KR-MED 2006 "Biomedical Ontology in Action"
November 8, 2006, Baltimore, Maryland, USA

"*Lmo-2* interacts with *Elf-2*" On the Meaning of Common Statements in Biomedical Literature

Stefan Schulz[1], **Ludger Jansen**[2]
[1] Department of Medical Informatics, Freiburg University Hospital, Germany
[2] Department of Philosophy, University of Rostock, Germany

Abstract

Statements about the behavior of biological entities, e.g. about the interaction between two proteins, abound in the literature on molecular biology and are increasingly becoming the targets of information extraction and text mining techniques. We show that an accurate analysis of the semantics of such statements reveals a number of ambiguities that is necessary to take into account in the practice of biomedical ontology engineering. Several concurring formalizations are proposed. Emphasis is laid on the discussion of biological dispositions.

Introduction

The study of so-called protein-protein interactions is essential for a better understanding of biological processes, from replication and expression of genes to the morphogenesis of organisms. Statements such as "*Lmo-2* interacts with *Elf-2*" – with *Lmo-2* and *Elf-2* being proteins – occur in biomedical literature abstracts with a very high frequency and represent, in many cases, the core message of a scientific paper.

There are several kinds of biomolecular interactions, e.g. binding, inhibition, activation, and transport. They all involve (1) at least two biomolecules and (2) the spatial vicinity of these, which leads to (3) a causal influence that they exert on each other.

Text mining, i.e the process of extracting structured knowledge from unstructured text, primarily targets statements such as these, and there is a major interest by the text mining community in obtaining ontological support for their information and knowledge extraction activities. This is one of the reasons why so-called bio-ontologies have emerged, and the use of formal ontological criteria has been repeatedly advocated in order to facilitate the process of automatic processing of domain information.

Much work in this area has already been done in the form of ontological investigations on material continuants, such as organs, cells, molecules [1, 2, 3]. However, there has been much less emphasis on biologically relevant functions and processes. Furthermore, biomedical ontology engineering has been mainly committed to traditions of semantic networks, lexical semantics, and cognitive science. Thus rather than being construed as describing real world entities by means of logical expressions, ontology has been understood as relating concepts (i.e. representations of word meanings) by means of conceptual relations. On this assumption, "*Lmo-2* interacts with *Elf-2*" would simply signify that there is some plausible linkage between the concepts (conceived of as mental representations) *"Interaction"*, *"Lmo-2"*, and *"Elf-2"*. As much as this approach might be adequate for communicating knowledge about the world by means of natural language or some kind of abstraction (e.g. semantic networks), it fails where exact statements and reasoning about biological entities such as molecules, functions, or pathways are required.

Interestingly enough, scientists and other human agents are perfectly able to communicate by means of such sentences, although there is only a vague consensus about the referents (the entities in the world) which are denoted by these linguistic expressions. Because of the ambiguities of natural language, a natural language statement like "*Lmo-2* interacts with *Elf-2*" may have more than one possible interpretation and thus more than one formalization in, say, first order predicate logic. In formally representing the meaning of such statements, we have thus to make explicit the ontological assumptions intended by the speakers or authors of that sentence.

In this paper we will demonstrate that even the formalization of an apparently simple but prototypical statement about protein interaction like "*Lmo-2* interacts with *Elf-2*" can yield totally different ontological assumptions.

Basic Ontological Assumptions

It is widely recognized that the construction of biomedical ontologies should obey strict logical and ontological criteria. To this end, several top-level ontologies have been devised, such as DOLCE [4], BFO [5], and GOL [6]. These ontologies mainly coincide in their fundamental division between continuants (endurants, e.g. material objects) and occurrents (perdurants, e.g. events, processes). The distinction is that occurrents have temporal parts (they are never fully present at a given time) and they are existentially dependent on continuants. Continuants are split into independent and dependent ones. Examples of independent continuants are material objects and spaces. Dependent continuants, on the other hand, are entities which inhere in something and are thus ontologically dependent on their bearer. Examples of dependent continuants are masses, colors, and tendencies: A particular mass may inhere in a particular molecule, a particular color may inhere in a particular flower. The tendency to divide may inhere in a cell and the tendency to relieve headache may inhere in an aspirin tablet. Tendencies are related to occurrents by the relation of *realization*. They are special kinds of dependent entities, in that they need not be realized in order to exist. There are cells which never divide, and aspirin tablets that never relieve a headache.

Our ontological framework for describing molecular interaction patterns includes entities of all these kinds. For instance, a protein molecule, which is a material continuant, has a disposition to perform a certain function, e.g. binding, which is a dependent continuant, and an actual realization of this disposition, *viz.* the process of binding a protein molecule, which is an occurrent.

Aware of the need to comply with existing standards for ontologies, especially in the light of the Semantic Web and the various specifications of Description Logic (DL) [7], we keep our logic simple. So we refrain from higher-order logics, as well as temporal or modal logics. We also use a parsimonious set of relations, following the OBO (Open Biological Ontologies) recommendation [8]. As a primitive formal relation we introduce the irreflexive, non-transitive and asymmetric instantiation relation *inst* which relates particular entities to their universal properties. In addition, we need a formal relation for class subsumption between universals, expressing scientific findings about relations between the kinds of things that are in the world. As scientific laws are meant to range not only over all present instances of a given kind, but also over all past and future instances and, moreover, also over merely possible instances [9], such a relation is not easily defined. We will here follow the OBO standard and introduce, to this end, the taxonomic subsumption relation *Is-a* by means of the *inst*[1] relation [8]. We will neglect the time parameter, which is not important for present purposes. On this basis, we define *Is-a* as a reflexive, transitive, and antisymmetric relation between universals A and B, as follows:

$$\begin{aligned} &Is\text{-}a(A,B) =_{def} \\ &\quad \forall x : (inst(x,A) \rightarrow inst(x,B)) \end{aligned} \tag{1}$$

Furthermore, we make the following ontological subdivision: When we deal with things of a certain kind, we have to distinguish between *individuals* belonging to this kind and *collectives* of individuals that belong to the same kind [1]. This very natural ontological distinction, which is mirrored by the singular / plural division in most natural languages, must be addressed wherever collectives or pluralities of individual objects occur. However, this distinction is often obscured when referring to mass entities (e.g. water vs. water molecules). Given the atomicity of material continuants, we do not admit material mass entities in our present framework, but consider them as collectives of particles instead.

---

[1] We use capitalized initial letters for the names of relations between universals as well as for the names of universals.

A collective is given, e.g., by all *Lmo-2* molecules involved in an experiment, as opposed to exactly one individual *Lmo-2* molecule. In the following, we will use the subscript "COLL" to refer to collectives. Thus, for each universal X we pricipally admit the existence of a corresponding collective X_{COLL}, the class of collections of instances of X. For instance, "$ProteinMolecule_{COLL}$" denotes the class of collectives of protein molecules as well as "$Lmo\text{-}2_{COLL}$" the class of collectives of *Lmo-2* molecules. We also admit collectives of occurrents, such as $Interaction_{COLL}$.

Concurrent interpretations I: Event Readings

Let us come back to our example: "*Lmo-2* interacts with *Elf-2*". Such statements are generally formulated by researchers who collect scientific evidence by empirical observations. These observations are commonly made in an indirect way, since the objects under scrutiny are below the threshold of visibility. For this reason measurement procedures of varying degrees of sophistication are applied, the results of which can be used to draw conclusions about the significance of an experiment. These conclusions may vary in their degrees of certainty. This certainty is affected by measurement errors as well as by errors in the design of the experiment which then may lead to false conclusions. If a statement like "*Lmo-2* interacts with *Elf-2*" is being uttered in a laboratory or written in a scientific paper or textbook, the minimal thing that can be inferred is that there are molecules of type *Lmo-2* and *Elf-2*. By way of contrast, this inference is not possible if such a sentence appears in a science fiction novel. As we are interested here in the scientific context only, we assume in what follows that there are universals *Lmo-2* and *Elf-2* that are kinds of protein molecules. These types of protein molecules do, of course, belong to the genus of protein molecules, which in turn are molecules, which are a kind of continuants. If one has an Aristotelian theory of universals, universals only exist if they are instantiated; that is, the existence of the universals *Lmo-2* and *Elf-2* implies that there are individual molecules that instantiate these universals. (Whoever has a different theory of universals may have to add this as a further assumption.) All the possible readings of "*Lmo-2* interacts with *Elf-2*" to be discussed in the remainder of this paper have thus the following as a common ground:

$$\begin{aligned}
&Is\text{-}a(Lmo\text{-}2, ProteinMolecule) \land \\
&Is\text{-}a(Elf\text{-}2, ProteinMolecule) \land \\
&Is\text{-}a(ProteinMolecule, Molecule) \land \\
&Is\text{-}a(Molecule, Continuant) \land \\
&\exists l, e : inst(l, Lmo\text{-}2) \land inst(e, Elf\text{-}2)
\end{aligned} \qquad (2)$$

Despite this common ground, the sentence remains highly ambiguous even within a scientific context. First we will discuss interpretations of "*Lmo-2* interacts with *Elf-2*" that interpret it as a report of events. Here are some possible interpretations of the sample statement that belong to this group:

1. One individual *Lmo-2* molecule interacts with one individual *Elf-2* molecule.
2. A collection of *Lmo-2* molecules interacts with one individual *Elf-2* molecule.
3. One individual *Lmo-2* molecule interacts with a collection of *Elf-2* molecules.
4. A collection of *Lmo-2* molecules interacts with a collection of *Elf-2* molecules.

Our sample statement appears to describe the fact that exactly one such interaction happened. Alternatively, it can describe the fact that a multitude of such interactions (as described in 1-4) happens, which would be the normal thing in many biochemical contexts. This adds up to eight different interpretations. But any of these interpretations is still ambiguous in a very important respect. With each of these interpretations, the speaker may mean either that such interaction(s) did actually happen, or the speaker may mean that the molecules in question have the disposition or the tendency to interact in such a way. This gives way to even more possible interpretations. Thus, "*Lmo-2* interacts with *Elf-2*" turns out to be a highly ambiguous sentence. We will now discuss the different possible interpretations of this sentence in turn and suggest methods for representing them formally.

Occurrents involving individual continuants

On the first interpretation, "*Lmo-2* interacts with *Elf-2*" describes the fact that an individual *Lmo-2* molecule interacts with an individual *Elf-2* molecule. A standard way to render such a situation formally would be the use of the existence quantifier of first order predicate logic:

$$\exists l, e : inst(l, Lmo\text{-}2) \wedge inst(e, Elf\text{-}2) \wedge interacts(l, e) \tag{3}$$

This formalization ensures that there is *at least* one individual *Lmo-2* molecule which interacts with *at least* one individual *Elf-2* molecule at *at least* one instant. This interpretation can now be modified in various ways. We could, e.g., add exclusivity postulates like in (4) that ensure that *exactly* one individual molecule of each kind are interacting with each other. Though such a solitary event might be rarely observed in experiments, there may be contexts where this is the intended meaning:

$$\begin{aligned}&\exists l, e : inst(l, Lmo\text{-}2) \wedge inst(e, Elf\text{-}2) \wedge \\ &\quad interacts(l, e) \wedge \\ &\quad \forall l^*, e^* : (inst(l^*, Lmo\text{-}2) \wedge inst(e^*, Elf\text{-}2) \wedge \\ &\qquad interacts(l^*, e^*)) \rightarrow (l^* = l \wedge e^* = e)\end{aligned} \tag{4}$$

Normally, however, this formalization will be much too strong an interpretation of our sample statement. For any statement of this form will be *false*, if at any other time another *Lmo-2* molecule interacts with an *Elf-2* molecule – or if at the very same time another *Lmo-2* molecule interacts with an *Elf-2* molecule at any other place. Therefore, we do not consider it as a useful interpretation of our sample sentence. We will, however, refer back to this formula and the exclusivity clauses used in it in the following discussion.

In Formulae 3 and 4 we have expressed the interaction event by means of a binary relation *interacts* between individual continuants. This relation on the level of instances is irreflexive (nothing ever interacts with itself), symmetric, and non-transitive. The OBO (Open Biological Ontologies) relation ontologies, however, recommends to restrict ourselves to a parsimonious array of basic relations. Therefore, we will eliminate the *interacts* relation, using the technique introduced by Davidson [10] to quantify over events. This means that we represent the interaction process as an occurrent entity in its own right rather than by the relation *interacts* as in Formulae 3 or 4. This move is made possible through our admission of occurrent entities, and it corresponds to common practice in biomedical ontologies. The relation between the particular process and the participating particular continuants is then given by the relation *has-participant* [8]. The *has-participant* relation is a relation between a particular occurrent and a particular continuant, in this order. It is irreflexive, asymmetric, and non-transitive. For nothing participates in a continuant and no occurrent participates in anything. Again, we dispense with a time index for sake of simplicity. Within a fully-fledged implementation, a time index should be included, as an occurrent may have different participants at different stages.

$$\begin{aligned}&\exists l, e, i : inst(l, Lmo\text{-}2) \wedge inst(e, Elf\text{-}2) \wedge \\ &\quad inst(i, Interaction) \wedge \\ &\quad has\text{-}participant(i, l) \wedge has\text{-}participant(i, e)\end{aligned} \tag{5}$$

This formalization makes it easier to represent occurrents with more than two participants, as with the representation in Formula 3, where we would have to deal with n-ary relations for n participants. According to this formal representation, "*Lmo-2* interacts with *Elf-2*" is to be understood as stating that there is at least one interaction process, in which at least one protein molecule of the given kinds is involved. It does not exclude that other molecules are involved in this very interaction process. If we want to secure that *Lmo-2* and *Elf-2* are the only participants of the molecular interaction, we have to employ exclusivity conditions similar to 4:

$$\begin{aligned}&\exists l, e, i : inst(l, Lmo\text{-}2) \wedge inst(e, Elf\text{-}2) \wedge \\ &\quad inst(i, Interaction) \wedge \\ &\quad has\text{-}participant(i, l) \wedge has\text{-}participant(i, e) \wedge \\ &\quad \forall x : (has\text{-}participant(i, x) \rightarrow \\ &\qquad inst(x, Lmo\text{-}2) \vee inst(x, Elf\text{-}2))\end{aligned} \tag{6}$$

If we want to keep the requirement of pairwise interaction, if have to add uniqueness conditions in the fashion of Formula 4 for this purpose:

$$\begin{aligned}
&\exists l, e, i : inst(l, Lmo\text{-}2) \wedge inst(e, Elf\text{-}2)\wedge \\
&\quad inst(i, Interaction) \wedge \\
&\quad has\text{-}participant(i, l) \wedge has\text{-}participant(i, e) \wedge \\
&\quad \forall x : (has\text{-}participant(i, x) \rightarrow \\
&\qquad inst(x, Lmo\text{-}2) \vee inst(x, Elf\text{-}2)) \wedge \\
&\quad \forall l^*, e^* : (inst(l^*, Lmo\text{-}2) \wedge inst(e^*, Elf\text{-}2) \wedge \\
&\qquad has\text{-}participant(i, l^*) \wedge has\text{-}participant(i, e^*)) \\
&\qquad \rightarrow (e^* = e \wedge l^* = l))
\end{aligned} \tag{7}$$

In contrast to Formula 4, such a formalization that quantifies over events is still much more realistic, because its truth is compatible with more than one interaction process of the same kind happening at the same time or at other times.

Occurrents involving collectives of continuants

As mentioned above, it is important to distinguish between individuals of a kind and collectives of individuals of that kind. Rector and Bittner [1] have accounted for this by introducing the formal relation *has-grain* which relates a collective c to each of its constituents e. In [3] this account has been further developed by introducing a collective universal X_{COLL} whose instances are constituted by two or more constituents which are instances of X:

$$\forall c : inst(c, X_{COLL}) \rightarrow \exists e_1, e_2, ..., e_n, n > 1 : \bigwedge_{\nu=1}^{n} inst(e_\nu, X) \wedge has\text{-}grain(c, e_\nu) \tag{8}$$

Note that *has-grain* is a subrelation of *has-part*. As a consequence, we identify a collection as a mereological sum of its constituents (regardless of their spatiotemporal arrangement), and not as a mathematical set. The reason for rejecting the set approach is two-fold. Firstly, because mathematical sets are extensional and therefore not robust with regard to the gain and loss of constituents. Secondly, because collectives should be of the same ontological category as their constituents: A collective of material objects should be a material object, and a collective of events should be an event. Sets, however, are abstract objects that do neither exist in space nor in time. We do not use the *has-part* relation, because participants in interactions may have parts that do not themselves participate in the interaction. A *Lmo-2* molecule, e.g., may participate in an interaction without every of its electrons being a participant in this interaction. Whereas *has-part* is transitive, *has-grain* is not. It is a irreflexive, asymmetric, and intransitive relation that holds between particular collectives and individuals.

We therefore modify our formalism substituting individuals by collectives:

$$\begin{aligned}
&\exists l, e, i : inst(l, Lmo\text{-}2_{COLL})\wedge \\
&\quad inst(e, Elf\text{-}2_{COLL}) \wedge inst(i, Interaction) \wedge \\
&\quad has\text{-}participant(i, l) \wedge has\text{-}participant(i, e)
\end{aligned} \tag{9}$$

Collectives of occurrents

Formalism 7 and 9 use the same occurrent type *Interaction* for different scenarios: In the first case, a particular interaction has individual protein molecules as participants, in the second case collectives of molecules. This ambiguity may be acceptable when we talk about such a generic process as interaction. It would not be tolerable in the case of a more specific one, such as *binding*. A binding can only happen between two individual molecules, not between two collectives of molecules. Thus, if we encounter a plurality of bindings within a plurality of molecules, it would not be admissible to describe this as a binding between two collectives of molecules but rather a collective of bindings between pairs instances of the kinds of molecules in question[2]. Thus we have to deal with a collective of processes rather than with collectives of continuants.

In order to represent such a situation, let us first introduce the collective interaction universal I_{COLL} which is constituted by individual constituents which are instances of I, analogously to Formula 8. Then we have to determine how each of the grain interactions look like. If they are pairwise interactions between an *Lmo-2* molecule and an *Elf-2* molecule, each of these interactions fits

---

[2] A counterexample is the interaction between solutes and solvents in a solution which necessarily involves colletives of both solvents and solutes.

Formula 7. Combining Formulae 7 and 8, we get:

$$\begin{aligned}
&\exists p, i_1, i_2, ...i_n, n > 1 : \qquad (10)\\
&\bigwedge_{\nu=1}^{n} (\mathit{inst}(i_\nu, I) \wedge \mathit{has\text{-}grain}(p, i_\nu) \wedge\\
&\quad \exists l_\nu, e_\nu : \mathit{inst}(l_\nu, \mathit{Lmo}\text{-}2) \wedge \mathit{inst}(e_\nu, \mathit{Elf}\text{-}2) \wedge\\
&\quad\quad \mathit{has\text{-}participant}(i_\nu, l_\nu) \wedge\\
&\quad\quad \mathit{has\text{-}participant}(i_\nu, e_\nu) \wedge\\
&\quad \forall x : (\mathit{has\text{-}participant}(i_\nu, x) \rightarrow\\
&\quad\quad \mathit{inst}(x, \mathit{Lmo}\text{-}2) \vee \mathit{inst}(x, \mathit{Elf}\text{-}2)) \wedge\\
&\quad \forall l^*_\nu, e^*_\nu : ((\mathit{inst}(l^*_\nu, \mathit{Lmo}\text{-}2) \wedge\\
&\quad\quad \mathit{inst}(e^*_\nu, \mathit{Elf}\text{-}2) \wedge \mathit{has\text{-}participant}(i_\nu, l^*_\nu) \wedge\\
&\quad\quad \mathit{has\text{-}participant}(i_\nu, e^*_\nu)) \rightarrow (e^*_\nu = e_\nu \wedge l^*_\nu = l_\nu)))
\end{aligned}$$

Concurrent Interpretations II: Dispositional Readings

The above interpretations stated the existence of one or more interaction events. However, messages of the style "*Lmo-2* interacts with *Elf-2*" very often do not focus on the accidental occurrence of an event but are rather meant to express some inherent property of the objects under investigation. On the one hand it is likely that a biologist would mean "An interaction between *Lmo-2* and *Elf-2* happened" while describing the outcome of a specific experiment. On the other hand a biology textbook would rather want to communicate something like "*Lmo-2* molecules have the disposition or tendency to interact with *Elf-2* molecules". This ambiguity, of course, matches Aristotle's famous distinction between act and potency, and Aristotle himself observed that "potency" is in itself an ambiguous term [11]. Thus the ambiguity of our sample statement increases even more, because the dispositional reading of our sample sentence is ambiguous in itself. Obviously, such a reading of "*Lmo-2* interacts with *Elf-2*" is intended to ascribe some causal or statistical property, a disposition or tendency. But even if this is the common ground of the dispositional reading, three questions remain open and have to be answered:

1. Which event is it exactly that the property in question is meant to cause?
2. What is thought to be the bearer of this property?
3. Which kind of property is in fact intended to be ascribed?

The first question can be answered by pointing to one of the many event readings we discussed (and formalized) thus far. Our answer to the second question will at least in part depend on our response to question 1. Are all instances of a given universal bearers of the disposition in question? Or only some of the instances? Are the individual molecules the bearers of the disposition, or rather collectives of such molecules? The third question, however, leads us in to the middle of the lively debate going on in philosophy on the ontology of disposition [12, 13, 14]. The dispositional properties most often discussed in the literature are so-called *surefire dispositions*: dispositions to react invariably in a certain way under specific circumstances. They are one candidate for an answer to question 3. From the point of view of knowledge representation, however, there are some problems connected with surefire dispositions. First, things may react differently in different circumstances. Thus to say that *Lmo-2* molecules have the disposition to interact with *Elf-2* molecules still leaves it open under which circumstances such an interaction will occur. We could account for this by explicitly mentioning the conditions of realization for each disposition. We may, of course, not *know* all these conditions, but this is an epistemic problem only. A more significant problem is that there may be infinitely many causally relevant conditions that have to be taken into account, and such an infinite list would be impossible for principled reason. We could try to circumvent this problem by adding (implicitly or explicitly) quantification phrases like "In all circumstances" or "In some circumstances". The *all*-phrase, however, will not do. For if a certain disposition would be realized under all circumstances, it will never be unrealized. Such cases may exist, but normally a disposition will only be realized under certain circumstances and not realized under others. When we use the *some*-phrase, on the other hand, many statements about dispositions for molecule interactions will become trivial, since nearly any molecule may interact with any other molecule in some peculiar way under certain (possibly very extreme) conditions. A usual way to deal with this problem is to introduce a set of standard or normal conditions [15]. In biology, this could mean that the "disposition to interact with *Elf-2* molecules" is only ascribed to *Lmo-2* if the interaction commonly occurs under biological conditions, such as physiological pH and temperature intervals. But the problem is not solved by referring to normal conditions. For, first, the problem that infinitely many conditions cannot be described in necessarily finite lists recurs with normal conditions. And, second, biomedical knowledge may also include the behavior of molecules in non-normal or even extreme circumstances, like low or high temperatures, exposure to intensive sunlight or atomic radiations. One option at this point would be to choose a different answer to question 3. Instead of ascribing surefire dispositions we could ascribe probabilistic dispositions, i.e. dispositions to do something (under certain circumstances) with a certain probability [16]. Such causal properties are also sometimes called "tendencies" [17] or "propensities" [18]. While with surefire dispositions a certain event will happen invariably in given circumstances, the event in question will only happen with a certain probability when a tendency is ascribed. It will, of course, be crucial to know with *which* probability the event will happen. Following standard procedures in mathematical probability theory, we can represent the quantities of the probabilities in question by real numbers between 0 and 1 satisfying the Kolmogorov axioms. In biomedical experiments, the observed result is often such a probability. Tendencies are thus of vital importance for the representation of biomedical knowledge [17]. There can, however, be several ontological groundings for such a probability. Suppose that we observed a hundred instances of a given universal U in situations in which all conditions necessary for the realization R of a certain disposition were present, but that in only fifty cases R happened, i.e. in only 50 % of all cases the disposition realized itself. There are several ontological scenarios that would explain this result. Here are two of them:

- (A) Every instance of U has a tendency to R with a probability of 0.5.
- (B) Every second instance of U has a surefire disposition to R; the other instances of U do not have any disposition to R.

Both of these scenarios would explain the assumed observations. Which of these scenarios we choose for our account of the observation will depend on other observations and causal assumptions. If we, e.g. knew that nearly always the same instances of U display R and nearly always the same instances of U do not display R, this would *prima facie* count as a reason to embrace (B). If, on the other hand, we know that the same instances of U sometimes do display R and sometimes do not display R, this would *prima facie* count as a reason to embrace (A). For such reasoning, however, we need background assumptions about the stability of the causal properties in question: how they can be stable over time, how (if at all) they can be acquired and how (if at all) they can get lost. Last but not least, (B) can indicate that the instances of the universal U differ in certain features, which are crucial to the ability to display R. An important example of this is the observation of modified proteins produced by mutated genes opposed to the observation of normal (wild-type) proteins. Considering all this, there is quite a long and complex list of enitities that we implicitly refer to when ascribing a disposition or tendency to a molecule:

- (independent) continuants (i.e. the bearer of the disposition),
- dependent continuants or occurrents (i.e. the realization),
- quantities (of probabilities), and
- state of affairs (of realization conditions).

Conclusion

Our deliberations shed light on the need for a more principled account of dispositions and processes in biomedical ontologies. Machine supported information extraction and knowledge acquisition techniques from scientific texts have become a cornerstone in molecular biology and genomics due to the increasing scientific productivity in this field.

The necessity of logic based ontologies for this purpose has been controversially discussed [19]. If we subscribe to a formally principled account as a basis for the semantic representation of the content of scientific texts then we have to take into account that the most common type of statements that are of interest in texts describing biochemical regularities do not have a clear and unambiguous meaning. Assertions of the type "A interacts with B" are generally more than accounts of a single event. Rather they refer to a plurality of events of the same kind, or an event involving pluralities (collectives) of participants. A universal interpretation such as "For each instance of A there is an interaction with some B" can easily be discarded. The need for universal quantifications can be satisfied by introducing dispositions: "Every A has the disposition to interact with some B." However, not every occurrence of the participation of some continuant in some process is proof of the existence of a related disposition.

Since interaction is a very general term, it is difficult to express a clear preference in favor of any of the proposed approaches without analyzing the nature of interaction on a molecular level, as well as the study of the "normal" behavior of biomolecules. The question when to ascribe a disposition or tendency – and which one – can not be discussed here (but cf. [16] on this).

We demonstrated that sentences like "A interacts with B" exhibit indeed a wide range of ambiguity. We offered several possible analyses to formally represent the different meanings of sentences of this type. Now, which one should we choose? One strategy would be to say: Which strategy you choose depends on the intended meaning of the particular occurrence of the sentence you deal with. For text mining purposes, however, that have to digest large amounts of texts in short periods of time and with as much automatization as possible, this strategy would be scarcely feasible. To cope with this situation, several strategies are conceivable. One strategy would be to choose the highest common factor of all interpretation – that what is included in all. Another strategy would be to set as a standard interpretation that is *most likely* the intended meaning. In order to determine which interpretation is the best candidate, empirical work on relevant text corpora may be helpful. This, however, is already beyond the scope of the present paper.

Acknowledgments:
This work was supported by the EU Network of Excellence *Semantic Interoperability and Data Mining in Biomedicine* (NoE 507505), the project *Forms of Life* sponsored by the Volkswagen Foundation, and the *Wolfgang Paul Award* of the Alexander-von-Humboldt-Foundation. We are indebted to Andrew D. Spear and to the anonymous referees of KR-MED for valuable comments.

Address for Correspondence:
Stefan Schulz, Department of Medical Informatics, Freiburg University Hospital, Stefan-Meier-Str. 26, 79104 Freiburg (Germany), phone: +49 761 203 6702, e-mail: stschulz@uni-freiburg.de

References

[1] Alan Rector, Jeremy Rogers, and Thomas Bittner. Granularity scale and collectivity: When size does and doesn't matter. *Journal of Biomedical Informatics*, 38, 2005.

[2] Stefan Schulz and Anand Kumar. Biomedical ontologies: What part-of is and isn't. *Journal of Biomedical Informatics*, 38, 2005.

[3] Stefan Schulz, Elena Beisswanger, Udo Hahn, Joachim Wermter, Anand Kumar, and Holger Stenzhorn. From GENIA to BioTop towards a top-level ontology for biology. *FOIS 2006 – International Conference on Formal Ontology in Information Systems*, 2006. Accepted for publication.

[4] Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. Sweetening ontologies with DOLCE. In Asunción Gómez-Pérez and V. Richard Benjamins, editors, *Proceedings of the 13th International Conference – EKAW 2002*, volume 2473 of *Lecture Notes in Artificial Intelligence*, pages 166–181. Berlin: Springer, 2002.

[5] Barry Smith and Pierre Grenon. The cornucopia of formal-ontological relations. *Dialectica*, 58(3):279–296, 2004.

[6] Barbara Heller and Heinrich Herre. Ontological categories in GOL. *Axiomathes*, 14(1):57–76, 2004.

[7] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook. Theory, Implementation, and Applications.* Cambridge, U.K.: Cambridge University Press, 2003.

[8] Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L. Rector, and Cornelius Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(5), 2005.

[9] Karl. R. Popper. *Logic of Scientific Discovery.* Hutchinson, London, 1959.

[10] Donald Davidson. The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*, pages 81–95. Pittsburgh, PA: University of Pittsburgh Press, 1967.

[11] Ludger Jansen. *Tun und Können. Ein systematische Kommentar zu Aristoteles' Theorie dispositionaler Eigenschaften im neunten Buch der Metaphysik.* Hänsel-Hohenhausen, Frankfurt am Main, 2002.

[12] Raimo Tuomela. *Dispositions.* Reidel, Dordrecht, 1978.

[13] Stephen Mumford. *Dispositions.* Oxford University Press, Oxford, 1998.

[14] Bruno Gnassounou and Max Kistler. *Les dispositions en philosophie et en sciences.* CNRS Editions, Paris, 2006. An English translation is to be published with Ashgate Publishers.

[15] Wolfgang Spohn. Begründungen a priori – oder: ein frischer Blick auf Dispositionsprädikate. In Wolfgang Lenzen, editor, *Das weite Spektrum der analytischen Philosophie. Festschrift Franz von Kutschera*, pages 89–106. de Gruyter, Berlin/New York, 1997.

[16] Ludger Jansen. Attribuer des dispositions. In Bruno Gnassounou and Max Kistler, editors, *Les dispositions en philosophie et en sciences*, pages 89–106. CNRS Editions, Paris, 2006.

[17] Ludger Jansen. The ontology of tendencies and medical information sciences. In Ingvar Johansson, Bertin Klein, and Thomas Roth-Berghofer, editors, *WSPI 2006: Contributions to the Third International Workshop on Philosophy and Informatics*, volume 14 of *IFOMIS Reports*, pages 89–106. IFOMIS, Saarbrücken, 2006.

[18] Karl. R. Popper. *A World of Propensities.* Thoemmes, Bristol, 1990.

[19] Sophia Ananiadou and Jun'ichi Tsujii. Thesaurus or logical ontology, which one do we need for text mining? *Language Resources and Evaluation, Springer Science and Business Media B.V.*, 39(1):77–90, 2005.