Symbol Grounding Revisited

Erich Prem
The Austrian Research Institute for Artificial Intelligence*
Schottengasse 3, A-1010 Wien, Austria
Tel. +43 1 5336112, FAX +43 1 5320652
erich@ai.univie.ac.at

Abstract

Symbol Grounding has originated in the domain of cognitive connectionism as an approach to a model of language acquisition.It has, however, transcended this restricted domain and its most prominent proponents now regard it as a technique which is of general importance to the connectionist-symbolic debate, especially for their integration into a common framework. This paper revisits the claims made by symbol grounders and summarizes different models which have been presented to date. We try to answer the question what Symbol Grounding really is all about and point to its parallels with classical logic and reasoning which have not received sufficient attention so far.

1 The Origin of Symbol Grounding

1.1 Overview

This paper is intended to evaluate recent developments within a specific subarea of symbolic-connectionist integration. This subarea consists of several connectionist approaches to the problem of grounding meaning in perception through somehow connecting sensory experience to symbolic forms. It contains a description and comparison of today's approaches as well as a fundamental analysis of Symbol Grounding (SG). The aim of this paper is *not,* however, to present yet another model of connectionist SG and the reasons for this should become clear soon.

Section 1 provides an overview of the history of Symbol Grounding and introduces the reader to areas of research within which Symbol Grounding is pursued. Section 2 presents a critical overview of more technical approaches to this problem, compares them and presents a generic SG architecture. Implications of these models which concern the question what SG actually is are presented in section 3. One of these consequences is a strong connection to symbolic (and transcendental) logic which we try to motivate in section 4.

1.2 The Symbol Grounding Problem

Based on a fundamental criticism of symbolic models, Stevan Harnad introduces "The Symbol Grounding Problem" in his 1990 paper. This nowadays prominent article presents the problem of how formal symbol systems can acquire a semantics which is not based on other symbol structures but on a system's own sensory experience. The problem begins with the observed fact that conventional rule based AI systems consist of specific representational structures (symbols) which only possess meaning with respect to other symbols in the program or because a program designer assigns meaning to them. In this last case Harnad would say they are "parasitic" on the meaning in our head. Such symbol structures have no meaning in themselves, i.e. no intrinsic, only derived meaning.

*__WWW-URL:__ `http://www.ai/oefai/nn/nngroup.html`

Unfortunately, this fact bears unpleasant consequences for such systems. They have been pointed out by Symbol Grounders along three dimensions, i.e. from a technical, a cognitive science and a linguistic perspective. The first is that system designers sometimes overestimate the actual meaning of such symbols (cf. [Harnad 90a]), which in turn results in system failures. Additionally, the total lack of any inherent semantics of such systems makes them implausible as any kind of model for cognitive phenomena. Thirdly, although the program consists of linguistic elements, namely the symbols themselves, these symbols are completely different in their characteristics from words in natural languages. Again, the reason lies in the lack of intrinsic meaning.

1.3 An Alternative Historical Perspective

Before turning to a technical analysis of SG, we shall now place it in a larger context of AI history. Criticism of a traditonal symbolic approach has continually arisen from different directions within science. Four such positions will be briefly examined below. Although the motivations for and subjects of these criticisms are quite different from one another, the symbol and its lack of grounding are continually attacked.

John Searle's 1980 article contains what is probably now the most prominent desctruction of the symbolic approach. Originally, his well-known *Chinese Room Argument* was an attempt to show that AI programs do not have cognitive states and intentionality. His argument is supposed to show that intelligent behavior and the processing of a merely formal program cannot be sufficient for reaching intentionality. This lack of intentionality also implies that the symbols used in such programs are meaningless. Searle himself says:

> [F]ormal symbol manipulations by themselves don't have any intentionality: They are quite meaningless; they aren't even *symbol* manipulations, since the symbols don't symbolize anything. [Searle 80, p.422]

Unfortunately this passage can be easily misunderstood. Indeed, the difference between original and derived intentionality has, especially in the light of some Symbol Grounding endeavours, been reduced to a purely semantic difference between intrinsic and merely interpretative meaning. But Searle's argument certainly goes beyond this restricted criticism. It is concerned with *intentionality*, not with meaning alone. Searle argues that in such simulations no causal properties of the brain, especially not of those brain states which we usually call intentional are considered. All that is taken care of are the merely formal results of these processes. Searle is also convinced that a computer simulation of all biochemical processes would not change his basic argument. Even such a simulation would not causally entail those properties of brain states which make them intentional.

> [N]o purely formal model will ever be by itself sufficient for intentionality, because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal powers except the power, when instantiated, to produce the next stage of the formalism when the machine is running.

Although Searle is quite specific here, his position has been sometimes misinterpreted in a way which is most obvious in SG discussions. The misunderstanding of Searle is about what these relevant "causal" powers are. In a major part of the SG literature the problem seems to reduce to the causal ("grounded") generation of program structures, i.e. to generating the meaning of symbols through the equipment of the computer with a measurement device that interacts with the environment. The goal is to generate symbols which are no longer totally arbitrary in their interpretation. In a very precise sense, the structures of the computer program are, of course, causal then. However, this is not how Searle originally uses the term. In fact, Searle himself has already clearly confirmed that this "transducer" argument is not what is meant by causal powers. Harnad (in his newer papers [Harnad 92]) has accepted Searles position in saying that Searle's argument is "absolutely right." Nevertheless, Searle's negation of the possibility that purely formal structures enable symbols to refer to objects can be regarded as one version of the SG problem.

A completely different attack on symbol structures has been pursued by Paul Smolensky in [Smolensky 88]. The direction of his argument is the explicit formulation of "conscious rules" in classical programs, in which knowledge is identified with concepts that are semantically interpretable (as English words). His nowadays well-known alternative to this approach is the replacement of symbols by "sub-symbols" (i.e. microfeatures), which have a different semantic interpretation compared to the symbolic concepts alone.

The main problem with Smolenkys approach is that he does not consider an immediate grounding of connectionist "sub-symbols" to be necessary. As long as they are in a sense more finely graded than the corresponding symbolic description and as long as they are shared among representational units of a higher level (ensuring distributedness), this is sufficient. Smolensky even points out the fact that in such a view of the problem it is extremely difficult to clearly formulate what the differences to the symbolic approach actually are. This is the case because in most *formal* analyses the difference between symbols and sub-symbols, being coneptually positioned at different levels, tends to disappear. Viewed from a Symbol Grounding point of view, Smolensky's contribution is mainly an attack of the symbol for technical reasons (for their lack of inherent context dependence) and for reasons which arise from the cognitive inplausibility of early AI rule based programs. Grounding plays a role in so far as it seems a natural way to generate these more finely graded, context dependent representations.

representative	*variant of the SG problem*	*AI criticism*
Dreyfus	decontextualization	existential-ontological critique
Searle	"Chinese Room Argument"	no (intrinsic) intentionality
Brooks	physical grounding hypothesis	missing practicality
Smolensky	microfeatures	no intuitive processor
(Harnad)	synt.semant.compositionality	"Hermeneutic Hall of Mirrors"

Figure 1: Variants of the Symbol Grounding Problem.

Again, a completely different account which has arisen in the rather technical domain of robotics has been introduced by Rodney Brooks [Brooks 91]. Having worked within the area of traditional robotics Brooks one day found himself questioning the implicit assumptions about the necessity of symbolic world-models in his robots. Brooks argues that after decades of research all the attempts to build a useful, well-working program which is able to model an aspect of the world in terms of symbolic descriptions and by means of directly measuring the world have been in vain. Neither are such symbolic models in themselves conceivable, nor have there been solutions to the problem as to how one should generate these models from raw sensory data. Brooks' approach instead consists in a more direct interaction of as many system control components with the world and with one another as necessary (called "subsumption architecture").

But Brooks goes even one step further in that this physical interaction is considered a *necessary* condition for intelligent behavior; this is the *Physical Grounding Hypothesis* [Brooks 90]. It is in the light of this rebuttle of purely symbolic abstractions that we have to understand the suggestion to use "the world as its own best model." As a consequence of this proposal Brooks had to develop extremely specific sensors. A good example is the soda can detector built into Toto [Brooks 91]. This special laser and video based sensor allows the system to find out the truth about the proposition "there is a soda can in front of the gripper." It is clear that even this highly specialized sensor can be fooled with respect to our human interpretation of this predicate.

Summarizing, the Brooksian attack is primarily directed towards simplistic assumptions in the development of world models for robots. These assumptions are all the more dangerous for being unspecified. The symbol is partly responsible for this, because its arbitrary interpretability hides the fact that in a final step it needs to refer to properties external to the system. In a more philosophical language one could say that the formal consistency of symbolic systems tends to conceal the necessity of material truth and the transcendental nature of a robot's world model. (We shall return to this fact below.)

Finally, we finish this overview of AI criticism related to Symbol Grounding with Hubert Dreyfus, who bases his line of argumentation on the existential-ontological program of Martin Heidegger [Dreyfus 84]. He argues that the design of conventional AI-programs parallels the ontological and epistemological project of explaining every phenomenon in the world by means of "occurrent" function predicates [Dreyfus 90, p.116], i.e. by concepts and relations of them. The main point for criticizing such an approach is that this view of the world necessarily reduces the phenomena under study to their scientific meaning. This consequently deprives these elements of the world of their inherent significance for humans.

The standard example of this idea is the AI description of a hammer. By means of putting together a number of symbolic descriptions of a hammer it is possible to say how it works, what it looks like, etc. What cannot be said, however, is what a hammer *is*. Its whole nature cannot be composed from "meaningless" value predicates and symbols. At one point in his criticism Dreyfus is surprisingly close to Harnad's formulation of the SG problem:

> Traditional ontology has always sought to understand the everyday world by finding something on the level of the occurent, such as substance, sense data, or representations in transcendental consciousness, that is supposed to be intelligible without reference to anything else, and then sought to show how everything else can be seen to be intelligible because it is built up out of these self-sufficient elements. [Dreyfus 90, p.122]

This last criticism of symbolic approaches to AI is not so much directed against the symbol in its more technical sense of being a means of representing knowledge. It is also not directed against the construction of "knowledge bases" by means of symbols because such a construction would be too difficult. It is, however, the world view which lies at the basis of such approaches that is attacked. It is the view of the world as a sum of decontextualised and therefore (for Dreyfus) meaningless elements which is questioned here.

The table in figure 1 tries to summarize these different AI criticisms, which are considered here as mere variants of the Symbol Grounding Problem.

2 Current Connectionist SG Models

In this section an overview of connectionist models of Symbol Grounding is given. Today's models can be summarized in three categories:

- pure object models
- verb models
- models which try to ground other aspects of natural language (like e.g. Terry Regier's model of grounding spatial terms [Regier 92], which will not be treated here).

2.1 Object models

A group of connectionist proposals addresses the problem of grounding object categories in perception. Typically, the models possess two types of input: one for (simulated) sensory data, another one for symbolic descriptions of the data. The networks are usually chosen so that the sensory input is categorized through un- or selfsupervised categorization algorithms like Kohonen networks. The resulting category representations are then associated with some symbolic description of the static input.

A typical and early example is the model of Cottrell et al. [Cottrell et al. 90] where faces of a small group of people (i.e. video images) are mapped onto their corresponding names. The goal is to ground the names in perception. Left part in figure 2 depicts the model. For each modality (faces and names) autoassociative networks exist, which extract the corresponding class features by means of error backpropagation. Each network, names and faces, is trained separately. In a second training phase associations from faces to names and vice versa are trained.

This procedure is very similar to the one employed in Chauvin's cognitive model of language acquisition in children [Chauvin 89]. The problem of assigning names to objects is taken to be a pattern recognition problem. Again, compressed representations of input patterns are generated through autoassociative error backpropagation. The naming phase (associating names and pictures) is separated for psychological plausibility. This separation of autoassociative invariant construction for the recognition of objects and the naming of these categories has now become the standard paradigm for this kind of cognitively motivated SG research. Philippe G. Schyns discusses the psychological motivations for this paradigm and presents his own model which is again of a very similar architecture [Schyns 91]. He calls this principle "label-concept" independence. It should be noted here that current models are extremely simple with respect to their computational capabilities. And although some different architectural principles are employed, the overall system behaviors are to a high degree comparable to each other.

One subtle difference is stressed in G.Dorffner's model of SG [Dorffner 92]. Whereas the above models tend to map similar inputs to similar outputs (labels), his model is capable of truly implementing the "tertium non datur" principle. Through the introduction of an intermediate categorization construct (C-layers) his model achieves that names of object categories (external embodiments of symbols) either refer to objects or do not do so at all.

Dorffner's model is probably the clearest implementation of the classic semiotic triangle. The sensory input to the system corresponds to the symbol's extension, the connection (C-layer) together with the generated category is the equivalent of the symbol's intension, i.e. it represents what Frege has called its "Sinn" [Frege].

2.2 Verb models

Connecionist models for grounding dynamic aspects of language in sensory experience do not exhibit such strong similarites like the ones presented in the previous section. This is mainly due to the fact that the processing of input sequences is still a novel research issue and comparatively difficult. Models for dynamic aspects of language have e.g. been presented by Cottrell [Cottrell et al. 90] or by Nenov and Dyer [Nenov & Dyer 88]. Cottrell's model is schematically depicted in figure 2.

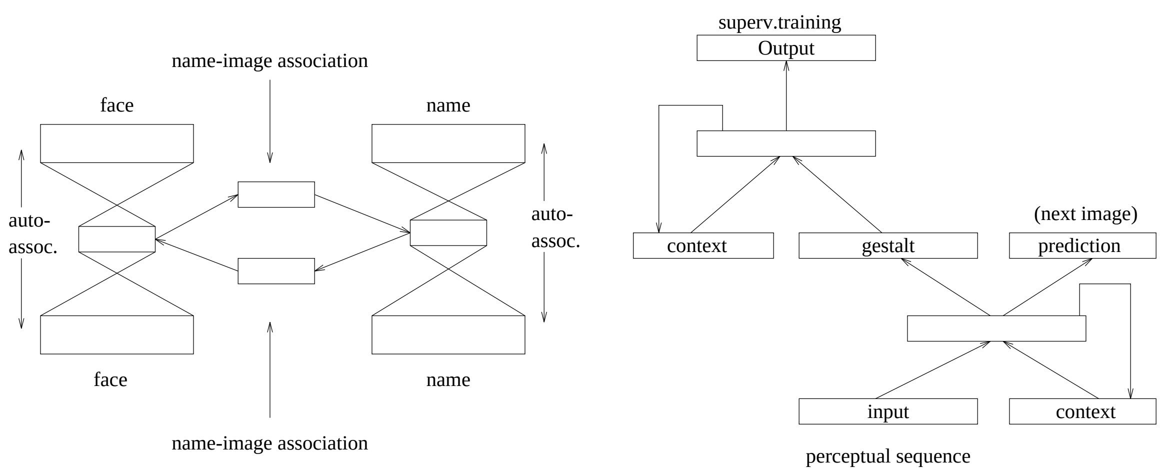

Figure 2: SG (left) and *movie description* (righht) networks (Cottrell et al.)

These models are capable of learning simple sentences like e.g. "ball bounces right." Their main difference to the object models is twofold: (i) The model does not only connect sensory inputs to a symbol, but it also learns to produce a sequence of such symbols. (ii) The input consists in some sort of time sequence, not only of a static input image.

2.3 Summary

If we now compare the different motivations for these different models of SG, we realize that there are two main goals:

- development of a model of aspects of natural language
- enriching formal symbol structures with some kind of sensor-based information

The main assumptions with which these goals are pursued are:

- Symbols are entities similar to words for sensory qualities.
- Semantics grows from a historical coupling of signs to sensory experiences.

The general picture arising from this background of SG models is depicted in figure 3.

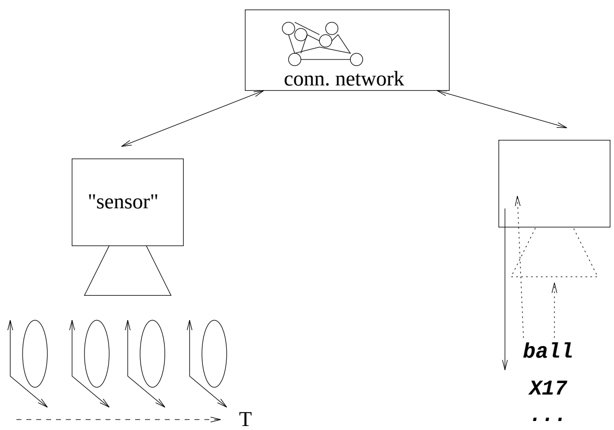

Figure 3: A generic SG architecture: Sensory data (or a time series of sensory data is mapped onto some symbol space through the employment of connectionist techniques.

3 Implications: what SG is and is not

The above figure shows that the program of SG as it is undertaken today consists in automatically developing structures in some sensor space (sometimes with an additional time dimension) which map onto symbols that are taught to the system. The hope is that after training, the SG system is capable of “uttering” the right symbols in a given situation.

3.1 SG and the Origin of Words

For those interested in linguistic research, this program is exactly what W.v.O. Quine [Quine 60] has called “radical translation.” The system which we are training is in the same situation as the linguist who tries to develop a translation for a totally unknown language. In Quine’s original example, the native utters “Gavagai!” whenever we would say that we see a rabbit running across the field. Quine points to the interesting fact that we can never be quite sure about what the native’s exclamation actually refers to. It could be the whole rabbit, the fact that the rabbit is running here and now, a sudden materialisation of ideal rabbitness, a part of the rabbit, etc.

Our connectionist SG system faces exactly the same problem: It must find out what the invariants in the sensory image are that make us utter specific symbols in specific situations. The central assumption is that the relevant information for our utterance *actually is in the sensory data.* Of course, this is a somewhat strange assumption in the area of natural languages. We strongly believe that our own linguistic utterances cannot be fully accounted for by sensory stimuli.

Our goals, emotions, preferences are equally involved in generating specific words. In fact this problem has been intensively discussed in the philosophical literature, e.g. by L. Wittgenstein or M. Heidegger. Even if the whole information for generating the right response would actually lie in raw sensory data, the problem of how to find the right invariants would still be enourmous.

Of course, objections like this to the program of SG are well known to linguistically oriented connectionists. The general response (cf. [Dorffner 93]) is that the reading of language as a merely descriptive activity is only a first crude approximation to what language really is. Other aspects of language, which have been characterized above or which are among laughing, quarreling, praying, telling jokes, etc.,[1] will be considered in future models of natural language. This program implies that these other aspects of language can be added once the primarily descriptive character of symbols (words) has been sufficiently clarified. This kind of research program does also imply something about the origin of language. It was again Stevan Harnad in [Harnad 93] who has made these implications explicit. He argues that

> [...W]ords originated as the names of perceptual categories. [...] Words originate by ostensive experience with concrete sensory categories. This "grounds" them psychophysically. They can enter into descriptions of higher categories, including abstract ones.

This last argument of Harnad is supposed to overcome the problem of "object names", which do not have any referent in the real world (like e.g. "unicorn"). In order to motivate an alternative view of SG, let me exaggerate this picture by using a polemic example of the origin of words. Consider the first few humans (i.e. apes) sitting around their fire-place (if developed yet). Suddenly one of our apes has the idea of associating a name with one of his surrounding objects and pointing to it or holding it up in his hand he shouts "banana!". The others (not stupid) immediately or after a few repetitions for the first time recognize bananas as bananas...

What makes this picture so preposterous is that the utterance of the first words is (i) deprived of all human purposes and (ii) limited to a strange way of descriptive talk. Therefore, a different picture of SG will be drawn in the next section.

3.2 SG and Descriptive Models

In SG models the only purpose of symbols is their reference to external objects, resp. to the sensory data generated by them. The linguistic signs which refer to these objects are a means to subdue these objects to a scientific investigation of their properties. The resulting objects have been decontextualized and are therefore easily managable so as to e.g. combine them into a knowledge based system. A theory of SG which largely adopts such a conception can only result in a specifically scientific model (in the sense of fig. 4) which is designed to reflect nature for epistemic purposes. I would like to argue that this process is equivalent to the automated development of a formal model of a natural system. ("Natural" because we are still talking about *grounded* systems, which are connected to the world through measurement devices.)

This interpretation of SG is not changed by the observation that SG models construct some kind of *subjective* model of their environment—based on the inherently statistical nature of neural network algorithms, because the system itself still generates a symbolic model of this environment, be it subjective or not. It is not argued here that all such models are comparable to each other, nor do we want to say that SG models implement constructivism. What they do, however, is to bring a manifoldness of sensory data into the unity of concepts and *express* these concepts by means of arbitrarily chosen forms.

It is essentially this feature which can be found in any mathematical (or formal) model of a natural system. Some kind of encoding procedure maps "objects" of the environment (usually through employing meters) onto symbols in a formalism. Implications in the formalism are then used to predict what natural law does to the natural system. This situation is depicted in figure 4.

(In fact, SG systems do a bit more than just labeling inputs. They even implement basic descriptive or predictive rules. However, this must be described elsewhere [Prem 94a].) In the next section the relations of SG to some aspects of formal logic are highlighted.

[1] Wittgenstein's famous examples.

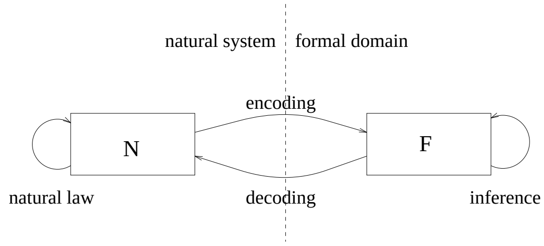

Figure 4: The modeling relation.

4 Symbol Grounding and Logic

At a first glance it may seem strange to construct a connection between SG and logic. This impression arises from the fact that logic nowadays nearly always means *formal* logic. However, if we go back to C.S. Peirce, we find a definition of logic as the science of the conditions which enable signs to refer. If one assumes now that symbols already refer to objects in the world, this science naturally reduces to syllogistic and set theoretic considerations. If we ask instead, how it comes that a given sign means this or that, we are engaged in what is traditionally termed *transcendental* logic. The following parallels between SG and logic are based on the assumption that in the original Arisotelian (and Platonic) tradition, reasoning is a process which is based on the structure of a hierarchical concept tree (arbor porphyriana), however with the intention of deriving *new* knowledge, not only that which has already been implicitely specified in the premises (i.e. which is not purely decuctive). (For a more detailed analysis cf. [Zeidler 92, p.125].)

According to Peirce, there are three types of signs: *Icons,* which are connected to their object because of some inner similarity. *Indexes,* which are "naturally" connected (i.e. connected through natural law, like e.g. a thermometer is a sign for temperature). Finally, *symbols* represent their objects only through "the idea of the symbol user." In a more fashionable terminology this means that symbols are arbitrary. In SG models this last characterization reduces to the requirement that any class of sensory data can be mapped onto any formal symbol. As we have seen, however, most SG models map similar inputs to similar outputs without any clear boundary. I.e. in Peircean terminology they would actually become "icons" for some "inner" similarity connects the sign and its object. Due to the continuous character of connectionst mappings, SG must necessarily exhibit such iconic features in grounding signs.

Let us now concentrate on how the sign becomes connected to its object. In the case of an icon, some "inner" quality is responsible for this. An "inner" property is one by which we identify the object as that kind of object. This similarity of representation and represented negotiates between the object and its sign; both possess the common property P. In more formal terms, let M be the set of object states, Q be the set of the sign states. Then it is the fact that both, M and Q, belong to a set P which makes them stand for each other. (M is P. Q is P. $\rightarrow$ M is Q.) This negotiation takes exactly the form of a logical inference. The position of P, the *terminus medius* corresponds to Aristotle's second figure [Aristotle, III 26b].[2] Peirce, also following [Aristotle, III 69a] has called this figure *abduction.* (That this process is the place in reasoning, where the "real" inventions are made, has also been claimed by [Simon 65, p.186]. Abduction was also called retroduction by Peirce.) In SG architectures this abductive basis P usually amounts to timely coincidence of the sign and its significatum.

The index is a bit more problematic, mainly because we are not so much used to working with it. According to Peirce, the index represents its object through natural correspondence. As opposed to the icon the object of the sign and the sign itself do not coincide in some "inner" but in an "outer" property. Helpful for understanding what an "outer" property is may be Peirce's hint that the index is also index without anyone interpreting it. Therefore, what is characteristic about it must lie in the object itself. I.e. the class of objects is partitioned into those refered to by the

[2] It is clear, of course, that in a purely *formal* approach, abduction is only possible with a negative conclusion.

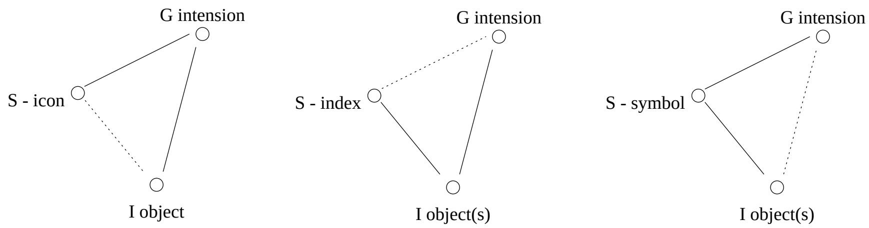

Figure 5: Different views of hierarchical concept trees (arbor porphyriana). Straight lines show relations given, dotted lines represent conclusions. G–general, S–special, I–individual.

sign and those not refered to. Viewed logically this process is similar to induction, which informs about the total set of objects with a specific property. According to Aristotle again, induction "in a certain sense" proofs the *terminus maior* for the *terminus medius* by means of the *terminus minor* [Aristotle, III 68b]. (See also figure 5.)

Finally, the symbol is a general representation of an object, with an arbitrary relation to the sign. This arbitrariness, which is central to all of AI, means that it is *the sign alone* which negotiates its meaning. In other terms, the symbol represents intension and extension, it alone ensures that one is in connection with the other. In a psychological interpretation this twofold representation becomes a picture of (i) a subjective interior world and (ii) an objective outer world. This observation lies at the bottom for viewing symbols as a means of communication. The symbol here is injected between the singular (the object) and the general (the "Sinn"). This is why symbolic negotiation is similar to deduction. Dorffner's insistence on the clear boundary between different symbols is nothing but the consequence of a set theoretic consideration. Objects possess a property or they don't. Symbols refer or don't, tertium non datur.

What can be learned from these parallels is that SG is not so far away from formal logical principles as it may consider itself to be. On the other hand these formal domains indeed have a transcendental basis which is addressed by Symbol Grounding endeavours, though unnoticed. The implications of this discovery go much further than could be described here (see [Prem 94a, Prem 94b] for further details). It is hoped, however, that the explication of these parallels shows the need for further discussing the relation of Symbol Grounding and logic.

5 Conclusions

Many different approaches to the problem of grounding symbols in perception exist. They differ in the concrete algorithms employed, but the overall architectures are highly similar. I have tried to convince the reader that using SG models for the purpose of cognitive linguistics implies an inherently simple view of semantics and, generally, of what the phenomenon of language really amounts to. Instead of such a simplistic view I have proposed here to regard SG as some kind of automated model construction. SG implements the very first part of such models, where symbols are used to formulate descriptive rules about what will happen in the natural environment. I have tried to show that the very idea of grounding is in surprising parallel to classical formal inference methods.

SG thus is a step towards grounding logic in empirical data. This means that the propositions formulated by a SG system are different from purely formal approaches in that they are based on material truth. But this is what was originally intended, was it not ?

6 Acknowledgements

The author wishes to thank G. Dorffner, A. Flexer, and G. Widmer for their comments on the manuscript and R. Trappl for his support of this work. The Austrian Resarch Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Research.

References

Aristotle. *Organon.*

Brooks, R.A. (1990). Elephants Don't Play Chess. In P.Maes (Ed.), *Designing Autonomous Agents.* Cambridge, MA: Bradford/MIT Press, 3–16.

Brooks, R.A., (1991). Intelligence without Representation. In Special Volume: Foundations of Artificial Intelligence, *Artificial Intelligence,* 47 (1–3).

Chauvin, Y. (1989). Toward a Connectionist Model of Symbolic Emergence. In *Proc. of the 11th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates, 580–587.

Cottrell, G.W., Bartell, B., Haupt, C. (1990). Grounding Meaning in Perception. In H. Marburger (Ed.), *Proc. of the German Workshop on AI: GWAI-90.* Berlin: Springer.

Dorffner, G. (1992). A Step Toward Sub-Symbolic Language Models without Linguistic Representations. In R. Reilly & Sharkey N. (Eds.), *Connectionist Approaches to Natural Language Processing, Vol.1.* Hove, East Sussex: Lawrence Erlbaum, Hove.

Dorffner, G., Prem, E. (1993). Connectionism, Symbol Grounding, and Autonomous Agents. In *Proc. of the 15th Ann. Meeting of the Cognitive Science Society* Boulder, CO, 144–148.

Dorffner, G., Prem, E., Trost, H. (1993). Words, Symbols, and Symbol Grounding. To appear in: Proceedings of the 16th Wittgenstein Symposium, Vol.2, Kirchberg am Wechsel/Austria: Austrian Ludwig Wittgenstein Society.

Dreyfus, H. (1984). *Why Computers Can't Think.* Macmillan.

Dreyfus, H. (1990). *Being-in-the-world.* Cambridge, MA: MIT Press, 115–127.

Frege, G. (1892). Über Sinn und Bedeutung. In *Zeitschrift fuer Philosophie und philosophische Kritik.* 100, 25–50.

Harnad, S. (1990a). Lost in the Hermeneutic Hall of Mirrors. In *J. Expt. Theor. Artif. Intell.* 2 (1990), 321–327.

Harnad, S. (1990b). The Symbol Grounding Problem. In *Physica D*, 42 (1990), 335–346.

Harnad, S. (1992). Artificial Life: Synthetic Versus Virtual. In *Artificial Life III* Santa Fe, June 1992.

Harnad S., (1993). The Origin of Words: A Psychophysical Hypothesis. In W.Durham, B.Velichkovsjy (Eds.), Muenster: Nodus Publ.

Nenov, V.I., Dyer, M.G. (1988). DETE: Connectionist/Symbolic Model of Visual and Verbal Association. In B.Kosko (Ed.), *IEEE International Conference On Neural Networks.* San Diego: IEEE, Vol. II, 17–24.

Prem, E. (1994a). *Symbol Grounding.* Die Bedeutung der Verankerung von Symbolen in reichhaltiger sensorischer Erfahrung mittels neuronaler Netzwerke. University of Technology, Vienna: PhD Thesis, forthcoming.

Prem, E. (1994b). *Symbol Grounding and Transcendental Logic.* Presented at the Workshop for Logic and Reasoning with Neural Networks, ICLP'94, Santa Margherita Ligure, Italy; TR-94-20, Austrian Research Inst. for AI, Vienna, Austria.

Quine, W.V.O. (1960). *Word and Object* Cambridge, MA:MIT Press.

Regier, T. (1992). *The Acquisition of Lexical Semantics for Spatial Terms.* University of California, Berkeley, CA: PhD Thesis.

Schyns, P.G. (1991). A Modular Neural Network Model of Concept Acquisition. In *Cognitive Science.* 15(4).

Searle, J.R. (1980). Minds, Brains and Programs. In *Behavioral and Brain Sciences.* 3, 417–457.

Simon, H.A (1965). The Logic of Rational Decision. In *The British Journal for the Philosophy of Science.* 63 (16), 169–186.

Smolensky, P. (1988). On the Proper Treatment of Connectionism. In *Behavioral and Brain Sciences.* 11(88), 1–74.

Zeidler, K.W. (1992). *Grundriß der transzendentalen Logik.* Cuxhaven: Junghans Verlag.