Spatio-temporal Ontologies and Attention

Kai-Uwe Carstensen
University of Freiburg/Germany

This paper presents a new proposal for the design of spatio-temporal ontologies which has its origin in cognitively motivated spatial semantics. It is shown that selective attention not only plays a central role in the characterization of spatial relations but that the representation of attentional aspects also leads to the possibility of defining an ontological upper structure which systematically covers both the spatial and the temporal domain.

Keywords: Spatial expressions, attention, spatial relations, ontologies

1 Introduction*

Computing the meaning of a given linguistic expression (e.g., an expression consisting of **spatial terms**) or generating natural language expressions out of meaning specifications (e.g., **spatial relations**) are central tasks in computational linguistics (cf. Carstensen et al., 2004), and computational semantics in particular. For historical reasons –especially because of the influence of formal logicians like Montague, Kamp and others– computational semantics has developed a strong bias towards formal logic that parallels the development in AI.

Simplifying a bit, the basic tenet of this kind of formal semantics is that one *can and has to* specify the meanings of linguistic expressions by a systematic mapping onto formal models of ("representations" of) the world (which are, as Smith, 1995:9 rightly points out, quite remote from (an adequate understanding of) the common-sense world). As to spatial relation expressions –the main topic of this paper–, they would be mapped onto spatial relations defined by a space calculus (e.g., the RCC calculus of Cohn et al., 1997).

This "objectivist" (Lakoff, 1987) approach is systematic and straightforward but has been accused of as running into an "**ontological problem**" (Carstensen,

1995b): the observation that formal theories are only as good as their underlying ontologies allow them to be –which is evidently not sufficient for handling relevant linguistic phenomena. In fact, although there exist various proposals for ontologies by now (see Bateman & Farrar, 2004 for an overview), very few of them can be used for computational semantics: "we take many proposals that have been made in the ontological and spatial representation tradition to be problematic. In essence the problem can be described as attempting to do too much with the wrong tools" (Bateman and Farrar, 2004:96).

On the (cognitive) linguistic side, this critique has a well-known history in the works of, e.g., Bierwisch & Lang, 1989; Jackendoff, 1983; Lakoff, 1987; Langacker, 1987; Talmy, 2000, which emphasize the role of intermediate, often schematic cognitive representations, e.g. idealizations (Herskovits, 1986), idealized cognitive models (Lakoff, 1987) or object schemata (Lang, Carstensen, & Simmons, 1991). However, although proposals within this cognitive semantics are usually intuitively appealing and well motivated by linguistic data, their theories are often less clearly stated and mostly not well formalized. It can be reasonably assumed that this is in part due to the poor understanding of the cognitive machinery underlying the hypothesized cognitive constructs. Thus, it seems to be mandatory to unveil the cognitive mechanisms linking language and space and to clarify the properties of the corresponding "linking ontology".

The proposal that will be made here is to highlight the role of **selective attention** as an interface between language and visuo-spatial representations (cf. Carstensen, 1995b, 1998, 2001), a view that receives slow but gradually increasing support in the literature (Hogan et al., 1998; Hogan & Diederich, 2001; Knott, 2001; Talmy, 2000; Tomlin, 1997)[1]. For example, it has already been shown that attention is necessary for the apprehension of spatial relations (Logan & Sadler, 1996) and that spatial language can be grounded in attentional processes (Regier & Carlson, 2001). However, these approaches take a "quantitative" point of view and therefore disregard the qualitative aspects of attentional operations and their role for modelling spatial knowledge. In stark contrast to that, I will emphasize the value and role of *qualitative* aspects of attention for the construction of spatio-temporal ontologies.

The paper is based on three general assumptions:

A1 Aspects of the ception[2] of the World (and not just of the world itself) are important for semantic modelling.

A2 Selective attention plays a central role in ceiving the world.

A3 Represented patterns of attentional engagement provide the categories for a genuinely abstract upper structure of ontologies.

---

[1] See also Marchetti's proposal for a new research paradigm of 'Attentional Semantics' (Marchetti, 2006).

[2] "ception" is Talmys generalization of "perception" and "conception", cf. Talmy, 2000.

While A1 anchors this paper in the cognitive semantic paradigm, A2 marks its unique contribution to this field. It is only because of A3, however, that aspects of attention deserve closer attention from the viewpoint of knowledge representation and natural language processing.

The structure of this paper is as follows. I will first motivate the *specific* impact and central role of selective attention for the representation of spatial relations and for spatial semantics. I will then present *general* aspects of an attentionally defined upper ontology that covers both spatial and temporal categories.

2   Attention and spatial relations

2.1   Motivation

Lang, Carstensen, and Simmons (1991) presented an approach in which the (in)-compatibility of dimensional terms (i.e., dimensional adjectives like *long, short, wide, high* etc.) and object nouns (*tree, river, field* etc.) is explained and modelled by a systematic mapping of spatial semantics to structured conceptual object representations (so-called object schemata). A small implementation of the model showed that expressions like *long pole* and **long ball* were correctly judged as acceptable and inacceptable, respectively. When I started out to extend this approach to also encompass distance terms like *near* and *far* and measure terms like *10kms*, I was confronted with the task of modelling **spatial relations between objects** so that *high above*, *far above* and *deep below* would be accepted by the system while **deep above* or **high below* would be ruled out as unacceptable expressions.

At that time, little research had been done on that topic and it turned out that existing proposals for the semantics of locative prepositions (or other closed-class spatial morphemes) and distance phrases were not suited for handling these combinatorics, and had to be adapted (cf. Carstensen, 1992). Some linguistic data even pointed to fundamental problems of the state-of-the-art modelling of spatial relations, however. Standard practice was to specify the semantics of spatial prepositions as a two-place predicate ('LOC') which represents the fact that the place of the located object (referent, trajector) is contained in a certain region wrt. the reference object (relatum, landmark).

(1) a. *(nahe) bei der Kirche (∼ (near)by the church)*

  b. *(nahe) an der Kirche (∼ close to the church)*

  c. *knot in the rope*

  d. *3m (weit) hinter der Kirche (3ms [far] behind the church)*

  e. **3m (weit) an/bei der Kirche (*3ms [far] by/close to the church)*

  f. *?nahe hinter der Kirche (?near behind the church)*

Wrt. the data in (1), this raises the following questions (cf. also Carstensen, 1995b, 2003 for a more elaborate discussion):

Q1 What are the exact criteria for distinguishing *an* from *bei* in (1a) and (1b) (note that German allows to express two grades of proximity, with *an* implying a smaller distance between the objects)?

Q2 Which regions are to be assumed (especially if looking at cross-linguistic variation, cf. Bowerman, 1996) and *what*, for example, could be the region the knot in (1c) is located in?

Q3 Why can most prepositions be modified with a distance phrase as in (1d)[3] although it is not clear how this could be done compositionally with the LOC-predicate (and, accordingly, how the corresponding compatibility would be computed)?

Q4 Why, on the other hand, is it not possible to add a measure term as in (1e) *although this does not conflict extensionally with the proximity region* at hand (that is, the referent's place is still contained in the region)? On which level is the combination ruled out? Note that, contrary to what might be assumed, the aspect of proximity cannot be the cause of the incompatibility as distance gradation is possible with *nahe*.

Q5 Why are descriptions like (1f) at least questionable (speakers' judgements differ in this case)?

It should be clear that the data in (1) cannot be regarded as "linguistic idiosyncrasies" to be handled as exceptions but should rather be viewed as test cases for theories of language and space. To my knowledge, none of the available theories provides a satisfactory answer to these questions. Only recently some vector-based approaches have emerged which come close to the proposal made in this paper (cf. Zwarts & Winter, 2000; Regier & Carlson, 2001). One must do justice to the field, of course. There was (and still is) a lively discussion of which spatial calculus to use, how many relevant levels of representation to assume (cf. Aurnague & Vieu, 1993), how spatial language relates to spatial cognition (Landau & Jackendoff, 1993) etc.

There is agreement wrt. the observation that spatial linguistic phenomena cannot be explained on the geometric level alone. "Schema"-approaches emphasize the role of schematization, abstraction, and idealization in the ception of space (e.g. Herskovits, 1986; Lakoff, 1987). "Functional" approaches point out (to a differing extent) the influence of functional relations (containment, contact, covering, support) on referent and relatum (cf., e.g., the "functional geometry" proposed in Garrod et al., 1999). In Talmy (2000), which comprises the seminal work

---

[3]Note that it is quite normal in German to use the adjective *weit* (*far*) together with the measure term (which seems to be different in English). It simply identifies the dimension at hand (as in *the rope is 1m long*), as opposed to descriptions where a measure term is missing (*long rope, far behind*, cf. Bierwisch & Lang, 1989; Carstensen, 1998).

of Talmy on language and space, both of these aspects are combined. It is significant that large parts of this book as well as newer work (on the "fundamental attentional system in language", cf. Talmy, forthcoming) are devoted to the role of attention for cognitive semantics.

I have some problems with Talmy's intuition-style argumentation and his –in the meantime extensive– use of attention, however. Although the number of publications on attention in the (neuro-)cognitive disciplines has risen steadily in the past decades (cf. Cavanagh, 2004:24) I miss references to specific insights gained there in Talmy's work. Since attention is a complex and non-unitary phenomenon (cf. Carstensen, 2001) and since "[e]veryone knows what attention is" (a much-quoted wording of William James, cit. ibid.), chances are that the term is over-used, leads to misunderstanding and/or looses its explanatory power. Besides that, although Talmy introduces a remarkable number of explanatory constructs (e.g., conceptual categories like 'plexity' and 'state of boundedness' and schematic systems like 'distribution of attention', cf. Talmy, 2000:76f.), he often leaves open their origin and role in the cognitive system (the same critique applies to other cognitive linguists as well).

My aims are much more humble: I want to provide answers for the questions raised above (and the specific data in (1)) which are based on specific explanatory cognitive constructs (in the same line as has been done in Landau and Jackendoff (1993) wrt. the impact of the 'what'/'where'-distinction). For this we need to take a closer look at a much more restricted sense/function of attention and the role it plays in how we ceive spatial relations.

2.2   Attention-based spatial relations

Based on evidence from language learning and development, Olson and Bialystok (1983) argued for the importance of distinguishing **implicit and explicit (spatial) relations** (a shorthand for: implicitly represented and explicitly represented (spatial) relations). For example, while it is easy even for very young children to correctly categorize certain things as 'lollipop' –which includes identifying the relation between a 'round thing' and a 'sticky thing'– they may not yet have this relation explicitly available for thinking and speaking. This distinction is supported by recent work in cognitive science. For example, Kosslyn writes with respect to the 'what'/'where' dichotomy: "Although the [what] system cannot represent explicit spatial relations, it must be able to represent implicit spatial relations; such relations are inherent in any pattern" (Kosslyn, 1994:421).

What are explicit relations and how do they get established? Consider the two objects in figure 1a. Obviously, there is a spatial relation between them which is also present in figure 1b. However, as is demonstrated clearly, mere presence of the –implicit– relation does not lead to its availability to the observer (or better, to the cognitive processes underlying the recognition of a certain spatial rela-
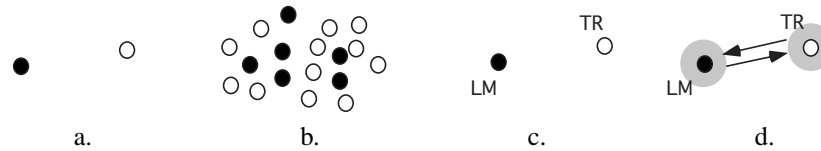
Figure 1.: Microperspectivization: making implicit relations explicit

tion). This becomes evident most dramatically in cases of so-called object-based neglect, where patients are not able to see even the spatial relation in figure 1a (although their visual areas are intact). They would report only the existence of one object as their attention is "stuck" to it (Behrmann & Tipper, 1994).

In general, it has been shown (cf. Theeuwes, 1993) that there is a serial stage at which objects in the visuo-spatial medium (the "visual buffer" of Kosslyn, 1994) are attentively selected one after another for further processing in the what-system. Exactly what gets attended at a certain point of time is jointly determined by the properties of the given and preprocessed entities of the "display" ($\rightarrow$ bottom up aspect) and stored patterns of attentional behaviour (attentional templates, $\rightarrow$ top-down aspect). As to the bottom-up aspect, it is controlled by two main principles: Differences in the display –that is, places of **change** between quale regions– attract the attentional window and thus determine salient entities (bounded regions or boundaries of regions) to be further processed, and inhibition of visited places/objects prevents immediate return to those entities.

Let me point out three important facts that follow from this condensed presentation of spatial selective attention. First, **boundaries** in the representational medium are essential for the working of (focused) **selective attention** and are definitional for something to be an object. This will become important later.

Second, it is **explicit** spatial relations which we need to look at and work with in theories of language and space, and it is the **shifts of attention** which establish these relations ("Computing relations requires directing attention", Logan, 1995:163). Formal approaches might therefore consider whether they model implicit or explicit relations, and approaches within cognitive linguistics might question, e.g., whether (some of) the information in an image schema reflects explicitly or implicitly represented spatial information. This also sheds new light on the discussion of the role of functional aspects in spatial relations: In some sense, all explicit relations are functional as they are not defined in geometric terms and as they can be influenced by various top-down/bottom-up factors, e.g., knowledge about typical object constellations via attentional templates.

Third, additional structural information is induced by attention-based relations. With the distinction between referent/trajector (TR) and relatum/landmark (LM) there is a well-known important asymmetry in linguistic spatial relations, often depicted as in figure 1c. As is shown in figure 1d, there are now two possibilities

of attention shifts between the objects of an implicit relation, and each of them imposes a certain perspective on it. In order to avoid confusion with the usual interpretation of the term 'perspective', I will call these shifts "**microperspectives**" to highlight their asymmetry. It is assumed that these microperspectives constitute the core of explicit spatial relations, which are verbalized, for example, as prepositions.

Speaking of a "core" presupposes that there is more to prepositional semantics than merely an attentional change. In fact, much recent research is devoted to the role of **frames of reference** (cf. Levinson, 1996) in characterizing spatial relations. However, it is important to point out that the implicit/explicit-dichotomy also applies here: As opposed to coordinate systems, it is the axes of the reference frames which are relevant in a qualitative attentional treatment of spatial relations. Axes allow discrete categorizations of attentional vectors (where there would otherwise be continuous differences) and therefore pave the way for the use of categorical linguistic expressions.

To sum up, my proposal is to take a more general, attention-based viewpoint on what a spatial relation available for the linguistic system is. This can be summarized as follows:

- The core of a spatial relation as an *explicit* relation is an attentional change between spatial objects
- the attentional change occurs in the spatial domain, i.e., attention operates on some spatial representation (which seems trivial, but note the possibility for abstraction)
- the attentional change is categorized with respect to a reference axis, if available
- the explicit relations are directed (i.e., are *microperspectives* of the implicit relation)

In other words, this proposal "aims at finding the attentional instructions conveyed by the meanings of words" (Marchetti, 2006:163), both for language interpretation and generation. The next subsection elaborates on this point.

2.3   Attentional parameters and spatial semantics

There is a set of prominent parameters according to which attention-based spatial relations can be conceptually categorized. The inter-individual similarity of this cognitive processing and representation is the assumed reason for the similarities in the linguistic inventories of different languages, while top-down influenced (e.g., culture-specific) variations in viewing and (micro)perspectivizing the same scene can explain the cross-linguistic differences in the meaning of spatial terms.[4]

---

[4]Note that these variations in (micro)perspective also apply to a single individual and a given scene so that an implicit relation can be described, for example, as *behind/near/3ms away from the church* without contradiction.

As I have worked out in detail elsewhere (Carstensen, n.d.), this also implies that understanding a locative description should be given a procedural account: spatial (relation) terms provide qualitative information, which by being instantiated in a hearer's working memory *retraces* the ception of the speaker, adjusting other aspects like scale and granularity on the fly (I have called this the 'localization as mental presentation' (LaMP)-view, which is a specific, attentional counterpart of *conceptualization*).

Some of the parameters discussed in the following will probably appear as features of a domain-specific spatial ontology while others will be part of the more general upper-level ontology presented later.

Type of attended object(s)

Abundant evidence suggests that on the perceptual level, attention is attracted by discontinuities of the so-called "map of locations" (Treisman, 1988). Qualitatively, at least[5] three types of attended objects (metaphorically, entities lying in the spotlight of attention) can be distinguished: whole visuo-spatial entities ("blobs"), boundaries between regions, or groups of blobs.

As to question Q1, there is a clear correspondence of the blob/boundary-distinction and the notoriously difficult differentiation between German *an* and *bei*: while *bei* expresses mere proximity between objects, *an* generally indicates a closer relation and is obligatory in cases where the landmark is large (living *an der See* (*by the sea*)) or, most importantly, where it is a noun denoting a boundary (*am Rand, an der Spitze/Ecke/Seite* (*on the border/side, at the peak/corner*)). Thus, *an* requires the visuo-spatial object that is designated by the landmark to be a boundary (while *bei* requires a blob). According to the LaMP view this means that the scale of the imagined *an*-constellation is bigger, which corresponds to a smaller distance between TR and LM in the world.

between is a preposition that evidently selects for a group-of-blobs landmark.

Type of attentional change

There are two general types of operations transforming one visuo-spatial attentional state into another: Attention can either be shifted, or it can be zoomed. While most locative prepositions express attention shifts, *in, between* and *among* express zooming. The answer to question Q2 is provided by the observation that attentional semantics can dispense with regions as elements of spatial relations and can account for linguistic data which are problematic for standard region based theories. For example, cases like *knot in the rope*, projections like *face in the mirror*, or abstract relationships *bug in the algorithm* are –without having to resort to functional containment– elegantly covered by assuming a zoom-in-type relation for *in*.

---

[5]Finer distinctions will be made below.

The answer to question Q3 is that the explicit relation provides a vector whose length extent can be measured (leading to transparent semantic compositionality of location and distance expressions).

Reference polarity.

With the two possible microperspectives of an implicit relation, there are two assignments of these roles to the source and goal object of a microperspective (see figure 1d). I will represent this difference in the following by a feature **reference polarity** (αrefpol) specified as follows:

+refpol: LM is source of a microperspective

-refpol: LM is goal of a microperspective

Reference polarity necessarily results in a subclassification of explicit spatial relations. Thus, it can be predicted that there may be two classes of spatial relation expressions based on this subclassification.

Actually, there is the well-known distinction between *projective* (*above, behind, left of* etc.) and *topological* spatial prepositions (*in*[6]*, at, on, by* etc.). It can therefore be argued that projective prepositions express positive reference polar (+refpol) relations and topological prepositions express negative reference polar (-refpol) relations. This explains the source orientation of these expressions (e.g., *away from*) as opposed to the goal orientation of topological prepositions (*close to, next to*).

By referring to reference polarity it can also be stated that only +refpol-relations are compatible with measure phrases (see question Q4) and that the polarity of the distance adjectives must match the reference polarity of the preposition (which is a peculiarity of these adjectives, cf. Carstensen, 2003) (question Q5).[7]

Reference system and axis

It has already been pointed out above that the axes of reference systems (as modular entities) play an important role in the categorization of spatial relations (as opposed to wholistic reference frames, cf. Lang, Carstensen, & Simmons, 1991). Spatial prepositions may be sensitive to *special* axes in *special* reference systems. For example, the differentiation between *above/below* and *over/under* can be attributed to their respective reference to an (inherently) spatial vertical axis (derived from the axis of gravitation) on the one hand, and a vertical axis of a visuo-spatial

---

[6]Note that *in* is classified differently according to the present theory as it is based on zooming and can furthermore be modified by *deep*.

[7]Actually, this is much more intricate. A complete explanation would require looking at measurement, comparison and the semantics of distance adjectives and gradation. See Bierwisch and Lang (1989); Carstensen (1998) for a more elaborate discussion.

representation on the other hand. If one refrains from using functional explanations (which I in fact do) like "covering" senses of *over/under*, one has to come up with a constructive and modular treatment of axes (e.g., axes defined via orthogonality to a boundary or surface, or axes induced by motion as for *behind the flying bullet*).

According to the present approach, most projective prepositions (i.e., +refpol relations) are somehow associated with reference systems. This is compatible with vector-based approaches, where the reference frame is imposed on the reference object (cf. Logan, 1995; Zwarts & Winter, 2000). It is assumed here, however (contrary to, e.g., Carlson & Deman, 2004), that this *only* holds for these prepositions (even with the exception of *away from*) and that reference systems are not involved in the ception of -refpol relations. Note that recent investigations grant the located object a more prominent role in the spatial relation apprehension process (cf. Burigo & Coventry, 2004), which is a step towards the representation and processing of -refpol relations as proposed here.

Microperspective's relation to axis

If –as has been argued– microperspectives are categorically related to spatial axes, then this relationship can be further specified: the *direction* of a microperspective may either be congruent or incongruent with the direction of an axis. It *must* be specified for the distinction between the so-called "mirror view" and "tandem view", i.e. relative reference frames with different assignments of the front/back-asymmetry (cf. Levinsons differentiation between *rotating* and *translating* coordinates, Levinson, 1996:143), and for corresponding assignments in the intrinsic reference frame (a car's front is the side towards the direction of its inherent OBServer axis, a desk's front is the one opposed to it, cf. Lang, Carstensen, & Simmons, 1991). This can be represented by simply coding whether (the direction of) the +refpol-microperspective, of, say, *behind* (or any corresponding term in another language), is *congruent* with (the direction of) the OBS axis or not.

Domain and level(s) of representation

The representation of explicit relations concerns different levels of cognitive representation and processing in some domain. Accordingly, both visuo-spatial and conceptual aspects are involved in the representation of spatial microperspectives. On the conceptual level, a spatial relation contains the categorical information about an attention change. On the visuo-spatial level, it codes the corresponding scan path (for distance measurement).

The next section shows that domain membership can be regarded as a feature in an attentional upper-level ontology.

3 Attention in ontologies

3.1 Motivating attentional criteria for ontologies

Having discussed the specific impact of attention for spatial representation and processing, the question arises whether attention also plays a role in other domains, and if so, in which sense. Part of the motivation for this comes from the author's work on the attentional semantics of *to follow* (cf. Carstensen, 1995a) and other work in cognitive linguistics (e.g., the work on metaphors beginning with Lakoff & Johnson, 1980), part from general interest in upper-level structures since their effective use in the Cyc-project (cf. Lenat & Guha, 1990). The main hypothesis is that attentional aspects are per se domain independent and therefore constitute part of a domain independent ontology. A secondary hypothesis would be that due to the abstract nature of attentional aspects, an attentional upper-level ontology can contribute to theories of metaphor in the sense that much of the burden of a structure mapping from source to target domain (cf., e.g. Gentner, 2001) is relieved by the possibility of both domains *sharing* some abstract structure. However, as this topic is clearly outside the scope of this paper, it will not be pursued further.

It is not quite easy to find a starting point for the discussion of attentional ontologies. I have decided to use DOLCE ("a Descriptive Ontology for Linguistic and Cognitive Engineering") which is described as having "a clear cognitive bias, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense. " (Masolo et al., 2003:13). In order to motivate my proposal, I want to begin with two important basic points of disagreement[8].

First, DOLCE makes a fundamental distinction between endurants and perdurants (simplified: between temporo-spatial entities appearing as non-temporal and temporal expressions in language, respectively). While this is standard procedure in the knowledge representation and natural language processing communities, it precludes having an abstract upper structure which contains aspects common to both domains. Any approach providing such a structure could be better suited for handling the –according to Lakoff and Johnson (1980)–omnipresent metaphoricity in language and cognition.

Second, DOLCE's *Amount of matter*, which is supposed to represent "Stuff" like "gold", "iron", "wood" etc., reflects part of the mass noun–count noun distinction in language (one can say *much gold* and *a few pencils*, but not **a few gold* or **much pencils*). With the name of the category, however, as well as with the examples they provide ("some air, some gold, some cement", p.10, and the "piece of clay" they use extensively), they refer to bounded entities. This conflicts with uses of mass nouns which refer to unbounded stuff itself, e.g. *the color*

---

[8] while I am fully aware of the fact that this might not do justice to the complexity of the approach.

of gold is..., *there is white smoke all around me*. Thus they incorrectly conflate unbounded stuff with bounded objects. This points to a general problem of extensional, "objectivist" theories: even in this vast universe, every stuff must be ultimately regarded as bounded (this is why in the Cyc ontology, the stuff of gold and the collection of all existing gold pieces were regarded as co-extensional, with their types being a subtype of each other–a very unpleasant result wrt. ontology design).

The solution to these problems proposed here is taking a radically cognitive stance in the design of ontologies. It reflects the insights gained about implicit and explicit representations and the role of attention. It also takes into account the fact that our ception of the world is piecemeal and partial, and that it it these view-specific, selective chunks of information that are expressed by natural language. Accordingly, the proposal is essentially based on **Boundedness**, both as the principle of cognitive operation just described and as the dominant factor in attentional selection (introduced in section 2.2).

If we acknowledge that the content of limited working memories –implicitly representing aspects of the outer world– is operated upon by selective attention, there are two options. Either there are boundaries or salient bounded entities available (as computed on the pre-attentional stage) which are then selected for categorization by focused attention (leading to **Object**-like categories), or the representational medium is devoid of salient entities, so that categorization is dependent on the qualities present (leading to **Stuff**-like categories).

Let us call such a working memory containing information of some domain (e.g., the spatial domain) and some level (visual, spatial, conceptual etc.) a **Frame**. We can now carve the structure of the (ception of the) common-sense world by using the **basic necessary** distinctions that need to be made for sorting possible frame instantiations, given the theoretical and empirical constraints about the working of selective attention. This gives rise to the attentional upper structure depicted in figure 2, which is explained in the following subsections.

3.2   The attentional ontology

We will restrict our interest to the explicit representation of spatio-temporal entities as defined by attentional criteria (**ADE**s), leaving other important aspects like *qualities*, *roles*, *regions*, *dimensions* etc. undiscussed. Below ADE, we then have a distinction between entities defined wrt. to a single frame (**SFE**) and those defined wrt. more than one frame (**MFE**). Superficially, this corresponds to dividing things into those definable in an instance of time and those requiring a time span (and **change**) for definition (that is, the endurants and perdurants of DOLCE mentioned before). Actually, there is quite a difference to these dichotomies as they are not defined wrt. objectively given instants and changes, but wrt. attentional ones. For example, spatial relations will be essentially micro-temporal
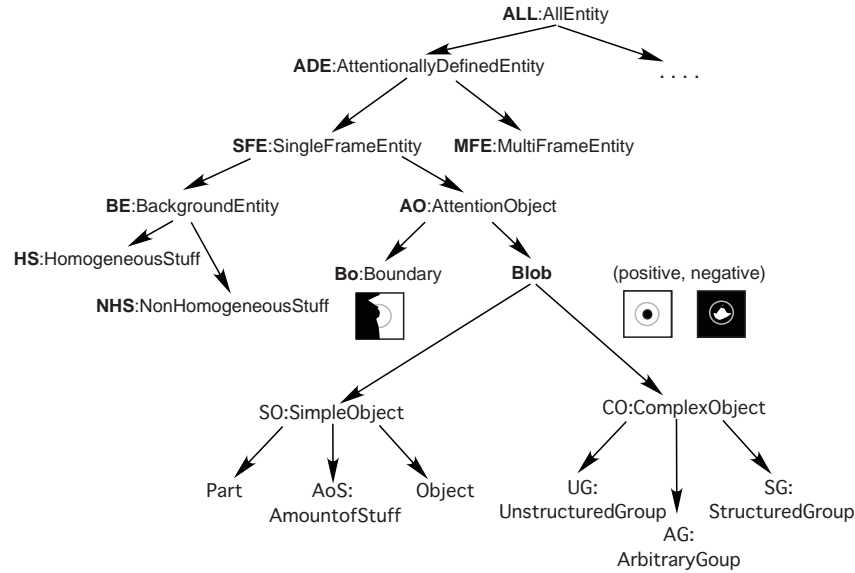
ALL:AllEntity

ADE:AttentionallyDefinedEntity   . . . .

SFE:SingleFrameEntity   **MFE**:MultiFrameEntity

BE:BackgroundEntity   **AO**:AttentionObject

HS:HomogeneousStuff   **Bo**:Boundary   **Blob**   (positive, negative)

NHS:NonHomogeneousStuff

SO:SimpleObject   CO:ComplexObject

Part   AoS:AmountofStuff   Object   UG:UnstructuredGroup   SG:StructuredGroup

AG:ArbitraryGoup

Figure 2.: An attentional upper ontology I

(MFE), and intrinsically temporal entities may be categorized as objects (and classified below SFE).

As shown in figure 2, the primary dichotomy under SFE is the one between entitites that are characterized by the background of the frame (**BE**) vs. those that are characterized as attractors of attention defining them as objects (**AO**). Note that this distinction abstracts over different domains, e.g. the spatial and the temporal domain. As a result, BE is a supertype of $SpatialStuff$ and $TemporalStuff$ ("sleeping", "standing"), yet with an intervening level consisting of **HS** ("gold", "standing on ones left leg") and **NHS** ("garbage", "playing"). The next differentiation is between boundaries (**Bo**) and bounded regions in the frame (**Blob**s): attention is either attracted to boundaries of objects or to whole entities. Below $Blob$ one can find typical, albeit domain independent, object categories, which are discussed in the next section.

MFEs (see figure 3) can be divided into entities characterized by a single change of frame (**SFC**) corresponding to a relation between two SFEs, and into entities characterized by a multiple change of frame (**MFC**) corresponding to processes of change. In each case, the change in background (**SBC**, e.g., *turning off the light*) and the attentional change between AOs constitute the next lower level of distinction.

SACs correspond to changes in focused selective attention and represent relations between entities. SACs therefore feature a source and a goal entity, both
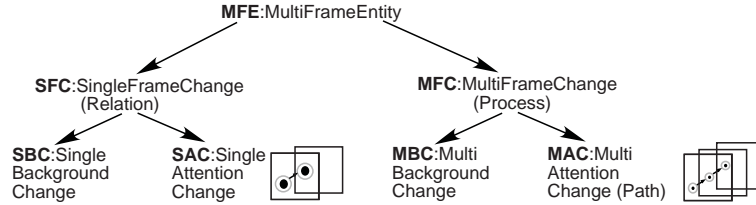
Figure 3.: An attentional upper ontology II

of which are AOs. In the spatial domain, they are typically expressed by prepositions and adverbs (i.e., *in, at, on, behind, in front of, above, below* etc.) while in the temporal domain they can be expressed by temporal prepositions and adverbs (e.g., *the day after tomorrow*, *before sunrise*) or conjunctions (e.g., *after I had written the paper,...*, *when she entered the room*, *I heard that before*). As has been shown in Carstensen (1998), gradation of dimensional or distance extent (*very long* etc.) and the comparison of different extents (*longer than*, *less than 3m wider than*, *nearly as far as* etc.) are another domain of SAC, which may also include quantification phenomena (*many/more people*, *nearly all of them* etc.).

MFC categorizes multiple changes between entities and therefore represents (iterative) processes. In the case of **MBC**, this may be processes involving gradually changing BEs (*getting dark*), while in the case of **MAC**s, the sequences of AOs are typically identified as *paths*.

Note that the MFC categories reflect primarily subjective, cognitive temporal changes that may or may not represent objective processes. For example, a MAC may represent paths of moving objects (*a man crossing the street*), but it may likewise represent abstract paths (*a power line crossing the street*), which corresponds to the differentiation between *tracking* and *scanning*. It is a result of attention research that we not only constantly scan (implicit representations of) objects in the world along their axial extensions or along their shape contours (boundaries), but that we also represent these acts of scanning as so-called *attentional templates*. Thus we learn aspects of shape not only as implicitly represented patterns, but also as explicitly represented attention paths. From this point of view, it comes as little surprise that we instantiate learned scanning patterns in form of corresponding tracking patterns which are expressed with reference to the originating shapes (*walk **across** the street*, *run a**round** the house* etc.).

3.3 Comparison with DOLCE

As has been demanded above, (nonbounded) stuff as BE is clearly distinguished from (bounded) amount of stuff (**AoS**) in the attentional ontology. There are a few other aspects in which this ontology differs from DOLCE.

For example, DOLCE distinguishes between $States$ and $Processes$ (corresponding to temporal stuff) using the homeomericity criterion[9] but does not distinguish further within $Amount\ of\ matter$, e.g., between "gold" and "garbage".

Bo corresponds in part to DOLCE's $Feature$ representing "edges", "borders", "corners" and the like. However, while "bumps" and "dents" are categorized as $Feature$s in Masolo et al. (2003), they are $Blob$s according to the present theory. The reason for this is that the attentional ontology is based on subjective criteria: "bumps" and "dents" are bounded objects and therefore $Blob$s, although they are objectively features of some surface (similar to "mountains" and "valleys" which are features of the earth's surface): linguistically, it is *the corner/edge of sth.* and not *the corner/edge in sth.* while it is *the bump/dent in sth.* as opposed to *the bump/dent of sth.*

As is depicted with the idealized iconic representations[10] in figure 2, whole entities can be either positive (the blobs are foregrounded entities) or negative (in essence, the blobs are bounded background). That is, "holes", "fissures", "valleys" (and "dents"!) etc. are negative blobs according to the present theory, and not $Feature$s, and are to be distinguished from $Boundaries$.

Below SO there is a distinction between three subtypes of simple objects. Ordinary **Object**s correspond to $PhysicalObject$s in other ontologies but are not specified wrt. a domain. **Part**s represent (arbitrary, i.e., *Fiat*) parts of an $Object$ (e.g., one of the two halves of a broken pencil, a piece of a cake) which are not themselves objects of some identifiable type but are dependent on some other object. AoS represent objects characterized by their bounding of stuff (that is, portions like "drop of water", "piece of wood", "handful of jewellery").

As to **CO** (in which case focused attention is attracted to more than one entity at the same time), there are subtypes of this category corresponding to the subtypes of SO in the attentional ontology. One can distinguish unstructured groups of blobs (**UG**s like a "flock of sheep" or a "swarm of bees") from structured groups (**SG**s like "teams", "choirs") and arbitrary groups (**AG**) which probably corresponds to $ArbitrarySum$s in DOLCE. Again, these properties also have temporal subtypes, e.g., the "summer music events in Hamburg" (UG) or the "winter olympics in Turin" (SG).

The upper-level attentional ontology takes a different stance on representing temporal/situational categories. As has been explained, static and dynamic $States$ are BEs while iterative $Process$es are MFCs. DOLCEs $Achievement$ denotes a change of state and is therefore a SFC. Its $Accomplishment$, however, cannot be easily integrated into the ontology because it is a complex category consisting of both an $Activity$ and an $Achievement$ (cf. Steedman, 2005 on temporal categories). I will therefore not discuss it further in this paper.

---

[9]Simplified: every part of an entity must be of the same type as the type of the entity.

[10]Note that one must not take these visuo-spatial items literally, as there are abstract entities that cannot actually be depicted (although they are subject to the same attentional constraints).

Interestingly, DOLCE lacks Steedman's $Point$, a category representing simple situational objects (corresponding to a $Perdurant Blob$: an atomic event which draws attention relative to its background of non-happening, e.g., *a flash of lightning*), and bounded activities (corresponding to $Extended\ Perdurant Blob$s: so-called $pofectives$, e.g., *run for a while*). In contrast to that, the domain-level attentional ontology will contain (as can be predicted from the foregoing explanation) even more temporal categories, for example $Negative Perdurant Blob$s representing "interruptions" and "pauses", or $Perdurant Boundaries$ representing the "beginning/end" of an event.

4   Conclusion

In this paper, I have tried to give tentative answers for the questions posed at the beginning, regarding spatial relation expressions, spatial relations, and their relationship. According to the attentional account, spatial semantics has to be based on a theory of explicit(ly represented) spatial relations (as opposed to implicit(ly represented) ones). Attentionally defined spatial relations are inherently different from geometrically defined ones and provide –with their features and their categorization with respect to salient axes, if +refpol– a wide range of explanation for linguistic phenomena. Therefore, they are an interesting alternative to functional explanations.

It has been shown that a systematic representation of some aspects of attentional engagement leads to an attentional ontology which can be viewed as the upper structure of spatio-temporal ontologies (with the potential to cover other domains, too). Spatial and temporal categories are seen here not as being differentiated by an early branching in the ontology but by their inherent domain-specific placement below domain-independent nodes of the upper structure.

To some reader, this outline of an attentional upper structure may occur simple or even simplistic, as it seems to ignore the intricate design choices, the complex ontological distinctions and their axiomatization, the elaborate methodology and the rigid formalization of up-to-date ontology theorization and engineering, which instead should have been considered. Simple? Yes (please mind my quite restricted research questions). Simplistic? The computational linguistic ideal HAL, capable of domain-crossing use of words, would perhaps answer as follows:

> "I'm sorry, Dave, I couldn't do that. It is my deep belief that, at the end, I would have been behind schedule after 19 pages. Maybe the time will come when nearly all of these aspects will be properly placed in my line of argument. I am highly looking forward to reaching this goal, Dave."

References

Aurnague, M., & Vieu, L. (1993). A three-level approach to the semantics of space. In C. Zelinsky-Wibbelt (Ed.), *The semantics of prepositions – from mental processing to natural language processing* (pp. 395–439). Berlin: Mouton de Gruyter.

Bateman, J., & Farrar, S. (2004). *Spatial ontology baseline* (SFB/TR8 internal report No. I1-[OntoSpace]: D2). Collaborative Research Center for Spatial Cognition, University of Bremen, Germany.

Behrmann, M., & Tipper, S. P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance xv : Conscious and nonconscious information processing* (pp. 109–169). MIT Press.

Bierwisch, M., & Lang, E. (Eds.). (1989). *Dimensional adjectives*. Springer Verlag.

Bowerman, M. (1996). Learning how to structure space for language: a crosslinguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 385–436). MIT Press.

Burigo, M., & Coventry, K. (2004, 5 – 7 August). Spatial language and reference frame assignment: the role of the located object. In *Proceeding of the 26th annual meeting of the cognitive science society.* Chicago.

Carlson, L. A., & Deman, S. R. V. (2004). The space in spatial language. *Journal of Memory and Language*, *51*, 418-436.

Carstensen, K.-U. (n.d.). *Location and distance in language: An attention-based approach.* (Paper presented at the 3rd ANNUAL LANGUAGE & SPACE WORKSHOP 'Resolution and Granularity in Language and Space'. ZIF, Univ. Bielefeld. 8./9.7.2002)

Carstensen, K.-U. (1992). Approaching the semantics of distance expressions: Problems and perspectives. In P. Bosch & P. Gerstl (Eds.), *Discourse and lexical meaning* (pp. 49–59).

Carstensen, K.-U. (1995a). Focused spatial attention and spatial semantics: The case of 'to follow'. In *Proceedings of the TSM'95* (pp. B 15–27). Toulouse.

Carstensen, K.-U. (1995b). Semantic and conceptual aspects of local expressions: Critical remarks on the 'state of the art'. In M. Bierwisch & P. Bosch (Eds.), *Semantic and conceptual knowledge* (pp. 117–126).

Carstensen, K.-U. (1998). A cognitive science approach to the semantics of spatial gradation. In P. Ludewig & B. Geurts (Eds.), *Lexikalische Semantik aus kognitiver Sicht – Perspektiven im Spannungsfeld linguistischer und psychologischer Modellierungen* (pp. 73–94). Gunter Narr Verlag (Tübingen).

Carstensen, K.-U. (2001). *Sprache, Raum und Aufmerksamkeit [Language, space, and attention].* Niemeyer Verlag.

Carstensen, K.-U. (2003). Asymmetries in spatial semantics. In C. Maienborn (Ed.), *(a)symmetries. papers in honor of ewald lang.* Stauffenburg Verlag.

Carstensen, K.-U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., & Langer, H. (Eds.). (2004). *Computerlinguistik und Sprachtechnologie: Eine Einführung* (2nd ed.). Spektrum Akademischer Verlag.

Cavanagh, P. (2004). Attention routines and the architecture of selection. In M. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 13–28). New York: Guilford Press.

Cohn, A. G., Bennett, B., Gooday, J. M., & Gotts, N. (1997). RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, *1*, 275–316.

Garrod, S., Ferrier, G., & Campbell, S. (1999). In and on: investigating the functional geometry of spatial prepositions. *Cognition*, *72*, 167-189.

Gentner, D. (2001). Spatial metaphors in temporal raesoning. In M. Gattis (Ed.), *Spatial schemas in abstract thought* (pp. 203–222). Cambridge, MA: MIT Press.

Herskovits, A. (1986). *Language and spatial cognition. an interdisciplinary study of the prepositions in english*. Cambridge Univ. Press.

Hogan, J. M., & Diederich, J. (2001). Encoding of spatial relations semantics through selective attention. In *Proceedings of the 4th australasian natural language processing workshop*. Sydney.

Hogan, J. M., Diederich, J., & Finn, G. D. (1998). Selective attention and the acquisition of spatial semantics. In J. Burstein & C. Leacock (Eds.), *Proceedings of the joint conference on new methods in language processing and computational language learning* (pp. 235–244). Somerset, New Jersey: Association for Computational Linguistics.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, Mass.: MIT Press.

Knott, A. (2001). Argument linking and spatial cognition. In *Proceedings of sempro 2001: Cognitively plausible models of semantic processing*. University of Edinburgh.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. MIT Press, Cambridge, MA.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Univ. of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Univ. of Chicago Press.

Landau, B., & Jackendoff, R. (1993). "what" and "where" in spatial language and spatial cognition. *Behavioural and Brain Sciences*, *16*, 217–265.

Lang, E., Carstensen, K.-U., & Simmons, G. (1991). *Modelling spatial knowledge on a linguistic basis: Theory - prototype - integration*. Springer. (Lecture Notes in Artificial Intelligence No. 481)

Langacker, R. (1987). *Foundations of cognitive grammar I: Theoretical prerequisites*. Stanford University Press.

Lenat, D., & Guha, R. V. (1990). *Building large knowledge-based systems: Representation and inference in the cyc project*. Addison-Wesley.

Levinson, S. C. (1996). Frames of reference and molyneuxs question: Crosslin-

guistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 109–169). MIT Press.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*, 103–174.

Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 493–529). MIT Press.

Marchetti, G. (2006). A presentation of attentional semantics. *Cognitive Processing*, *7*, 163-194.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., & Oltramari, A. (2003). *Wonderweb deliverable d18: Ontology library* (Deliverable No. D18). ISTC-CNR, Italy.

Olson, D. R., & Bialystok, E. (1983). *Spatial cognition*. Lawrence Erlbaum Associates.

Regier, T., & Carlson, L. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*, 273–298.

Smith, B. (1995). Formal ontology, common sense and cognitive science. *International Journal of Human-Computer Studies*, *43*, 641–667.

Steedman, M. (2005). *The productions of time* (Draft tutorial notes about temporal semantics No. 5.0). Univ. of Edinburgh.

Talmy, L. (2000). *Toward a cognitive semantics, vol. i*. Cambridge et al.: The MIT Press.

Talmy, L. (forthcoming). Attention phenomena. In D. Geeraerts & H. Cuyckens (Eds.), *Handbook of cognitive linguistics.* Oxford University Press.

Theeuwes, J. (1993). Visual selective attention: A theoretical analysis. *Acta Psychologica*, *83*, 93–154.

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge University Press.

Treisman, A. (1988). Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, *40*, 201–237.

Zwarts, J., & Winter, Y. (2000). Vector space semantics: a model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information*, *9*, 169-211.