

Metaphors and Anti-Unification

Helmar Gust and Kai-Uwe Kühnberger
Institute of Cognitive Science, University of Osnabrück
Katharinenstr. 24, 49069 Osnabrück, Germany
{hgust|kkuehnbe}@uos.de

Ute Schmid
Department of Mathematics and Computer Science, University of Osnabrück
Albrechtstr. 28, 49069 Osnabrück, Germany
schmid@informatik.uni-osnabrueck.de

Abstract

A classical approach of modeling metaphoric expressions uses a source concept network that is mapped to a target concept network. Both networks are usually represented as algebras. We present a representation using the mathematically sound framework of anti-unification. We will interpret metaphors as a special case of analogical reasoning for which anti-unification was already successfully applied. Anti-unification has the advantage that a common structural description of the involved concept networks can be given and the connection between source domain and target domain are more clearly specified.

Keywords: Metaphors, Analogies, Anti-Unification

1 INTRODUCTION

There are not many spelled-out examples of mathematical models for metaphors. Although conceptual discussions about metaphors on the one side and classifications of metaphors on the other exist to a large extent, there are not many frameworks that try to analyze the formal properties of metaphorical expressions.

An influential example for a formally more or less spelled-out theory of metaphors is Indurkha 1992. In this account, Indurkha represents the target concept network and the source concept network as algebras, i.e. as pairs $\mathfrak{A} = \langle A, \Phi \rangle$ resp. $\mathfrak{B} = \langle B, \Psi \rangle$ where A and B are sets of concepts and Φ and Ψ are sets of operations defined on A respectively B (each with a certain arity). The relation connecting target and source (called metaphorical relation) is considered again as an algebra defined on the product of \mathfrak{A} and \mathfrak{B} . Indurkha's framework, on the one hand, models nicely certain examples taken from analogies in formal languages, namely Hofstadter's string examples of proportional analogies. In these examples, relational properties of strings occurring in the form $(A : B) :: (C : ?)$ are considered (compare Hofstadter & The Fluid Analogies Research Group 1995). On the other hand, there is neither a single explicitly formalized application of this framework to metaphors in natural language nor an explicitly formalized application to certain other non-trivial types of analogical reasoning different from string examples. We think that a major reason for this fact is the quite unnatural usage of algebras in Indurkha's framework. A second reason may be the complicated task to allow changes and errors in the establishment of the metaphoric relation: Whereas formal languages provide a precisely defined field of investigation, natural language metaphors and analogies in other domains are notoriously vague and therefore subject to errors, misunderstandings, and wrong conceptualizations.

We propose a different approach for modeling metaphoric expressions using anti-unification. A basic claim of our conceptual modeling of metaphors is the role analogical reasoning plays in an appropriate conceptualization. We think that a large number of metaphors can be reduced to

analogical reasoning and the same frameworks that can be used for analogical reasoning can be used for many types of metaphors as well.¹ A reason for this is the fact that many metaphors can be reformulated explicitly as analogies. For example, consider the following metaphoric expression (1):

(1) *Gills are the lungs of fish.*

Clearly (1) can be reformulated as a proportional analogy by making the association of the underlying concepts explicit:

(2) *Gills are to fish as lungs are to mammals.*

Although (2) specifies the relation between gills and lungs explicitly, whereas (1) does not, the meaning of (1) and (2) are equivalent provided an appropriate context is given in which (1) and (2) occur. An example, for such a context would be the zero context. An important claim of this paper is that certain metaphors can be modeled using the same method that can successfully be used for analogies, namely anti-unification.

The idea behind the anti-unification framework is basically to use a general structural description of both, target and source, in order to give an analysis of a metaphorical expression. In other words, the common structure needs to be identified. The theory of anti-unification can be used for giving such a general structural description, determining the common structure of the involved domains, provided there exists a term algebra formalizing the involved concept networks. Such a structural description (together with substitutions) is usually called anti-instance. This approach is known as calculating the most special generalization (compare Plotkin 1969) or as anti-unification (compare Reynolds 1970). A precise formulation of the theory as well as applications to certain types of analogies in naive physics was developed in Gust et. al. 2003a.

The remaining parts of this paper have the following structure: In Section 2, we will discuss the similarities and differences between analogies and metaphors. A development of the basic ideas of the framework of anti-unification together with an spelled-out example of an predictive analogy from naive physics will be presented in Section 3, followed by a an application of this framework to metaphors in Section 4. Finally in Section 5 some ideas concerning further work and a conclusion will be presented.

2 METAPHORS AND ANALOGIES

2.1 SIMILARITIES BETWEEN METAPHORS AND ANALOGIES

Analogies occur in a variety of domains, as well as in different forms. In order to classify certain aspects and properties of analogical reasoning we propose the following classification of analogies (Indurkha 1992, Schmid et. al. 2003):

- **Proportional analogies:** They have the general form $(A : B) :: (C : ?)$. These analogies were studied in the domain of natural language ("*Lungs are to humans as gills are to [fish]*"), with respect to geometric figures (Evans 1968, O'Hara 1992), and in string domains (Hofstadter & The Fluid Analogies Research Group 1995). Algebraic accounts for proportional analogies were developed and successfully applied.
- **Predictive analogies:** These analogies explain a new domain (target) by specifying similarities with a given domain (base), i.e. by transferring information from the base to the target (cf. Gentner 1983). Examples are metaphoric expressions from the domain of natural

¹An independent support of this claim is Gentner et. al. 2001.

language ("*Electricity is the water in an electric circuit*") as well as complex conceptualizations of physical correlations between seemingly different domains (compare Section 3 for a detailed example).

- **Analogical problem solving:** This type of analogical reasoning is used to solve a problem by transferring a solution from a well-known domain to a less well-known domain. An example is the usage of a LISP program as a starting point for developing new program code (Anderson & Thompson 1989).

It is not claimed that the above classification of analogies is complete, nor is it claimed that it is the only possible one. Rather such a classification can be useful to specify different properties of analogies: Whereas proportional analogies are determined by their form, predictive analogies can vary across domains, and analogical problem solving transforms *solutions* of problems from the source to the target.

It is often mentioned that metaphors are strongly connected to analogical reasoning (Gentner et. al. 2001, Indurkha 1992). As the classification above makes clear certain metaphors can be considered as proportional analogies and predictive analogies. Dependent on contextual features one and the same metaphor can be proportional or predictive. Consider the following metaphor:

(3) *Electrons are the planets of the atom.*

In a situation where a high school teacher is lecturing students elementary atom physics, (3) can be interpreted as a predictive analogy: By an understanding of this analogy, the students learn a new conceptualization of the atom. On the other hand, (3) can be simply interpreted as a proportional analogy, for example, in a situation where we examine properties of types of analogical reasoning.

In order to understand the metaphor represented by (3), it is necessary to figure out that electrons play the functional role in the atom as planets do in the solar system (Rutherford analogy). Hence, an association between the functional role *revolves_around* between planets and the sun and a functional role *revolves_around* between electrons and the nucleus needs to be established. Anti-unification can be used to find a common generalization of this functional role, namely the *revolves_around* relation between two variables that can be instantiated by the corresponding concepts.

Predictive analogies were studied in the realm of naive physics. Quite explicitly spelled-out examples of predictive analogies in naive physics can be found in Gust et. al. 2003a, Gust et. al. 2003b, and Schmid et. al. 2003. In these papers, the authors give the theoretical foundation of anti-unification for axiomatic theories and examine, for example, the analogy between a heat-flow system and a water-flow system: Like water flowing from a beaker to a vial in the case both are connected and the pressure in the beaker is higher than the pressure in the vial, heat flows via a bar connecting a cup with hot coffee and an ice-cube. As can be seen in Gust et. al. 2003a, relevant concepts and laws of physics can be represented by a many-sorted term algebra and the analogical inference can be drawn from anti-instances of a sufficiently specified conceptualization of the water-flow system. A PROLOG program designed for solving predictive analogies in physics with anti-unification can be found in Gust et. al. 2003a as well. We propose to use these ideas for a modeling of metaphors as well. In order to get a flavor for the overall framework, we will present a spelled-out example of a predictive analogy in Section 3.

2.2 DIFFERENCES BETWEEN METAPHORS AND ANALOGIES

Although metaphors and analogies show a lot of similarities they require a different modeling. Assume the metaphoric expression (3) is given establishing an analogy between electrons and planets. In comparison to predictive analogies in naive physics, the metaphoric expression (3) is both, simpler and more complicated:

- Example (3) is simpler because several, sometimes quite complicated laws of physics need not to be formalized in order to get a conceptualization and understanding: In many relevant discourse contexts, a metaphorical relation establishes a simple analogy between particular properties of the involved concepts. In this respect metaphors require just some (often clearly preferred) transferred facts from the source to the target, whereas in the case of predictive analogies whole theories need to be considered.
- Metaphor (3) is more complicated than predictive analogies, because there is no clear correct or incorrect way of modeling: Dependent on the context, the usage, and the intention of the speaker metaphors can mean many different things, whereas in physics, a predictive analogy either *is* in accordance to a given conceptualization of the laws of physics or *is not*.

The two mentioned aspects have certain consequences for our representation. In particular, the modeling of predictive analogies as described in Gust et. al. 2003b and Schmid et. al. 2003 cannot be applied one-to-one to metaphoric expressions: In the physical realm there are theories that govern analogical reasoning, but no such theories are, in general, available for metaphoric expressions.² Furthermore, whereas theories in physics provide a rich basis on which reasoning can be established, no such theories are available for metaphors. Instead of theories, at most lexical meanings of certain concepts can be presupposed. To make things even worse, such lexical meanings can be quite different across speakers and therefore no common ground can be assumed. For example, the concept of a planet for a physicist is clearly embedded in a physical theory where certain laws govern the behavior of such an entity in a solar system. On the other hand, for a native speaker of natural language, the concept of a planet is less clear: the speaker knows examples (like Earth), perhaps she knows that planets occur always together with a sun, and perhaps she knows that planets usually revolve around this sun. But this seems to be the maximal set of facts that can be assumed concerning the assumptions that govern metaphorical expressions.

3 USING ANTI-UNIFICATION FOR PREDICTIVE ANALOGIES

3.1 A SIMPLE EXAMPLE FROM PHYSICS: THE RUTHERFORD ANALOGY

We will use the theory of anti-unification to model predictive analogies in naive physics. Anti-unification is formally founded on the mathematics of term algebras (cf. Plotkin 1969). We extend this framework to anti-unify not only terms but whole theories (for the details of the theoretical background compare Gust et. al. 2003a). Because of the formally sound mathematical framework it is possible to represent precise statements of a state of affairs and reasoning about the quality of solutions. Furthermore, efficient algorithms can be derived from the framework in a straightforward way. We will describe how anti-unification works using a simple example of a predictive analogy. We want to represent the situation where a conceptualization of the solar system can be used to get a new conceptualization of the Rutherford atom model. The following ideas summarize the account in Gust et. al. 2003b where the interested reader is referred to for a more explicit discussion of the topic. Another spelled-out example in the realm of naive physics can be found in Schmid et. al. 2003. Furthermore, PROLOG programs can be found in Gust et. al. 2003a that solve these predictive analogies. In this section, we will only give a rough overview how anti-unification can be used to model predictive analogies.

The solar system is conceptualized by a model \mathfrak{M}_1 as depicted in Table 1: *Planets* and *sun* are considered to be objects. With respect to these objects certain observable properties (or features) are measurable by performing experiments: the mass of an object, the distance between two objects, and a force between two objects, called gravity, as well as the centrifugal force between two objects - provided an object o_1 is following a circular path around an object o_2 . Additionally, certain facts and laws about objects are given governing the behavior of the system.

²Although those theories are considered as qualitative (not quantitative) and essentially causal in nature (and not associative), they give a large amount of information about the involved domains.

Table 1: Modeling the physics of a solar system (\mathfrak{M}_1)

<p><i>types</i> <i>real, object, time</i></p> <p><i>entities</i> <i>planet : object</i> <i>sun : object</i></p> <p><i>functions</i> <i>observable mass: object × time → real × {kg}</i> <i>observable dist: object × object × time → real × {m}</i> <i>observable gravity: object × object × time → real × {N}</i> <i>observable centrifugal: object × object × time → real × {N}</i></p> <p><i>facts</i> <i>revolves_around(planet, sun)</i> <i>mass(sun) > mass(planet)</i> $\forall t : time : gravity(planet, sun, t) > 0$ $\forall t : time : dist(planet, sun, t) > 0$</p>	<p><i>laws</i> $\forall t : time, o_1 : object, o_2 : object :$ $dist(o_1, o_2, t) > 0 \wedge$ $gravity(o_1, o_2, t) > 0$ \rightarrow $\exists force : force(o_1, o_2, t) < 0 \wedge$ $force(o_1, o_2, t) = centrifugal(o_1, o_2, t)$</p> <p>$\forall t : time, o_1 : object, o_2 : object :$ $dist(o_1, o_2, t) > 0 \wedge$ $centrifugal(o_1, o_2, t) < 0$ \rightarrow $revolves_around(o_1, o_2)$</p>
--	--

Now we will consider the atom model given by model \mathfrak{M}_2 (Table 2). The conceptualization of the atom does not contain a comparable amount of information as the conceptualization of the solar system, because otherwise the establishment of a creative analogy would not be necessary. As objects *electron* and *nucleus* are given. Observable properties are the electric charge of objects as well as masses of objects. Additionally we assume that the Coloumb force between two objects can be measured. Concerning facts governing electron and nucleus, we presuppose that the electron as well as the nucleus have a mass and an electric charge. The latter is the reason why there is a Coloumb force attracting the two objects. Notice that gravity as well as the Coloumb force have the same direction, i.e. both forces attract electrons and the nucleus (represented by *gravity* (*electron, nucleus, t*) > 0 and *coloumb*(*electron, nucleus, t*) > 0). As long as we are interested in a qualitative analysis of the atom model, it is sufficient to consider only one force, namely the one with the grater magnitude, i.e. the Coloumb force. Last but not least, we are able to perform experiments, in order to test whether analogical transfers yield experimentally valid results. One of these experiments is essentially an abstract representation of the Rutherford experiment, i.e. an experiment that shows that electrons and nucleus have a distance from each other greater than 0.

Notice that the predicate *revolves_around* has no corresponding predicate in the target domain. Simply transferring this fact to the target would be possible in principal, but there is no way to test in an experiment whether this predicate applies in the target domain. A better modeling is to give an explanation why these concepts can be used in the target domain. This can be achieved by performing an experiment measuring that $dist(electron, nucleus, t) > 0$ and by applying a general (transferred) law from the base that results in the fact *revolves_around*(*electron, nucleus*).

3.2 THE ANALOGICAL TRANSFER

Anti-unification is the attempt to find generalizations (anti-instances) of the two models \mathfrak{M}_1 and \mathfrak{M}_2 . Let us first consider the classical case of term anti-unification: Anti-unifying two first-order terms t_1 and t_2 of a term algebra means to construct a third term t and two substitutions Θ_1 and Θ_2 such that $t_1 = t\Theta_1$ and $t_2 = t\Theta_2$. This can be naturally extended to the general case: A substitution Θ assigns values to variables.

An important concept in this framework is subsumption. A term s subsumes a term t relative to a given equational theory E if it holds (Burghardt & Heinz 1996):

$$s <_E t : \iff \exists \Theta : E \vdash s\Theta = t$$

A term t is called an anti-instance of a set of terms T if t subsumes all t' of T . An equational theory E is introduced because we want to be able to represent equivalences between expressions of the form $a < b$ and $b > a$. Although this is not crucial for our considerations here, such

Table 2: Modeling the physics of the atom model (\mathfrak{M}_2)

<p><i>types</i> <i>real, object, time</i></p> <p><i>entities</i> <i>electron: object</i> <i>nucleus : object</i></p> <p><i>functions</i> <i>observable mass: object \times time \rightarrow real \times {kg}</i> <i>observable dist: object \times object \times time \rightarrow real \times {m}</i> <i>observable electric_charge: object \rightarrow real \times {eV}</i> <i>observable coloumb: object \times object \times time \rightarrow real \times {N}</i></p>	<p><i>facts</i> <i>mass(nucleus) > mass(electron)</i> <i>electric_charge(electron) < 0</i> <i>electric_charge(nucleus) > 0</i> <i>$\forall t : time : coloumb(electron, nucleus, t) > 0$</i></p> <p><i>experiment</i> <i>$\forall t : time : dist(electron, nucleus, t) > 0$</i></p>
--	--

equivalences can be important for implementations of the framework.

In a concrete situation usually one is confronted with a whole bunch of anti-instances. Then, it is natural to ask for the set of those anti-instances that are most specific, complete, and minimal (Gust et. al. 2003a). These anti-instances can be identified with generalizations that represent structural descriptions of certain objects.

The sketched first-order case of anti-unification is simple and straightforward. But for our purposes we need a weak form of second-order anti-unification. We make this clear using the following example: Consider the following equivalence (where the expression on the left side is an expression of equational theory E_1 and on the right side there is an expression of equational theory E_2):

$$f(h(x, h(a, b))) \leftrightarrow g(h(a, b))$$

Syntactic anti-unification results in the anti-instance $F(h(X, Y))$ where the substitutions Θ_1 and Θ_2 are given as follows:

$$\begin{aligned} \Theta_1/\Theta_2 & : F \mapsto f/g \\ & X \mapsto x/a \\ & Y \mapsto h(a, b)/b \end{aligned}$$

Although F is a second-order variable, those second-order phenomena are not problematic. A more complicated second-order generalization is discussed in Schmid et. al. 2003. We come back to our example of the Rutherford analogy. The following table summarizes the anti-instances of the anti-unification process (we use e and n as shortcuts for *electron* resp. *nucleus* and p and s for *planet* resp. *sun*):

Table 3: Anti-Instances of our Modeling

Base	Target	A
$mass(s) > mass(p)$	$mass(n) > mass(e)$	$mass(Y) > mass(X)$
$rev_around(p, s)$	$rev_around(e, n)$	$rev_around(X, Y)$
$gravity(p, s, t) > 0$	$coloumb(e, n, t) > 0$	$F(X, Y, t) > 0$
$dist(p, s, t) > 0$	$dist(e, n, t) > 0$	$dist(X, Y, t) > 0$

Applying appropriate substitutions to the anti-instances yield again facts of our models \mathfrak{M}_1 and \mathfrak{M}_2 . Here are the corresponding substitutions Θ_1 and Θ_2 for the anti-unification process with the property that $Base = A\Theta_1$ and $Target = A\Theta_2$:

$$\begin{aligned} \Theta_1/\Theta_2 & : X \mapsto planet/electron \\ & Y \mapsto sun/nucleus \\ & F \mapsto gravity/coloumb \end{aligned}$$

Notice that by transferring the laws of the base domain to the target domain we get hypothetical laws in the target domain as well. These laws are not simply mapped one-to-one to the target but accordingly to the governing anti-instances. Table 4 specifies the result of transferring the laws from the base to the target domain (governed by the anti-instances):

Table 4: Hypotheses of the Target Domain

<p><i>laws</i></p> $\forall t : time, o_1 : object, o_2 : object :$ $dist(o_1, o_2, t) > 0 \wedge$ $coloumb(o_1, o_2, t) > 0$ \rightarrow $\exists force : force(o_1, o_2, t) < 0 \wedge$ $force(o_1, o_2, t) = centrifugal(o_1, o_2, t)$	$\forall t : time, o_1 : object, o_2 : object :$ $dist(o_1, o_2, t) > 0 \wedge$ $centrifugal(o_1, o_2, t) < 0$ \rightarrow $revolves_around(o_1, o_2)$
---	---

A remark concerning the laws of the base domain should be added. These laws are transferred to the target domain with their respective interpretation. Just because we can apply these laws, it is possible to *deduce* that an electron is revolving around a nucleus, i.e. we can give an explanation why the electron is revolving. Hence, modeling the Rutherford atom model in this way provides a possibility to model the creative (or generative) aspect of predictive analogies.

4 THE MODELING OF METAPHORS

Having set up the machinery so far, the application of anti-unification to metaphoric expressions can be developed in this section. The structure of this section is as follows: First, we will restrict the types of metaphors to those from the physical realm having a particular form. Second, we will roughly point out the differences of the Rutherford analogy between the physical realm and the corresponding metaphorical expression used in ordinary discourse situations and we will continue to model metaphoric expressions in a naive way. Third, we will refine our naive modeling by specifying which conceptualization must be assumed for modeling the metaphor.

4.1 TYPES OF METAPHORS

There are many different classifications of metaphors. Furthermore, there are many closely related concepts to metaphors in linguistic theories like idioms, forms of irony, similes and the like. In the following, we will discuss roughly which types of metaphors we will consider in this section, namely metaphors that connect noun phrases directly by a form of *to be*. Examples of this type of metaphors are the following ones:

- (4)(i) *Electrons are the planets of the atom.*
- (ii) *Electricity is the water of an electric circuit.*
- (iii) *Lawyers are sharks.*
- (iv) *Juliet is the sun.*

Another type of metaphoric expressions assigns a particular attribute to a concept (noun phrase) that typically would not be considered as applicable in a conventional interpretation. Reasons for the conventionally non-applicability of the attribute are usually sort problems: A liquid can have a color, it can be oily or transparent, it can flow, be cold, or be warm and the like, but usually a liquid cannot be soft like in (5)(i).

- (5)(i) *A soft wine.*
- (ii) *A cold warrior.*
- (iii) *A warm acknowledgment.*

In this paper, we will not consider metaphors of the type exemplified in (5), although we think that it is, in principal, possible to generalize and extend the present account to this type

of metaphors as well. Whereas examples like the ones in (4) and (5) are as simple as possible, this is quite often not true for metaphors occurring in poetic contexts. Examples like the often cited poem *Fog* by Carl Sandburg where *fog* is metaphorically correlated to cats (which is a quite non-obvious connection) are much more complicated to analyze than examples (4) and (5) (The poem is quoted according to Indurkha 1992):

*The fog comes
on little cat feet.*

*It sits looking
over harbor and city
on silent haunches
and then moves on.*

We will not consider those poetic metaphors in our modeling either. Another type of expressions that are related to metaphors are idioms. They are usually considered as lexicalized metaphors, i.e. as metaphors that are already transformed into conventionally interpreted expressions. Idioms will not play a role in this paper. Although it would be interesting to investigate how it is possible for a particular expression to change from a creative metaphor to a conventional metaphor, i.e. an idiom, we will not provide an account for these language change phenomena. Furthermore, examples of irony and similes will not count as metaphors here and will not be considered in this section. Last but not least, we do not analyze or discuss many types of unconventional usage of concepts in natural languages as metaphorical. For example, if Tom says to Jim that his fish is in the living room referring to the wooden fish he bought in Singapore, then this non-conventional usage of the concept fish has in our opinion nothing to do with metaphors.

A further aspect concerns the domain of discourse. As a matter of fact, metaphors seem to occur in nearly every domain of natural language discourse.³ For our purposes in this paper, we will exemplify anti-unification just in the domain of naive physics, in order to show the general possibility to apply this framework. The extension of this account to a variety of domains and types of metaphors will be postponed to another paper.

4.2 THE RUTHERFORD ANALOGY AS A PREDICTIVE ANALOGY AND AS A METAPHOR

Consider again the Rutherford analogy as given by the natural language description in (4)(i). Our modeling in Subsection 3.1 and Subsection 3.2 presupposes a logical reformulation of physical theories – clearly with the restriction that the representation is qualitative not quantitative. These theories were used to anti-unify laws and facts from the base and the target domains. Whereas this is a natural setting for modeling predictive analogies we think that in the case of metaphorical expressions this approach is not appropriate: Understanding a metaphor like (4)(i) does generally not presuppose broad knowledge about facts and laws of physical theories. A rather simple and straightforward solution would be just to apply some thinning procedure to the models \mathfrak{M}_1 and \mathfrak{M}_2 used to represent the predictive analogy case and to use this simplified version for modeling the two realms in the metaphor case. We choose a slight variation of a simplified model of Subsection 3.1 where the crucial preconditions that need to hold in order to apply a law are coded in a more or less vague predicate *approp_constellation*. The following table summarizes the model of the base domain of the Rutherford metaphor.

³Possibly there are certain restrictions across languages. For example, in certain North-Australian languages the usage of metaphors seems to be more restricted (compare Goddard forthcoming).

Table 5: Naive Modeling of the Base Domain of the Rutherford Metaphor

<p><i>types</i> <i>object</i></p> <p><i>entities</i> <i>planet: object</i> <i>sun: object</i></p> <p><i>functions</i> <i>mass: object</i> \rightarrow <i>real</i> \times $\{kg\}$</p>	<p><i>facts</i> <i>revolving_around(planet, sun)</i> <i>mass(sun) > mass(planet)</i></p> <p><i>laws</i> <i>mass(sun) > mass(planet) \wedge</i> <i>approp_constellation(planet, sun)</i> \rightarrow <i>revolving_around(planet, sun)</i></p>
--	--

Notice that we only assumed that there is some knowledge about the masses of the involved objects as well as the fact that planets revolving around the sun. The law adds a relatively vague claim about sufficient conditions about solar systems in order to guarantee that planets orbit around the sun.

A very similar modeling can be proposed for the target domain. Because of the fact that the target should be newly conceptualized transferring information from the base to the target, we do not need to assume a lot of facts or laws in the target domain. Table 6 specifies some aspects how a naive modeling could look like.

Table 6: Naive Modeling of the Target Domain of the Rutherford Metaphor

<p><i>types</i> <i>object</i></p> <p><i>entities</i> <i>electron: object</i> <i>nucleus: object</i></p>	<p><i>functions</i> <i>mass: object</i> \rightarrow <i>real</i> \times $\{kg\}$</p> <p><i>facts</i> <i>mass(nucleus) > mass(electron)</i></p>
--	---

Understanding metaphor (4)(i) can be interpreted as establishing successfully this *revolving_around* relation between *electron* and *nucleus*. The standard procedure of anti-unification can do the job. Using the facts *mass(sun) > mass(planet)* in the base domain and *mass(nucleus) > mass(electron)* in the target domain we can easily establish a connection between *planet* and *electron* on the one side and *sun* and *nucleus* on the other by finding appropriate anti-instances of these facts. Again we apply the procedure that laws and facts that occur in M_1 but not in M_2 need to be transferred from M_1 to M_2 . This is a crucial step in the whole set-up and marks a creative aspect of the reasoning, because new information is transported from one (well-understood) model to another (less understood) model. The association of *planet* and *electron* on one side and *sun* and *nucleus* on the other together with a transfer of the *revolving_around* relation yields the desired result: The most specific generalization of source and target establishes the following fact (interpreted as the most specific generalization of the corresponding fact in the base domain): *revolving_around(X, Y)*. With this generalization the following substitutions Θ_1 and Θ_2 are associated:

$$\begin{aligned} \Theta_1/\Theta_2 : \quad X &\mapsto \textit{planet} / \textit{electron} \\ &Y \mapsto \textit{sun} / \textit{nucleus} \end{aligned}$$

The described modeling is relatively simple and straightforward but shows also certain problems. First, it is not clear why it should be possible at all to infer the desired conclusion using simple lexicalized meanings of the involved concepts. In particular, whether some abstract feature like mass need crucially be assumed in the modeling or not is not absolutely clear. Second, there

are often preferred interpretations of metaphors where a particular property of a concept guides the whole correspondence process between base and target. Nothing in the modeling designates a certain property as the preferred one. In particular with respect to transferred laws and facts, this can become important: whereas in the physics realm, experiments can be performed in order to test whether a particular transfer is successful or not, this is not possible in the realm of metaphoric expressions. Consequently, we omitted a testing procedure (experiment) in the modeling of the target. Preferred attributes would give a hint how we could get a reliable solution without available testing procedures. Third, the conceptualization is in a certain sense too general: a law in the modeling of the base domain (like in Table 5) does not play any role in an understanding process of such a metaphor. In total, it seems to be the case that the naive (and simple) way to model metaphor (4)(i) causes several non-trivial problems. In the next subsection, we will refine our modeling of the Rutherford metaphor.

4.3 LEXICALIZED CONCEPTS

In our view, metaphors operate on a basis that corresponds to a large extent to lexical meanings of the involved concepts. Furthermore, the meaning of concepts does not include such an enormous amount of abstract information as in the case of predictive analogies in naive physics. In particular, what is missing is a *theory* of a particular domain under consideration: The lexical meaning of a concept most often does not involve a spelled-out conceptualization comparable to current scientific theories. Rather one or more preferred properties are often associated with metaphors governing the new non-conventional meaning of the involved target concepts.⁴

We need to make things more precise concerning metaphor (4)(i): Considering the concept *planet*, this concept is specified via a binary relation between *planet* and *sun* where additionally a conceptualization of *sun* must be available:

- The concept *sun* is a lexicalized entity. As a possible conceptualization *sun* can have the following properties: it occurs with other objects (like *planets*) and builds the center of a more complex system that includes *sun* and these other objects. Important is that the expansion of this system is finite, i.e. that it is nothing that is arbitrarily extended.
- A relation R to another object *sun* defines the concept *planet*. R is a two-ary relation $R(x, y)$ together with a certain sort restriction with respect to x and y . The idea is that sort restriction only allows y to be of sort *object* (and not of sort *real* or *time* or anything like this, because *sun* is an object not a real number). $R(x, y)$ is necessary to assume because an object *sun* is not mentioned in (4)(i) and $R(x, y)$ can introduce an object like *sun*. Hence, this conceptualization of *planet* introduces a concept *sun* and links both concepts together via a two-ary relation.
- Preferred properties are assigned to *planet*. For our purposes the preferred one is *revolving_around* the *sun*. (In other metaphors preferred properties can be *heavy*, *round* and the like.)

It is less clear, how much information concerning the gravitational center of solar systems needs to be introduced at this point. This changes dramatically with respect to the intended speaker: Whereas children usually do not know anything about gravitation and a central gravitation system, people well-educated in physics – although non-specialists – can know a lot about these things. For us, it is sufficient to assume that *sun* has the properties specified above.

Although the above remarks give a first idea how a modeling of the situation could look like there is one point missing: What is $R(x, y)$ in this account? Is it necessary that the relation $R(x, y)$ covers the *revolving_around* relation? Or is it rather the case that from R and the conceptualization of *sun* the *revolving_around* property can be implied? Both possibilities do not seem to be too counterintuitive. We assume here that the *revolving_around* constraint is

⁴Notice that we restrict our attention to context independent metaphors. Taken context into account would clearly make things more complicated.

separately covered by a preferred property of *planet* and $R(x, y)$ is used to introduce the concept *sun* as well as linking *sun* and *planet* both concepts together. But clearly this seems not to be the only possibility. To summarize the above considerations, the modeling of the base domain of the Rutherford metaphor (4)(i) can look like follows:

Table 7: Modeling of the Base Domain

<i>types</i>	<i>facts</i>
<i>object</i>	<i>revolving_around(planet, sun)</i>
<i>entities</i>	<i>R(planet, sun)</i>
<i>planet: object</i>	
<i>sun: object</i>	

On the target side, essentially two objects are mentioned: *electron* and *nucleus*. Because both concepts occur together we can assume that a relation $R'(x, y)$ holds between electron and nucleus and this relation is restricted with respect to the involved sorts.

Table 8: Modeling of the Target Domain

<i>types</i>	<i>facts</i>
<i>object</i>	<i>R'(electron, nucleus)</i>
<i>entities</i>	
<i>electron: object</i>	
<i>nucleus: object</i>	

Using this modeling of the domains, it is possible to apply the anti-unification process. Anti-unification yields the desired generalization establishing the fact *revolving_around(electron, nucleus)* in the target domain in a straightforward way.

4.4 THE GENERAL CASE

The example in the above subsection makes crucial assumptions about the involved domains. For the general case of metaphors in the physical realm, it is clear that such assumptions play an important role for the understanding of metaphors: Often it is not only the lexical meaning of the involved contexts but furthermore simply world knowledge of the hearer that makes an understanding of metaphors possible. This is particularly true for metaphors in the physical realm.

Here are some prerequisites that are needed for a successful establishment of a metaphor:

- Concepts of the base domain need to be introduced and (or) linked by a relation R
- Concepts of the target domain need to be linked via a relation R'
- Designated properties need to be assigned to elements of the base domain
- These designated properties can be transferred to the target using the anti-unification machinery

Clearly most of the knowledge that is required to understand metaphors is covered in the designated properties of concepts in the base domain. In general, several of these properties needs to be assumed. For example, *planet* could be in a similar role as *sun* in a metaphor like (6):

(6) *An electron in a hydrogen atom is the moon of this atom*

In this example, *planet* needs to be introduced and linked to *moon*. The relevant properties of *planet* change dramatically: Now the gravitation center is considered to be the planet. What is needed for an appropriate modeling is a list of designated properties of the involved concepts. Such a list must contain relevant information concerning possible properties of concepts that can play a role in metaphors. Furthermore, it seems to be reasonable to rank these properties according to their importance. Clearly, this is a highly empirical problem, but it is necessary to get a further idea for a semantics of metaphors on the one hand and a tractable algorithm on the other.

5 CONCLUSION AND FURTHER WORK

In this paper, we showed that on the one hand, metaphors and analogies are closely related to each other but on the other both show significant differences at the same time. We introduced the framework of anti-unification as a means to model generalizations of certain theories together with the correlated instantiations of the underlying expressions in order to model predictive analogies from naive physics. Finally we proposed possibilities to adapt the framework of anti-unification to applications involving metaphors of natural language (restricted to a certain type of metaphor). We think that – although the extent to which this theory can be successfully applied to is not finally clear – it is nevertheless a promising approach that deserves further considerations.

It is clear that the proposal in this paper is only a first step towards a theory of metaphors. Here are some points that are planned for further research: First, the development of a formal semantics for the present approach should make the link between the algorithmic properties of anti-unification and the meaning of the generalizations visible. Second, a reasonable fragment of a natural language domain preferable from the realm of naive physics needs to be specified. Finally, the existing PROLOG programs that are currently extended to model variations of domains and heuristics in searching appropriate anti-instances for examples of predictive analogies in physical theories need to be applied to metaphors as well.

REFERENCES

- Anderson, J. & Thompson, R. (1989), Use of analogy in a production system architecture, in: *Similarity and analogical reasoning*, editors: Vosniadou & Ortony, Cambridge, pp. 267-297.
- Burghardt, J. & Heinz, B. (1996) *Implementing Anti-Unification Modulo Equational Theory*, Technical Report, Arbeitspapiere der GMD, vol. 1006.
- Dastani, M. (1998), *Languages of Perception*, ILLC Dissertation Series 1998-05, <http://www.cs.uu.nl/~mehdi/publications/thesis.ps.ps>.
- Evans, T. (1968), A program for the solution of a class of geometric-analogy intelligence-questions, in: M. Minsky (ed.), *Semantic information processing*, MIT press, pp. 271-353.
- Gentner, D. (1983), Structure-mapping: A theoretical framework for analogy, *Cognitive Science* 7, pp. 155-170.
- Gentner, D. Bowdle, B. Wolff, P. & Boronat, C. (2001), Metaphor is like analogy, in *The analogical mind: Perspectives from cognitive science*, editors: Gentner, Holyoak, Kokinov, Cambridge MA, pp. 199-253.
- Goddard, C. (forthcoming), The ethnopragmatics and semantics of 'active' metaphors, *Journal of Pragmatics* (Special Issue on 'Metaphor across languages', ed. by G. Steen), forthcoming.
- Gust, H., Schmid, U. & Kühnberger, K.-U. (2003a), *Anti-unification of axiomatic systems*, <http://www.cogsci.uni-osnabrueck.de/~helmar/analogy1.ps/>.
- Gust, H., Kühnberger, K.-U. & Schmid, U. (2003b), Solving Predictive Analogy Tasks with Anti-Unification, *Proceedings of the Joint International Conference on Cognitive Science 2003 (ICCS ASCS 2003)*, Sydney.

- Hofstadter, D. & The Fluid Analogies Research Group (1995), *Fluid concepts and creative analogies*, New York.
- Indurkha, B. (1992), *Metaphor and Cognition*, Dordrecht, The Netherlands, Kluwer.
- O'Hara, S. (1992), A model of the redescription process in the context of geometric proportional analogy problems, in: *Proc. Int. Workshop on Analogy and Inductive Inference (AII'92)*, Springer, pp. 268-293.
- Plotkin, G. (1969), A note on inductive generalization, in *Machine Intelligence*, vol. 5, pp. 153-163.
- Reynolds, J. (1970), Transformational systems and the algebraic structure of atomic formulas, in *Machine Intelligence*, vol. 5, pp. 135-151.
- Schmid, U., Gust, H., Kühnberger, K.-U. & Burghardt, J. (2003), An Algebraic Framework for Solving Proportional and Predictive Analogies, in *Proceedings EuroCogSci 03*, Osnabrück.