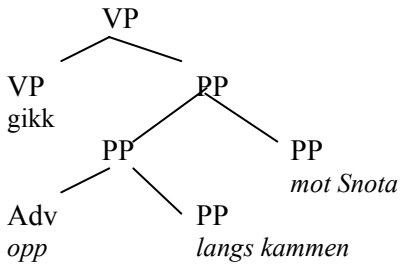
Seemingly, then, multiplicity of occurrence is not by itself a proof that at a 'top' level, more than one constituent is involved. In (16b), this one constituent in turn may be analyzed as an argument. So, perhaps such an analysis could be applicable also for (15)?

A difference between (16b) and (15) is that in (16b), all the adverbs/PPs are read as specifying one and the same location, whereas in (15), each PP specifies a new stretch of movement. It is to be noted that the 'moving on' sense induced by *run* does not by necessity entail that all adverbs/PPs express different stretches - in the sequence *down via the park* in (15), *down* and *via the park* may well qualify one and the same stretch, and similarly for the directionals in (17):

(17) Jon gikk opp langs kammen mot Snota
Jon went up along ridge-DEF towards Snota

Thus, even in the domain of iteration of directionals, one has to recognize the possibility that two or more consecutive directionals co-specify one and the same stretch, or 'leg', analogously to the way the PPs/adverbs in (16b) co-specify one and the same location. For co-specificational clusters like these, it seems that one may well assume a corresponding syntactic clustering - (17), for instance, might receive an analysis like in (18) (although further investigations are warranted to explore what are the possible head-adjunct divisions in such structures - (18) is just one example), and a similar PP cluster would be used in representing the location argument in (16b):

(18)



The crucial point that we would like to make is that in the case of iteration of PPs/adverbs tied to verbs like *put*, the clustering analysis is the *only* analysis relevant, whereas for directionals accompanying a verb like *run*, the clustering analysis is only one of the options, and the other is a successive adjunction analysis as illustrated in (15), reflecting a reading where consecutive stretches are being 'consumed'. For this latter construction type, an adjunction analysis seems the more reasonable option. Consequently, for verbs like *run*, whether or not a directional modifier is simple, as in (15), or a cluster as in (18), it either way attaches as an *adjunct*, whereas for a verb like *put*, a locative or directional constituent is necessarily an *argument*.[12]

In the formal analysis of directional adjuncts, a situation like that depicted in (18) will have a semantic representation where the ARG0s of all the PPs or adverbs clustered together have identical value - this represents the circumstance that they are all the *same leg*. (And correspondingly for a structure like in (16b), ARG0 identity will represent identity of location.) For structures like (15), in contrast, the ARG0s will be distinct, displaying the status of the PPs/adverbs as expressing *distinct legs*.

[12] There is in principle a further position construing directionals as arguments which would not entail that a sequence of directionals necessarily shares stretch/leg specification, namely successive application of lexical rules, expanding the verb valence step by step. Such an approach does not seem to make any empirical gains, although might be of relevance in connection with the considerations mentioned in section 5.2.

5.2 The status of DIRARG

The position taken here, to the effect that multiple directionals can reflect multiple adjuncts, was seen to necessitate the introduction of an extra attribute under the verb's HOOK, namely DIRARG, to which each adjunct will refer in tying its ARG1 to the right argument of the verb. It may be noted that if directionals were instead arguments, always abstractly specified in the valence frame of a verb, then this pairing of ARG1 value with the right verb argument would be done inside the valence specifications of each verb, and then the use of a DIRARG attribute would not be required. Let us consider some factors that might count in the evaluation of this feature.

5.2.1 Verbs of 'co-movement'

Verbs like *follow, chase, pull* and others are commonly interpreted to the effect that the subject argument and the object argument perform the same movement. Would such a situation entail that such verbs have two DIRARGs, and that the ARG1 of each directional is somehow tied to both of those DIRARGs?

We believe that although these verbs clearly need some sort of representation of 'co-movement' in their lexical semantics, this co-movement is not so strict that it warrants a representation of the type alluded to. Typical uses of a verb like *follow*, for instance, tend to fixate on one of the arguments at the time. For instance, in *follow the guests out***,** the actual situation is commonly one where only the object - the guests - actually end up outside (the host may stay inside the doorstep). In *I have followed Lenin to where I am today*, analogously, the 'move' described only qualifies the subject. Thus, it seems that verbs of this type have only one DIRARG, but that they alternate in usage as to which argument is tied to ; either way, the general factor of 'co-movement' is always represented, but at a different level than that of DIRARG coindexation. Thus, the verbs of this type do not seem to necessitate a proliferation of DIRARGs.

5.2.2 Controlled adjuncts

Constructions with directionals may be compared to constructions such as those in (19), where the *as*-predicate may seem in principle able to pick any of the argument functions of the head verb as controller (cf. Beermann 1997):

(19)	a.	They arrived as winners	(subject)
	b.	They burned her as a witch	(direct object)
	c.	They gave Jon the responsibility as the captain	(indirect obj.)

If directionals, as adjuncts, warrant a DIRARG attribute for effecting the control relations, what might constructions like (19) warrant? It may be noted that these control relations are less dependent on specific semantic properties of the verbs involved, and our tentative view is that these control relations should be dealt with at a semantic postprocessing level where coreference resolution more generally may take place. Thus, there is no reason to supplement DIRARG with further features under HOOK for these constructions.

5.3 'Contextual' arguments like 'speaker', 'hearer', 'landmark'

Our semantic representations involving notions like *'speaker', 'hearer', 'landmark',* etc, are obviously non-resolved as concerns anaphora and deixis, and in this respect provide only templates for further development of the grammars involved. Compared with designs of context representation like that in Pollard and Sag (1994), it may be noted that where they have 'SPEAKER' as an attribute under CONTEXT, our notion *speaker* emerges as a type used as value of SORT. A motivation for introducing 'speaker' in this way is that in connection with adverbs like *hit* and *her*, what is contextually really involved is a notion of 'most salient point': in a given discourse, this could be resolved as a place which is described as having been reached, whereas its default subtype would be 'speaker'. For this reason, it is reasonable to have the notion 'speaker' construed as a type in a small hierarchy; this of course does not conflict with using the notion also as an attribute.

6 Conclusion

Although the grammars described push their semantic analysis somewhat beyond what is currently instantiated in most MRS based analysis, the formal devices employed are, apart from the one feature DIRARG discussed in the previous section, confined to those contained in the Matrix inventory. This ensures a smooth interface to the other components of a grammar. The semantic analysis itself follows rather standard assumptions from the general literature, in its invocation of a 'mover' as essential in the analysis of directionals, in its treatment of telicity, and in its decomposition of contextually determined locative and directional adverbs. Being still a sentence based grammar, the marking of contextually salient features awaits the incorporation of context anaphoric resolution, but brings the sentential analysis to a point where it can hopefully interact with devices performing such resolution.

References

Beermann, D. (1997) Syntactic Discontinuity and Predicate Formation. Ph.D. diss., Tilburg University.

Beermann, D, J.A. Gulla, L. Hellan and A. Prange (2004) 'TrailFinder: an NLP web application for hiking information'. Presented at ScanMatrix meeting, Univ. of Gothenburg, October 2004.

Bender, E., D. Flickinger and S. Oepen (2002) 'The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars.' In: *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics,* 8-14, Taipeh, Taiwan.

Copestake, A. (2002) *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.

Copestake, A., D. Flickinger, I. Sag and C. Pollard (to appear) 'Minimal Recursion Semantics: An Introduction' (ms, subm. for publication).

Copestake, A., A. Lascarides and D. Flickinger (2001) 'An Algebra for Semantic Construction in Constraint-based Grammars'. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics* (ACL 2001), Toulouse.

Davis, A. 2001. *The Structure of the Hierarchical Lexicon*. CSLI Publications, Stanford.

Flickinger, D. (2000) 'On building a more efficient grammar by exploiting types', *Natural Language Engineering 6(1): 15-28,* Cambridge University Press.

Flickinger, D., E. Bender and S. Oepen (2003) 'MRS in the LinGO Grammar Matrix: A Practical User's Guide.' Deep Thought Project Report D3.3.

Ginzburg, J., and I. Sag (2001) *Interrogative Investigations.* CSLI Publications.

Gonella, D. and G. Mazzini (2003) 'The Italian Grammar.' Deep Thought Project Report D3.4.

Hellan, L., and P. Haugereid. 2003.) 'The NorSource Grammar - an excercise in the Matrix Grammar building design'. In: *Proceedings of Workshop on Multilingual Grammar Engineering, ESSLLI 2003*.

Jackendoff, R. (1990) *Semantic Structure*. MIT Press.

Müller, St. (2002) *Complex Predicates.* CSLI Publications.

Pollard, C., and I. Sag (1994) *Head Driven Phrase Structure Grammar.* Chicago University Press, Chicago.

Sag, I., T. Wasow and E. Bender (2003) *Syntactic Theory. A Formal Introduction*. CSLI Publications, Stanford.

Siegel, M. and E. M. Bender (2002): 'Efficient Deep Processing of Japanese'. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization. Coling 2002 Post-Conference Work-shop*. Taipei.

Smith, C. (1991, 1997) *The Parameter of Aspect.* Kluwer.