Grounded Models as a Basis for Intuitive and Deductive Reasoning: the Acquisition of Logical Categories

Josefina Sierra-Santibáñez[1]

Abstract. Grounded models differ from axiomatic theories in establishing explicit connections between language and reality that are learned through language games. This paper describes how grounded models are constructed by visually grounded autonomous agents playing different types of language games, and explains how they can be used for intuitive reasoning. It proposes a particular language game that can be used for simulating the generation of logical categories (such as negation, conjunction, disjunction, implication or equivalence), and describes some experiments in which a couple of visually grounded agents construct a grounded model that can be used for spatial reasoning.

1 INTRODUCTION

This paper builds up on earlier work on the relation between language acquisition and conceptualization in visually grounded robotic agents [1, 9, 11, 4] by addressing the issue of the acquisition of logical categories and proposing a particular language game (called the *evaluation game*) that can be used for simulating the generation of logical categories, such as negation, conjunction, disjunction, implication or equivalence. Logical categories are very important for intellectual development, because they allow the generation of structured units of meaning and they set the basis for deductive thinking.

It also describes how grounded models [4] can be used for simulating a form of intuitive reasoning that explains why we accept certain axioms as intuitively true without further argumentation or how we come up with some simple theories of *common sense reasoning* [2].

The rest of the paper is organized as follows. First, we review the concepts of sensory channel, category and categorizer [6], which are used for conceptualizing perceptual information. Then, we consider the process of truth evaluation, show how logical categories can be constructed by identifying sets of outcomes of the evaluation process, and explain how concepts can be generated combining logical and perceptually grounded categories. Next, we describe how grounded models are constructed by autonomous agents as they play different types of language games, and present some experiments in which a couple of visually grounded agents construct a grounded model that can be used for spatial reasoning. Finally, we explain how grounded models can be used for intuitive reasoning.

2 PERCEPTUALLY GROUNDED CATEGORIES

We consider an experimental setting similar to that used in *The Talking Heads Experiment* [9]: a set of robotic agents ("talking heads") playing language games with each other about scenes perceived through their cameras on a white board in front of them. This section describes how the agents conceptualize the perceptual information they obtain by looking at a particular area of the white board.

2.1 Sensory channels

The agents look at one area of the white board by capturing an image of that area with their cameras. First, they segment the image into coherent units in order to identify the objects that constitute the context of a language game. Next, some *sensory channels* (implemented by low level visual processes) gather information about each segment, such as its horizontal or vertical position. In this paper, we assume that there are only two primitive sensory channels.

- H(o), which computes the x-midposition of object o.
- V(o), which computes the y-midposition of object o.

The values returned by the sensory channels H and V are scaled with respect to the area of the white board captured by the agents cameras so that its range is the interval (0.0 1.0). Consider the three objects numbered in figure 1, object 1 has the values H(O1)=0.2, V(O1)=0.8, object 2 has the values H(O2)=0.2, V(O2)=0.2, and object 3 has the values H(O3)=0.8, V(O3)=0.2.

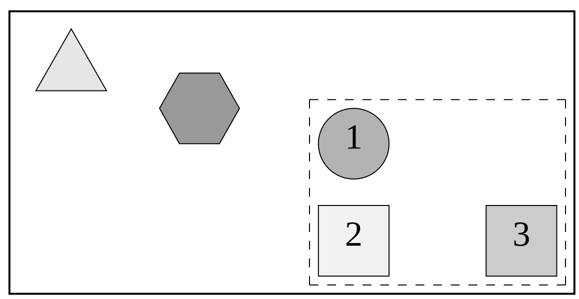

Figure 1. A typical configuration of the white board. The area of the white board captured by the agents cameras is the lower right rectangle.

In addition to two primitive sensory channels, there are other sensory channels constructed from them.

- HD(o1,o2) computes the difference of the x-midpositions of objects o1 and o2, i.e., HD(o1,o2)=H(o1) – H(o2).
- VD(o1,o2) computes the difference of the y-midpositions of objects o1 and o2.
- EQ(o1,o2) is defined as a predicate, i.e., a function which takes the value 1 (true) if the values of the primitive sensory channels H and V are equal for objects o1 and o2, and 0 (false) otherwise.

For example, the sensory channel VD has the value 0.6 when it is applied to the pair of objects (O1,O2), i.e., VD(O1,O2)=0.6. The range of the sensory channels HD and VD is (-1.0 1.0). The range of the sensory channel EQ is the discrete set of Boolean values $\{0, 1\}$.

[1] Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Madrid, email: Josefina.Sierra@ii.uam.es

2.2 Categories

The data returned by the sensory channels are values from a continuous domain (except for sensory channel EQ). To be the basis of a natural language conceptualization, these values must be transformed into a discrete domain. One form of categorization consists in dividing up each domain of output values of a particular sensory channel into regions and assigning a *category* to each region [6]. For example, the range of the H channel can be cut into two halves leading to a distinction between [LEFT] (0.0 < H(x) < 0.5) and [RIGHT] (0.5 < H(x) < 1.0). Object 3 in figure 1 has the value H(O3)=0.8 and would therefore be categorized as [RIGHT]. Similarly, the VD channel can be cut into two halves leading to a distinction between [ABOVE] (0.0 < VD(x) < 1.0) and [BELOW] (-1.0 < VD(x) < 0.0).

It is always possible to refine a category by dividing its region. Thus an agent can divide the bottom region of the H channel (categorized as [LEFT]) into two subregions [TOTALLY-LEFT] (0.0 < H(x) < 0.25), and [MID-LEFT] (0.25 < H(x) < 0.5). The categorization networks resulting from these consecutive binary divisions form *discrimination trees* [6].

Following the notation introduced in [9], we label categories using the sensory channel they operate on followed by the upper and lower bounds of the region they carve out. Thus [TOTALLY-LEFT] is labeled as [H 0.0 0.25], because it is true for a region between 0.0 and 0.25 on the H channel. We also use first order logic predicate notation to emphasize the fact that perceptually grounded categories correspond to n-ary predicates of first order logic. For example, we use the unary predicate [H 0.0 0.25](o) to refer to the category [H 0.0 0.25], and the binary predicate [VD 0.0 1.0](o1,o2) to refer to the category [VD 0.0 1.0].

2.3 Categorizers

At the same time the agents build categories to conceptualize perceptual information, they construct cognitive procedures (called *categorizers*) that allow them to check whether these categories hold or not for a given tuple of objects. Categorizers give grounded meanings to symbols [1] by establishing explicit connections between internal representations (categories) and reality (perceptual input as processed by sensory channels) [6]. These connections are learned through language games [12], and allow the agents to check whether a category holds or not for a given tuple of objects. Most importantly, they provide information on the perceptual and cognitive processes an agent must go through in order to evaluate a given category.

The behavior of the categorizers associated with the perceptually grounded categories used in this paper can be described by linear constraints. We use the notation [CAT]$^C(\vec{x})$ to refer to the categorizer which is capable of recognizing whether category [CAT]$(\vec{x})$ holds or not for a given tuple of objects. For example, the behavior of the categorizer associated with category *above* can be described as follows: [VD 0.0 1.0]$^C(x,y) \equiv 0.0 < \mathrm{V}(x) - \mathrm{V}(y) < 1.0$.

3 LOGICAL CATEGORIES

We consider now the process of truth evaluation, and describe how logical categories can be constructed by identifying sets of outcomes of the evaluation process. Logical categories are important because: (1) they allow the generation of structured units of meaning, which correspond to first order logic formulas; (2) they set the basis for deductive reasoning.

3.1 Evaluation channel

The *evaluation channel* (denoted by E) is a cognitive procedure capable of finding the categorizers of a tuple of categories, applying them to a tuple of objects, and observing their output. If $\vec{c} = (c_1, \ldots, c_n)$ is a category tuple and $\vec{o} = (o_1, \ldots, o_m)$ an object tuple, $E(\vec{c}, \vec{o})$ is a tuple of Boolean values $(v_1, \ldots, v_n)$, where each v_i is the result of applying c_i^C (the categorizer of c_i) to $\vec{o}$ (i.e., $v_i = c_i^C(\vec{o})$). For example, $E(([V\ 0.0\ 0.5](x), [H\ 0.5\ 1.0](x)), O1) = (0,0)$, because $O1$ (object 1 in figure 1) is neither on the lower part nor on the right part of the white board area captured by the agents' cameras.

3.2 Logical categories and concepts

Although the evaluation channel can be applied to category tuples of any arity, we will consider only unary and binary category tuples[2]. The range of the evaluation channel for unary tuples of categories is the set {1,0}, and its range for category tuples of length two is the set of Boolean pairs {(0,0),(0,1),(1,0),(1,1)}. The subsets of these sets correspond to the meanings of all the connectives of propositional logic (i.e., negation, conjunction, disjunction, implication and equivalence), plus the meanings of some conjunctions found in natural languages. For example, the sentence $c_1 \vee c_2$ is true if the result of evaluating the pair of categories (c_1, c_2) is a Boolean pair which belongs to the set $\{(1,1), (1,0), (0,1)\}$.

The agents construct *logical categories* by identifying subsets of the range of the evaluation channel. The evaluation game creates situations in which the agents discover such subsets, and use them to distinguish a subset of object tuples from other subsets of object tuples in a given context. The representation of logical categories as sets of Boolean tuples is, in fact, equivalent to the *truth tables* used in classical logic for describing the semantics of propositional connectives.

The notation used for describing logical categories consists of the symbol E (for evaluation channel) followed by a schematic representation of the set of Boolean tuples for which each category holds. Thus, [E 0](c) corresponds to $\neg c$, [E 11-10-01](c1,c2) to $c1 \vee c2$, [E 11](c1,c2) to $c1 \wedge c2$, [E 11-01-00](c1,c2) to $c1 \rightarrow c2$, and [E 11-00](c1,c2) to $c1 \leftrightarrow c2$.

Logical categories describe properties of concepts, therefore it is natural to apply them to perceptually grounded categories in order to construct structured units of meaning. For example, the concept [E 0]([V 0.0 0.5](x)) can be constructed by applying the logical category [E 0](c) (i.e., $\neg c$) to the category [V 0.0 0.5](x) (i.e., down(x)). If we consider perceptually grounded categories as atomic concepts, we can define the notion of *concept* (structure of meaning) by induction as follows: (1) a perceptually grounded category is a concept; (2) if $l(\vec{x})$ is an n-ary logical category and $\vec{c}$ is an n-ary tuple of concepts, then $l(\vec{c})$ is a concept[3].

3.3 Logical categorizers

The categorizers of logical categories are cognitive procedures that allow determining whether a logical category holds or not for a tuple of concepts and an object tuple. As we have explained above, logical categories can be associated with subsets of the range of the evaluation channel. The behavior of their categorizers can be described therefore by constraints of the form $E(\vec{c}, \vec{o}) \in S_l$, where l

[2] We omit the parentheses when the evaluation channel is applied to category tuples of length one.

[3] A concept is, therefore, any free quantifier first order logic formula that can be constructed from perceptually grounded categories.

is a logical category, S_l is the subset of the range of the evaluation channel for which l holds, E is the evaluation channel, $\vec{c}$ is a tuple of concepts, and $\vec{o}$ is an object tuple. For example, the constraint $E((c1, c2), \vec{o}) \in \{(1, 1)\}$ describes the behavior of the categorizer for the logical category [E 11](c1,c2) (i.e., $c1 \wedge c2$).

The evaluation channel can be naturally extended to evaluate generic concepts using the categorizers of logical and perceptually grounded categories. The following is an inductive definition of the evaluation channel $E(\vec{c}, \vec{o})$ for generic concepts.

1. If $[CAT](\vec{x})$ is a perceptually grounded category, then $E([CAT](\vec{x}), \vec{o}) = [CAT]^C(\vec{o})$.
2. If $l(\vec{c})$ is a concept, where l is a logical category, $\vec{c}$ is a tuple of concepts and S_l is the subset of the range of the evaluation channel for which l holds, then $E(l(\vec{c}), \vec{o}) = 1$ if $E(\vec{c}, \vec{o}) \in S_l$, and 0 otherwise.

4 CONSTRUCTING GROUNDED MODELS

In [4], a *grounded model* is defined as the set of concepts and categorizers constructed by an agent at a given time in its development history. This section describes how grounded models are constructed by visually grounded autonomous agents playing different types of language games. First, the agents play language games of the sort described in the book *The Talking Heads Experiment* [9]. These games allow them to construct perceptually grounded categories, categorizers and a shared lexicon for referring to them. Once the agents have learned a shared lexicon for perceptually grounded categories, they start playing evaluation games, which allow them to construct logical categories and a shared lexicon for referring to such categories.

In this section, we describe the evaluation game, and refer the reader to [9] for a detailed description of the language games used for learning perceptually grounded categories, in the context of an interesting research on the origins of language and meaning [7].

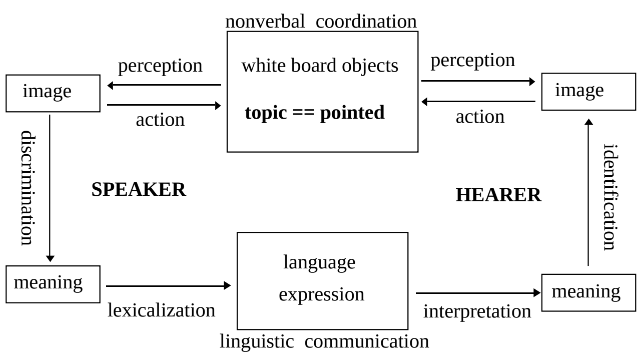

Figure 2. The semiotic square (taken from [9]) summarizes the main cognitive processes involved in a language game.

The *evaluation game* is played by two agents. One agent plays the role of *speaker*, and the other agent plays the role of *hearer*. The main cognitive processes involved in the game are summarized in figure 2.

Perception First, the speaker looks at one area of the white board and directs the attention of the hearer to the same area[4]. Both, speaker and hearer, segment their images of that area of the white board into coherent units, which roughly correspond to different objects pasted on the white board. Then, the speaker chooses a *context* for the game. This is a set of object tuples of the same arity. The speaker indicates the hearer the set of object tuples that constitute the context by pointing to them. Then, the speaker picks up a subset of object tuples from the context which we will call the *topic*. The rest of the object tuples in the context constitute the *background*. Both, speaker and hearer, use their sensory channels to gather information about each object tuple in the context, and store that information so that they can use it in subsequent stages of the game.

For example, in evaluation game 999 (described below) the speaker chooses as context the set $\{(O1), (O2), (O3)\}$, where $O1, O2$ and $O3$ represent the objects numbered in figure 1. The topic chosen by the speaker is the subset of object tuples $\{(O1), (O3)\}$, and the background the subset $\{(O2)\}$.

Discrimination The speaker tries to find a pair of categories which distinguishes the topic from the background. That is, a pair of categories $(c1, c2)$ such that its evaluation on the topic is different from its evaluation on any object tuple in the background. The evaluation of a tuple of concepts on a set of object tuples (e.g., the topic) is defined as follows. Let E be the evaluation channel, $\vec{c}$ a tuple of concepts, and S a set of object tuples, $E(\vec{c}, S) = \{E(\vec{c}, \vec{o})\}_{\vec{o} \in S}$, where $E(\vec{c}, \vec{o})$ is as defined in previous sections. The speaker tries to find, then, a pair of categories $(c1, c2)$ such that $E((c1, c2), \text{topic}) \bigcap E((c1, c2), \text{background}) = \emptyset$.

The pair of categories ([V 0.0 0.5](x),[H 0.5 1.0](x)) (i.e., (down(x),right(x))) distinguishes the topic from the background in evaluation game 999, because E(([V 0.0 0.5](x),[H 0.5 1.0](x)),{O3,O1})= {(1,1),(0,0)}, E(([V 0.0 0.5](x),[H 0.5 1.0](x)),{O2})= {(1,0)}, and {(1,1),(0,0)} $\bigcap$ {(1,0)}=∅.

If the speaker cannot find a discriminating pair of categories, the game fails. Otherwise, it tries to find a logical category which is associated with the set of Boolean pairs T resulting from evaluating the pair of categories on the topic. If it does not have any logical category associated with this set, it creates a new logical category of the form [E T](c1,c2) and adds it to its current grounded model. The concept constructed by applying this logical category to the categories in the discriminating pair characterizes the topic as *the set of object tuples in the context that satisfy the concept*.

For example, in evaluation game 999, the speaker characterizes the topic {(O1),(O3)} as the set of object tuples in {(O1),(O2),(O3)} that satisfy the concept [E 11-00]([V 0.0 0.5](x),[H 0.5 1.0](x)).

Lexicalization The speaker examines its lexicon in order to find an expression associated with the logical category. The *lexicon* of an agent contains associations of the form (C,W,R), where C is a category, W is a linguistic expression, and R is the rate of the association (i.e., a number between 0 and 1 which corresponds to the confidence the agent has on the usefulness of that association). If there are several expressions associated with the logical category, the speaker chooses an expression of an association with highest rate. If there are no associations for the logical category C, the speaker may invent a new expression W with probability R_C, and add an association of the form $(C, W, 0)$ to its current lexicon. The constant R_C is the *creativity* rate of the agent. In our experiments R_C is 0.9 for all agents. For example, if the speaker has no associations for the logical category [E 11-00](c1,c2), it may invent a new expression (e.g., *iff*) and add the association ([E 11-00](x,y),iff(x,y),0) to its lexicon.

In this paper, we assume that concepts (i.e., internal representations) and linguistic expressions have identical structures, so that the agents do not have to construct and learn a syntax for their language (see [8] and [10] for some experiments on the origins of syntax in visually grounded agents). This simplification allows us to concentrate on the issue of the acquisition of logical categories from the point of view of their semantical function.

[4] Robots direct the attention of others to specific areas or objects on the white board by informing each other of the direction where they are focusing.

Next, the speaker constructs a sentence using the lexicalizations of the logical category and the categories in the discriminating pair. At this stage, it is assumed that the agents have learned a shared lexicon for perceptually grounded categories already. For example, the speaker may construct the sentence *iff(down(x),right(x))* to express the concept [E 11-00]([V 0.0 0.5](x),[H 0.5 1.0](x)), assuming the associations ([E 11-00],iff(x,y),R1), ([V 0.0 0.5](x),down(x),R2) and ([H 0.5 1.0](x),right(x),R3) are part of its lexicon already.

Once the speaker has constructed a sentence which expresses the concept that characterizes the topic, it communicates that sentence to the hearer.

Interpretation The hearer interprets the sentence using the associations between (logical or perceptually grounded) categories and linguistic expressions in its lexicon. That is, it tries to reconstruct the concept encoded by the sentence. If the hearer does not have any association for the outer most expression of the sentence, the game fails and a repair process takes place. The speaker points to the topic so that the hearer can guess the logical category it has used to conceptualize the topic, and acquire an association between that logical category and the expression used by the speaker. This happens with probability R_A. The constant R_A is the *assimilation* rate of the agent. In our experiments R_A is 0.9 for all agents.

For example, in evaluation game 999 the hearer may interpret the sentence *iff(down(x),right(x))* as the concept [E 11-00]([V 0.0 0.5](x),[H 0.5 1.0](x)), if its lexicon contains the right associations between categories and linguistic expressions. Otherwise, it may acquire the association ([E 11-00](x,y),iff(x,y),0) as a result of a repair process.

Identification The hearer tries to find the referent, i.e., the set of object tuples in the context that satisfy the meaning of the sentence. For example, the set of object tuples that satisfy the meaning of *iff(down(x),right(x))* in evaluation game 999 is {(O1),(O3)}. If the meaning of the sentence is true for all the object tuples in the context or for none of them, the game fails and a repair process similar to that described in the previous step takes place. Failures such as these may happen when speaker and hearer associate the same expression with different logical categories.

Coordination The hearer points to the referent, i.e., to the set of object tuples it identified in the previous step. The game succeeds if the referent guessed by the hearer is equal to the set of object tuples the speaker had in mind (the topic). If the game succeeds, speaker and hearer increment the rates of the associations they used for lexicalizing and interpreting the logical category by an amount I, and decrement the rates of competing associations which were discarded by both agents during the processes of lexicalization and interpretation. If the game fails, only the rates of the associations used by speaker and hearer are decremented, and a repair process similar to that of the previous step takes place. In our experiments, the amount I by which rates are modified is 0.1.

We summarize below the main steps of evaluation game 999, which has been used as an example in our previous discussion.

```
Game number: 999. Speaker: A1. Hearer: A2.
Topic: {(O1),(O3)}. Background: {(O2)}.
Speaker conceptualizes topic as:
[E 11-00]([V 0 0.5](x),[H 0.5 1](x)).
Speaker lexicalizes concept as:
iff(down(x),right(x)).
Hearer interprets sentence as:
[E 11-00]([V 0 0.5](x),[H 0.5 1](x)).
Hearer points to: {(O1),(O3)}.
Speaker says: OK. The game is a success.
```

4.1 Experiments

We have run some experiments in order to see whether a couple of visually grounded autonomous agents can learn a set of perceptually grounded categories for spatial reasoning, and a set of logical categories which allow them to construct logical formulas from perceptually grounded categories and to evaluate those formulas using the categorizers of perceptually grounded and logical categories.

In the first experiment, the agents play language games of the sort described in [9]. The experiment shows that the agents can construct: (1) a grounded model which includes the categories *up, down, right, left, above, below, rightof and leftof*; and (2) a shared lexicon for referring to them, in less than 250 language games.

In the second experiment, the agents play evaluation games. These games require the existence of a shared lexicon for perceptually grounded categories and are, therefore, played after the first experiment has been completed. The experiment shows that the agents: (1) extend their grounded models with logical categories for all the logical connectives used in propositional logic; and (2) construct a shared lexicon for referring to them. In this experiment, the agents reach a lexical coherence of 0.97 after playing 3000 games. The *lexical coherence* [9] is a measure of the similarity of the agents' lexicons.

5 INTUITIVE REASONING

A natural question is to ask what the agents have really learned by playing language games. In first place, they have constructed categorizers for perceptually grounded categories (such as up, down, right, left, above, below, rightof and leftof), and for logical categories (such as negation, disjunction, conjunction, implication or equivalence). In second place, they have acquired a shared vocabulary for referring to those categories. And, in third place, they have learned to evaluate generic concepts (i.e., free quantifier first order logic formulas) using the categorizers of logical and perceptually grounded categories.

According to the result of the evaluation process, the concepts that can be constructed by an agent can be classified into three categories: (1) *intuitive truths*, which are true for every object tuple; (2) *intuitive falsehoods*, which are false for every object tuple; and (3) *regular concepts*, which are true for some object tuples and false for others.

Intuitive reasoning is a process by which the agents discover relationships that hold among the categorizers of basic concepts. For example, an agent may discover that the concept $above(x, y) \wedge above(y, x)$ is an intuitive falsehood, because the categorizers of $above(x, y)$ and $above(y, x)$ never return true at the same time. Similarly, it can discover that the concept $above(x, y) \rightarrow \neg above(y, x)$ is an intuitive truth, because the categorizer of $above(y, x)$ returns false whenever the categorizer of $above(x, y)$ returns true.

Intuitive reasoning can be used then for determining whether a concept is an intuitive truth, an intuitive falsehood or a regular concept of a grounded model. It may work as a process of constraint satisfaction in natural agents, by which they try to discover whether there is any combination of values of their sensory channels that satisfies a given concept. It is not clear to us, how this process of constraint satisfaction can be implemented in natural agents. It may be the result of a simulation process by which the agents generate possible combinations of values for their sensory channels and check whether they satisfy a given concept. Or it may be grounded on the physical impossibility of firing simultaneously some categorizers due to the way they are implemented by physically connected neural networks in natural agents.

To get an idea of the set of intuitive truths that are implied by the grounded model G constructed by a couple of agents in the previous experiments, consider the first order logic theory T_S. The language of T_S consists of two binary predicate symbols $A(x, y)$ (above) and $R(x, y)$ (rightof). Its set of non-logical axioms is as follows.

$$A(x, y) \rightarrow \neg A(y, x) \quad (1)$$

$$A(x, y) \wedge A(y, z) \rightarrow A(x, z) \quad (2)$$

$$A(x, y) \wedge A(x, z) \wedge \neg R(y, z) \wedge \neg R(z, y) \wedge y \neq z \rightarrow A(y, z) \vee A(z, y) \quad (3)$$

It is easy to see that every theorem of T_S is an intuitive truth of grounded model G, because every axiom of T_S is an intuitive truth of G, and the intuitive truths of G are closed under the inference rule of resolution.

When the behavior of the categorizers for basic concepts (i.e., perceptually grounded categories) can be described by linear constraints, intuitive reasoning can be implemented as a process of linear constraint satisfaction. This is so, because the behavior of the categorizer of every concept can be described by a disjunction of linear constraint systems which can be computed by replacing every category by a linear constraint describing the behavior of its categorizer in the concept, and computing the disjunctive normal form of the constraint system resulting from the substitution. Once this transformation has been done, showing that a concept is an intuitive truth is equivalent to proving that the disjunction of constraint systems associated with the concept is true for every value of the sensory channels, and this can be done by proving that the disjunction of constraint systems associated with the negation of the concept is unsatisfiable.

For example, it can be shown that axiom 2 (which states that the relation above – A(x,y) – is transitive) is an intuitive truth of grounded model G by checking that the following disjunction of linear constraint systems is unsatisfiable for every value of x, y and z in the interval (0.0 1.0). This formula has been obtained by: (1) replacing every category by a linear constraint describing the behavior of its categorizer in the concept associated with the negation of axiom 2; (2) replacing every instance of V(x), V(y) and V(z) by x, y and z in the expression resulting from step 1; and (3) computing the disjunctive normal form of the expression resulting from step 2.

$$\{0 < x\text{-}y,\ x\text{-}y < 1.0,\ 0 < y\text{-}z,\ y\text{-}z < 1.0,\ x\text{-}z \leq 0\} \vee$$

$$\{0 < x\text{-}y,\ x\text{-}y < 1.0,\ 0 < y\text{-}z,\ y\text{-}z < 1.0,\ 1.0 \leq x\text{-}z\}$$

It is easy to check that this disjunction of linear constraint systems is unsatisfiable. A disjunction of constraint systems is unsatisfiable if every disjunct is unsatisfiable, and each disjunct is a linear constraint system that can be checked by a linear constraint solver, such as the one implemented in Sicstus Prolog.

The rest of the axioms of T_S can be shown to be intuitive truths of G by constraint satisfaction as well. That is, the couple of agents of the previous experiments can discover that the relation *above* is *antisymmetric, transitive,* and *a total order on objects located on the same horizontal position* by intuitive reasoning. In this sense, one can say that grounded model G provides an explanation for the fact that most people accept these axioms as intuitively true, and use them for building logical theories for *common sense* reasoning [2].

It can also be proved that intuitive reasoning in grounded models in which the behavior of the categorizers for basic concepts can be described by linear constraints is closed under resolution. That is, if two concepts are intuitive truths of a grounded model, its resolvent is an intuitive truth of the grounded model as well. This is a consequence of the *soundness* of the inference rule of resolution, and the fact that the linear constraints describing the behavior of the categorizers of a grounded model constitute a model (in the sense of model theory semantics) of every intuitive truth of the grounded model[5].

Every theorem of T_S is, therefore, an intuitive truth of grounded model G, but not every intuitive truth of G is a theorem of T_S, because grounded model G has categorizers for concepts (such as up(x), down(x), right(x), left(x), below(x,y) or leftof(x,y)) which are not even included in the language of T_S. For example, the formulas $up(x) \rightarrow \neg down(x)$ or $up(x) \wedge down(y) \rightarrow \neg above(y, x)$ are intuitive truths of grounded model G, but not theorems of T_S.

6 CONCLUSIONS

Grounded models [4] differ from axiomatic theories in establishing explicit connections between language and reality that are learned through language games [9, 12]. These connections, which we call categorizers, give grounded meanings to symbols [1] by linking them to the portion of reality they refer to.

In this paper, we have explained how categories and categorizers are constructed by visually grounded autonomous agents using some conceptualization mechanisms and language games proposed in [9].

We have considered then the process of truth evaluation, proposed a language game that can be used simulating the generation of logical categories, and showed how logical categories and categorizers allow the construction and evaluation of generic concepts of the same complexity as free quantifier first order logic formulas.

Finally, we have described some experiments in which a couple of visually grounded autonomous agents construct a grounded model that can be used for spatial reasoning, and we have explained how grounded models can be used for intuitive reasoning.

ACKNOWLEDGEMENTS

I would like to thank Luc Steels for many interesting conversations on the topics of the origins of language and grounded representations.

REFERENCES

[1] S. Harnard, 'The symbol grounding problem', *Physica D*, **42**, 335–346, (1990).

[2] J. McCarthy, *Formalizing Common Sense. Papers by John McCarthy*, Ablex Publishing Corporation, New York, 1990.

[3] J. Sierra-Santibáñez, 'Grounded models', in *Working Notes of the AAAI–2001 Spring Symposium on Learning Grounded Representations*, pp. 69–74, (2001).

[4] J. Sierra-Santibáñez, 'Grounded models as a basis for intuitive reasoning', in *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-01*, pp. 401–406, (2001).

[5] J. Sierra-Santibáñez, 'Grounded models as a basis for intuitive reasoning: The origins of logical categories', in *Papers from the AAAI–2001 Fall Symposium on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*, pp. 101–108, (2001).

[6] L. Steels, 'Perceptually grounded meaning creation', in *Proceedings of the International Conference on Multi-Agent Systems*, (1996).

[7] L. Steels, 'The synthetic modeling of language origins', *Evolution of Communication*, **1(1)**, 1–35, (1997).

[8] L. Steels, 'The origins of syntax in visually grounded agents', *Artificial Intelligence*, **103**, 1–24, (1998).

[9] L. Steels, *The Talking Heads Experiment. Volume 1. Words and Meanings*, Special Pre-edition for LABORATORIUM, Antwerpen, 1999.

[10] L. Steels, 'The emergence of grammar in communicating autonomous robotic agents', in *Proceedings of the European Conference on Artificial Intelligence, ECAI-00. Invited talk.*, pp. 764–769, (2000).

[11] L. Steels, 'The role of language in learning grounded representations', in *Working Notes of the AAAI–2001 Spring Symposium on Learning Grounded Representations*, pp. 80–85, (2001).

[12] L. Wittgenstein, *Philosophical Investigations*, Macmillan, NY, 1953.

[5] The linear constraints describing the behavior of categorizers are considered now as predicates that are true for those object tuples which satisfy them.