Grounded Models as a Basis for Intuitive Reasoning: the Origins of Logical Categories

Josefina Sierra-Santibáñez
Escuela Técnica Superior de Informática
Universidad Autónoma de Madrid
28049 Madrid, Spain
Josefina.Sierra@ii.uam.es

Abstract

Grounded models (Sierra 2001b) differ from axiomatic theories in establishing explicit connections between language and reality that are learned through language games (Wittgenstein 1953). This paper describes how grounded models are constructed by autonomous agents as a side effect of their activity playing different types of language games (Steels 1999), and explains how they can be used for intuitive reasoning. It proposes a particular language game which 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.

Introduction

In (Sierra 2001b), we introduced grounded models and compared them to axiomatic theories of mathematics. This paper describes how grounded models are constructed by autonomous agents as a side effect of their activity playing different types of language games (Steels 1999). It builds up on previous work on the relation between language acquisition and conceptualization in visually grounded robotic agents (Steels 2001), by addressing the issue of the acquisition of logical categories, and proposing a particular language game (called the *evaluation game*) which 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 (Piaget 1985), because they allow the generation of structured units of meaning, and they set the basis for intuitive reasoning at the level of propositional logic.

Grounded models are not proposed as substitutes for axiomatic theories. On the contrary, they provide an explanation for our ability to come up with some of these theories by intuitive reasoning. They explain, for example, why we accept certain axioms as intuitively true without further argumentation. Grounded models should rather be considered as precursors of axiomatic theories, since these theories require considerable linguistic competence for their formation.

The rest of the paper is organized as follows. First, we define the concepts of sensory channel, category and categorizer (Steels 1996), 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 by combining logical and perceptually grounded categories. Next, we describe how grounded models are constructed as a side effect of the agents activity playing different types of language games, and present some experiments in which a couple of agents learn a grounded model that can be used for spatial reasoning. Finally, we introduce the notion of intuitive reasoning and show how grounded models can be used for it.

Perceptually Grounded Categories

We assume a scenario similar to the physical setting of *The Talking Heads Experiment* (Steels 1999), i.e., a set of robotic "talking heads" playing language games with each other about scenes perceived through their cameras on a white board in front of them. Figure 1 shows a typical configuration of the white board with several geometric figures pasted on it.

We consider now the first steps of a language game, which are intended to conceptualize the perceptual information obtained by the agents after looking at same area of the white board.

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 color, horizontal or vertical position. In the experiments described in this paper, we assume that there are only two primitive sensory channels.

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

The values returned by the sensory channels H and V are scaled so that its range is the interval (0.0 1.0). Consider

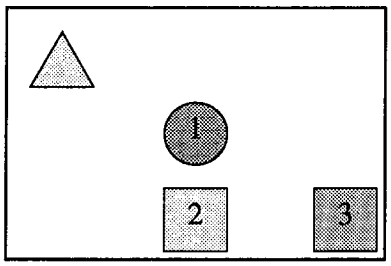

Figure 1: Each object in the scene is characterized by its values on two primitive sensory channels: H and V.

the three objects numbered in figure 1, object 1 has the values H(O1)=0.5, V(O1)=0.5, object 2 the values H(O2)=0.5, V(O2)=0.2, and object 3 the values H(O3)=0.8, V(O3)=0.2.

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 0.3 when it is applied to the pair of objects (O1,O2), i.e., VD(O1,O2)=0.3. 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 $\{0, 1\}$.

Categories

The data returned by 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 values of a particular sensory channel into regions and assigning a *category* to each region (Steels 1996). 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 characterized 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 could 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* (Steels 1996).

Following the notation introduced in (Steels 1999), 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].

Categorizers

At the same time the agents build categories to conceptualize perceptual information, they construct cognitive procedures that allow them to check whether these categories hold or not for a given tuple of objects. A *categorizer* is a cognitive procedure capable of determining whether a category applies or not. For example, the behavior of the categorizer for category [V 0.0 0.5](o) can be described by a function that takes the value 1 if $0.0 < \mathrm{V}(o) < 0.5$, and 0 otherwise.

Categorizers give grounded meanings to symbols by establishing explicit connections between internal representations (categories) and reality (perceptual input as processed by sensory channels) (Steels and Vogt 1997). These connections are learned through language games (Wittgenstein 1953), and allow agents to check whether a category is true or not for a given tuple of segmented 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 $[\mathrm{CAT}]^C(\vec{x})$ to refer to the categorizer which is capable of recognizing whether category $[\mathrm{CAT}](\vec{x})$ holds or not for a given tuple of objects. It is important to realize however that categorizers are cognitive procedures, not linear constraints or implemented functions. We only use linear constraints to model their behavior. We do not assume, therefore, that natural agents do learn such constraints, but that they build them in their perceptual and cognitive systems.

The linear constraints below describe the behavior of the categorizers associated with the categories *up, down, above,* and *below.*

$$[\mathrm{V}\ 0.5\ 1.0]^C(x) \equiv \quad 0.5 < \mathrm{V}(x) \wedge \mathrm{V}(x) < 1.0$$

$$[\mathrm{V}\ 0.0\ 0.5]^C(x) \equiv \quad 0.0 < \mathrm{V}(x) \wedge \mathrm{V}(x) < 0.5$$

$$[\mathrm{VD}\ 0.0\ 1.0]^C(x,y) \equiv \quad 0.0 < \mathrm{VD}(x,y) \wedge \mathrm{VD}(x,y) < 1.0$$

$$[\mathrm{VD}\ \text{-}1.0\ 0.0]^C(x,y) \equiv -1.0 < \mathrm{VD}(x,y) \wedge \mathrm{VD}(x,y) < 0.0$$

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 process of truth evaluation. Logical categories are important for a number of reasons: (1) they allow the generation of an infinite set of concepts, which corresponds to the set of free quantifier first order formulas that can be constructed from perceptually grounded

categories; (2) they set the basis for intuitive reasoning at the level of propositional logic.

Evaluation Channel

The *evaluation channel* (denoted by E) is capable of observing the internal state of the agent. It is applied to a category tuple and an object tuple, and returns a tuple of Boolean values resulting from evaluating each category on the object tuple. It evaluates categories by finding their categorizers, applying them to a given object tuple, and observing their output.

Let $\vec{c} = (c_1, \ldots, c_n)$ be a category tuple and $\vec{o} = (o_1, \ldots, o_m)$ an object tuple, the result of applying the evaluation channel E to $\vec{c}$ and $\vec{o}$ is $E(\vec{c}, \vec{o}) = (v_1, \ldots, v_n)$, 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. That is, $[V\ 0.0\ 0.5]^C(O1) = 0$ and $[H\ 0.5\ 1.0]^C(O1) = 0$.

Logical Categories and Concepts

The range of the evaluation channel is infinite, since it consists of all the Boolean tuples of any arity. In the experiments described in this paper, we will focus on unary and binary tuples only[1]. The range of this channel for unary tuples of categories is the set {1,0}, where 1 and 0 correspond to the logical categories *true* and *false*, respectively. The range of the evaluation channel for category tuples of length two is the set {(0,0),(0,1),(1,0),(1,1)}. If we assign a logical category to each element of this set, we obtain one of the categories used in classical logic (conjunction), and several conjunctions used in natural language, such as neither (0,0), but (1,0), or although (0,1) [2]. If we consider however subsets of the range of the evaluation channel (instead of individual elements) we obtain the meanings of all the connectives used in propositional logic, i.e., negation, conjunction, disjunction, implication and equivalence.

For example, we say that 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 same can be said for the sentence $c_1 \leftrightarrow c_2$ and the set of Boolean pairs {(1,1),(0,0)}.

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 objects from others in a given context. The representation of logical categories, such as conjunction, disjunction or implication, as sets of Boolean tuples is in fact equivalent to the *truth tables* used in classical logic for describing the semantics of propositional connectives. Trivial subsets such as the empty set, {(1),(0)} or {(1,1),(1,0), (0,1),(0,0)} are not associated with categories, because they are not informative. Subsets which combine unary and binary Boolean tuples, such as {(1,1),(0)} are not considered either.

The notation used for describing logical categories consists of the symbol E (for evaluation channel) followed by an schematic representation of the set of Boolean tuples for which each category holds. Thus, [E 1](c) corresponds to true(c), because it is true if the result of applying the evaluation channel to c belongs to the set of Boolean tuples {(1)}. [E 0](c) represents the logical category false(c), or the logical connective not(c), because it is true if the result of applying the evaluation channel to c belongs to the set {(0)}. The logical connective or(c1,c2) (*disjunction*) is represented by the logical category [E 11-10-01](x,y), which is true if the result of applying the evaluation channel to the pair of categories (c1,c2) belongs to the set {(1,1),(1,0),(0,1)}.

A key aspect of logical categories is that they describe properties of categories, as opposed to perceptually grounded categories which describe properties of objects. It is therefore natural to apply logical categories to perceptually grounded categories to construct structured units of meaning, which we call concepts. 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., not(c)) to the category [V 0.0 0.5](x) (i.e., down(x)). The concept [E 11-10-01]([[V 0.5 1.0](x),[H 0.5 1.0](x)) (i.e., or(up(x),right(x))) can be constructed similarly by applying the logical category or(c1,c2) to the categories up(x) and right(x).

If we consider perceptually grounded categories as atomic concepts, we can observe that: (1) logical categories cannot only be applied to perceptually grounded categories but to all sorts of concepts; and (2) the result of applying a logical category to a single concept or a pair of concepts is again a new concept. Logical categories allow generating then an infinite set of concepts, which corresponds to the set of free quantifier first order formulas that can be constructed from perceptually grounded categories. The notion of *concept* as structure of meaning constructed by an agent can then be defined inductively 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.

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 is a logical category, S_l is the subset of the range of the evaluation channel for which l holds, $\vec{c}$ is a tuple of concepts, and $\vec{o}$ is an object tuple. That is, by functions which take the value 1 (true) if the result of evaluating the tuple of concepts to which the logical category is applied belongs to the subset of the range

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

[2] The association between Boolean pairs and natural language connectives is approximate: *neither* is used when the two sentences that follow it are negative; *but* is usually preceded by an affirmative sentence and followed by a negative one; and *although* is sometimes preceded by a negative sentence and followed by an affirmative one.

of the evaluation channel for which the category holds, and 0 (false) otherwise.

The following constraints describe the behavior of the categorizers for the logical connectives *not, and, or, if* (implication) and *iff*. The categorizer which is capable of checking whether category [CAT]($\vec{c}$) holds is denoted by [CAT]$^C(\vec{c})$. The variables c, $c1$ and $c2$ range over concepts. The variable $\vec{o}$ ranges over object tuples.

$$[\text{E } 0]^C(c) \equiv E(c, \vec{o}) \in \{(0)\}$$

$$[\text{E } 11]^C(c1, c2) \equiv E((c1, c2), \vec{o}) \in \{(1, 1)\}$$

$$[\text{E 11-10-01}]^C(c1, c2) \equiv E((c1, c2), \vec{o}) \in \{(1, 1), (1, 0), (0, 1)\}$$

$$[\text{E 11-01-00}]^C(c1, c2) \equiv E((c1, c2), \vec{o}) \in \{(1, 1), (0, 1), (0, 0)\}$$

$$[\text{E 11-00}]^C(c1, c2) \equiv E((c1, c2), \vec{o}) \in \{(1, 1), (0, 0)\}$$

Because the categorizers of logical categories use the evaluation channel, this channel can be naturally extended to evaluate generic concepts (i.e., free quantifier first order formulas) recursively using the categorizers of logical and perceptually grounded categories. Atomic concepts (perceptually grounded categories) are evaluated by applying their categorizers to the object tuple given as input to the evaluation channel. Nonatomic concepts of the form $l(\vec{c})$ are evaluated by applying the categorizer of the logical category l to the result of evaluating the concept tuple $\vec{c}$ on the object tuple given as input to the evaluation channel. 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.

Constructing Grounded Models

In (Sierra 2001a), we defined a *grounded model* 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 as a side effect of the agents activity playing different types of language games. First, the agents play language games of the sort described in the book *The Talking Heads Experiment* (Steels 1999). These games allow them to construct perceptually grounded categories and a shared lexicon for referring to such categories. 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 them.

We describe the evaluation game in this section, and refer the reader to (Steels 1999) for a thorough discussion of the language games used for learning perceptually grounded categories, and their crucial role on the origins of language

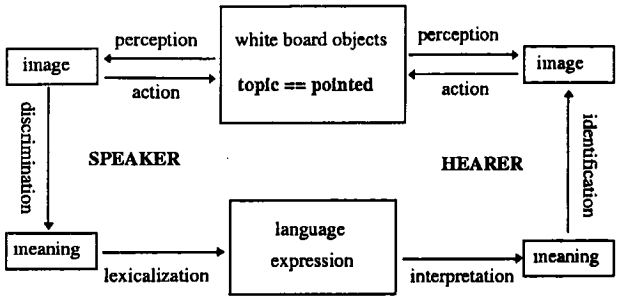

Figure 2: The semiotic square summarizes the cognitive processes involved in a language game.

(Steels 1997). The *evaluation game* is a language game which focuses on the evaluation of concepts and the characterization of sets of objects by means of logical compositions of perceptually grounded categories. The semiotic square (Steels 1999) (figure 2) 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*.

Perception First, the speaker looks at one area of the white board and directs the attention of the hearer to the same area[3]. Both, speaker and hearer, segment their images 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 language game. This is a set of object tuples of the same arity. In the experiments we have carried out so far, the object tuples are of arity one or two. The speaker indicates the hearer the set of object tuples that constitute the context by pointing to them. Different protocols exist for pointing to unary and binary object tuples. Then, the speaker picks up a subset of object tuples from the context which we will call the *topic*. In our experiments, the topic contains a variable number of object tuples which goes from one to three. 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 language game.

For example, in evaluation game 1 (described below) the speaker chooses as context the set of object tuples $\{(O1), (O2), (O3)\}$, where $O1, O2$ and $O3$ represent the numbered objects 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.

[3]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.

That is, a pair of categories $(c1, c2)$ such that their evaluation on the topic is different from their 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. We use the function symbol E to denote the evaluation channel, and give this definition as an extension of previous ones. Let $\vec{c}$ be 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 $E((c1, c2), \text{topic}) \bigcap E((c1, c2), \text{background}) = \emptyset$.

For example, consider the pair of categories ([V 0.0 0.5](x),[H 0.5 1.0](x)) (i.e., (down(x),right(x))). This pair of categories discriminates the topic from the background in evaluation game 1, because the evaluation of the pair of categories on the topic does not intersect with the evaluation of the pair of categories on the background. That is, E(([V 0.0 0.5](x),[H 0.5 1.0](x)),{O1,O3})= {(0,0),(1,1)}, E(([V 0.0 0.5](x),[H 0.5 1.0](x)),{O2})= {(1,0)}, and {(0,0),(1,1)} $\bigcap$ {(1,0)}=$\emptyset$.

If the speaker cannot find a discriminating pair of categories, the game fails. Otherwise, the speaker tries to find a logical category which is associated with the set of Boolean pairs tv 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 tv](c1,c2), together with its categorizer, and adds it to its current grounded model.

Continuing with the example of evaluation game 1, if the speaker does not have a logical category associated with the set of Boolean pairs {(1,1),(0,0)} (resulting from evaluating the discriminating pair of categories ([V 0.0 0.5](x),[H 0.5 1.0](x)) on the topic {(O1),(O3)}), it can create a new logical category of the form [E 11-00](c1,c2) and add it to its 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 for which the concept is true.* For example, in evaluation game 1, the speaker may characterize the topic as the set of object tuples in the context which are on the lower part of the image if and only if they are also on the right part of it. The topic {(O1),(O3)} is the set of object tuples in {(O1),(O2),(O3)} which satisfy the formula *iff(down(x),right(x))*. This formula is internally represented by 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 candidates, the speaker chooses the expression of the association with highest rate. If there are no associations for the logical category C, the speaker may invent a new expression W with probability CR, and add an association of the form $(C, W, 0)$ to its current lexicon. The constant CR is the *creativity* rate of the agent. It determines the probability with which the agent may enrich the language by creating new words. In our experiments CR is set to 0.9 for all agents.

For example, if the speaker has no associations for the logical category [E 11-00](c1,c2) in its current lexicon, it may invent a new expression (e.g., *iff*) and add a new association of the form ([E 11-00](x,y),iff(x,y),0) to its lexicon. In this paper, we assume that internal representations and linguistic expressions have identical structures, so that the agents do not have to worry about learning grammar rules (see (Steels 1998) and (Steels 2000) for some experiments on the origins of syntax in visually grounded agents). This simplification allows us to concentrate on the issue of the origins of logical categories from the point of view of their semantical function.

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, so that the learning process can focus on the construction of a shared lexicon for logical categories.

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 some of the expressions used in 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 AS. The constant AS is the *assimilation* rate of the agent. It indicates the probability with which an agent adopts expressions used by other agents. In our experiments AS is set to 0.9 for all agents.

For example, in evaluation game 1 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 the 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 which satisfy the meaning of *iff(down(x),right(x))* in evaluation game 1 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 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, respectively. If the game fails, only the rates of the associations used by speaker and hearer are decremented, and a repair process similar to that described for the previous step takes place. In our experiments, the amount I by which rates are modified is set to 0.1.

We summarize below the main steps of evaluation game 1, which has been used as example in our description of the evaluation game.

```
Game number:  1
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.
```

Experiments

We have run some experiments in order to see whether a couple of 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 (Steels 1999). These games allow them to learn perceptually grounded categories and a shared lexicon for referring to those categories. Figure 3 shows the evolution of the lexical coherence in a series of 500 games. The *lexical coherence* (Steels 1999) is a measure of the similarity of the agents' lexicons. It can be observed that they reach total lexical coherence after 250 games. The spatial categories and lexicon learned after 500 games are shown in table 1.

In the second experiment, the agents play evaluation games. These games are played after a series of 500 language games, during which the agents learn a shared lexicon for perceptually grounded categories that they needed for playing evaluation games. Figure 4 shows the evolution of the lexical coherence in a series of 10000 evaluation games. First, the agents learn unary categories, such as true and false. This is done very quickly. Then, they learn binary categories, such as conjunction, disjunction, implication or equivalence. This takes longer, about 3000 games to reach a lexical coherence of 0.97. In addition to logical connectives, such as *not, and, or, if* or *iff*, the agents learn other categories which correspond to natural language conjunctions, such as but, neither, however and others. The logical categories and lexicon learned after 10000 games are shown in table 2.

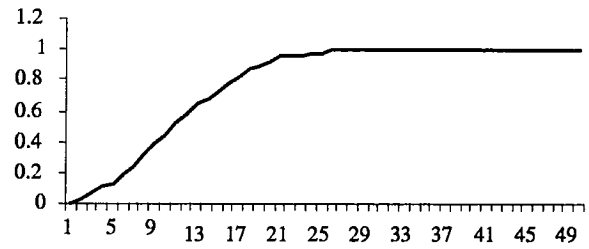

Figure 3: This graph shows the evolution of the lexical coherence in a series of 500 language games. It can be observed that the agents reach total lexical coherence after 250 games.

Table 1. Lexicon: spatial categories.

Expression	Category	English	Rate
seba	[VP 0.0 0.5]	down(x)	1
bi	[VP 0.5 1.0]	up(x)	1
dilebo	[HP 0.0 0.5]	left(x)	1
na	[HP 0.5 1.0]	right(x)	1
sule	[VD -1.0 0.0]	below(x,y)	1
ba	[VD 0.0 1.0]	above(x,y)	1
belebubi	[HD -1.0 0.0]	leftof(x,y)	1
le	[HD 0.0 1.0]	rightof(x,y)	1

Table 2. Lexicon: logical categories.

Expression	Category	English	Rate
lebasu	[E 0]	not(x)	1
nadebabu	[E 1]	true(x)	1
sosulu	[E 11]	and(x,y)	1
deloda	[E 10]		1
nodasosu	[E 01]		1
lo	[E 00]		1
benusu	[E 11-10]		1
si	[E 11-01]		1
luse	[E 11-00]	iff(x,y)	1
babesu	[E 10-01]	xor(x,y)	1
nidi	[E 10-00]		1
nu	[E 01-00]		1
di	[E 11-10-01]	or(x,y)	1
ne	[E 11-10-00]		1
sede	[E 11-01-00]	if(x,y)	1
sabeleli	[E 10-01-00]		1

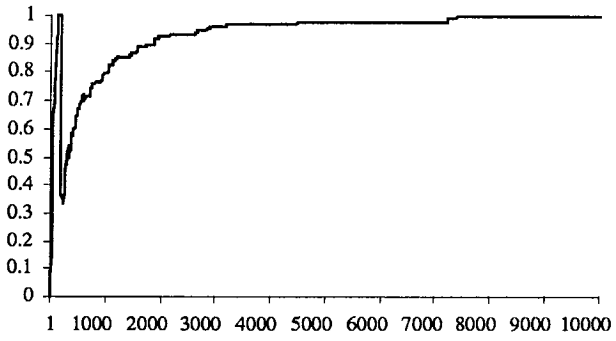

Figure 4: This graph shows the evolution of the lexical coherence in a series of 10000 evaluation games. First, the agents learn unary connectives ($\neg$), reaching total coherence very soon. Afterwards, they start learning binary connectives ($\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$), reaching a lexical coherence of 0.97 after 3000 games.

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 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, since the categorizer of $above(y,x)$ returns false whenever the categorizer of $above(x,y)$ returns true.

Intuitive reasoning can then be used 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 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 nonlogical axioms is as follows.

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

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

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

$$R(x,y) \rightarrow \neg R(y,x) \quad (4)$$

$$R(x,y) \wedge R(y,z) \rightarrow R(x,z) \quad (5)$$

$$R(x,y) \wedge R(x,z) \wedge \neg A(y,z) \wedge \neg A(z,y) \wedge y \neq z \rightarrow R(y,z) \vee R(z,y) \quad (6)$$

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., 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 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 in grounded model G; (2) replacing every instance of VPOS(x), VPOS(y) and VPOS(z) by x, y and z in the expression resulting from step 1; and (3) computing the disjunctive normal form of the result of 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 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 grounded model G by constraint satisfaction as well. That is, the couple of agents used in the previous experiments can discover that the relations *above* and *rightof* are *antisymmetric, transitive* and *total orders on objects located in the same horizontal or vertical positions* 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 (McCarthy 1959), (McCarthy 1990).

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 property of *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 (Shoenfield 1967)) of every intuitive truth of the grounded model[4].

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 $up(x) \wedge \neg down(x)$ or $up(x) \wedge down(y) \rightarrow \neg above(y, x)$ are intuitive truths of grounded model G, but are not theorems of T_S.

Conclusions

Grounded models (Sierra 2001b) differ from axiomatic theories in establishing explicit connections between language and reality that are learned through language games (Wittgenstein 1953). These connections, which we call categorizers, give grounded meanings to symbols by linking them to the portion of reality they refer to.

In this paper, we have explained how categories (i.e., symbolic representations) and categorizers (grounded meanings) can be constructed by autonomous agents connected to their environment through sensors and actuators using some conceptualization mechanisms and language games proposed in (Steels 1999).

We have then considered 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 formulas.

Finally, we have described some experiments in which a couple of visually grounded 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.

[4]The linear constraints describing the behavior of the categorizers are considered, in this case, as predicates which are true for those object tuples that satisfy them.

Acknowledgments

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

References

McCarthy, J. 1959. Programs with Common Sense. In Mechanization of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory, 77–84.

McCarthy, J. 1990 *Formalizing Common Sense. Papers by John McCarthy.* Edited by Vladimir Lifschitz. Ablex Publishing Corporation, New Jersey.

Piaget, J. 1985. *The Equilibration of Cognitive Structures: the Central Problem of Intellectual Development.* University of Chicago Press, Chicago.

Shoenfield, J R. 1967. *Mathematical Logic.* Addison-Wesley Publishing Company.

Sierra-Santibáñez, J. 2001. Grounded Models. In Working Notes of the AAAI–2001 Spring Symposium on Learning Grounded Representations, 69–74. Stanford University, Stanford, California.

Sierra-Santibáñez, J. 2001. Grounded Models as a Basis for Intuitive Reasoning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence.* Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence, Inc.

Steels, L. 1996. Perceptually Grounded Meaning Creation. In *Proceedings of the International Conference on Multi-Agent Systems.* Menlo Park, Calif.: AAAI Press.

Steels, L. 1997. The Synthetic Modeling of Language Origins. *Evolution of Communication* 1(1):1–35.

Steels, L. 1998. The Origins of Syntax in Visually Grounded Agents. *Artificial Intelligence*, 103:1–24.

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

Steels, L. 2000. The Emergence of Grammar in Communicating Autonomous Robotic Agents. In *Proceedings of the European Conference on Artificial Intelligence 2000.* Horn, W. (ed.), IOS Publishing, Amsterdam.

Steels, L. 2001. The Role of Language in Learning Grounded Representations. In Working Notes of the AAAI–2001 Spring Symposium on Learning Grounded Representations, 80–85. Stanford University, Stanford, California.

Steels, L., and Vogt, P. 1997. Grounding Adaptive Language Games in Robotic Agents. In *Proceedings of the European Conference on Artificial Life 97.* The MIT Press, Cambridge Ma.

Wittgenstein, L. 1953. *Philosophical Investigations.* Macmillan, New York.