Understanding is discussed with reference to an agent-action-objective model. The formalism developed controls the processes of differentiation and integration underlying understanding. The starting point is a culturally agreed-upon expectancy or prescription that defines a particular form of understanding. Central to an investigation of how scientists model understanding is how they conceive symbols. The collective cognitive imperative in the scientific community states that symbols are abstract codes whose definitions are wholly arbitrary and, hence, can only be understood by those who have a common outlook. Scientists working with the development of computer technology, especially within the field of artificial intelligence, have repeatedly pointed out that they have constructed models of understanding the natural language of significance for cognition-oriented research. Therefore, focus is on classical models. It appears that self-reference is an important constituent component in any model or theory of understanding. (TJH)
The Cult of Understanding

Bernhard Bierschenk
The Cult of Understanding

Bernhard Bierschenk

1986 No. 15

Communications should be sent to:
Cognitive Science Research
Paradisgatan 5
Lund University
S-223 50 Lund, Sweden

Coordinator: Bernhard Bierschenk
Department of Psychology
Abstract

In this article, understanding is discussed with reference to an Agent-action-Objective model. The formalism developed controls the processes of differentiation and integration underlying understanding. The starting point is a culturally agreed-on expectancy or prescription which defines a particular form of understanding.

Central to an investigation of how scientists model understanding is how they conceive symbols. The collective cognitive imperative in the scientific community states that symbols are abstract codes, whose definitions are wholly arbitrary and because of this can only be understood by those who have the same common outlook.

Scientists working with the development of computer technology, especially within the field of artificial intelligence, have repeatedly pointed out that they have constructed models of understanding natural language of significance for cognition-oriented research. Therefore, some by now classical models have been made the foundation of this discussion. Finally, the discussion points at the importance of self-reference as constituent component in any model or theory of understanding.
The ecological approach to studies of human cognition and performance is a trend in both cognitive- and natural science-oriented research. It stresses the indisputable fact that human cognition develops only in cooperation with the environment. Two concepts are basic in ecological studies of cognition, namely, intention and orientation. Genetic epistemology (Piaget (1978), developmental psychology (Werner & Kaplan, 1963), or analytical biology (Sommerhoff, 1950; Wolsky & Wolsky, 1976) have demonstrated that the general cooperation of the organism with its environment requires the teleonomic concept in order for the organism to express an "intended and oriented schematization" (Bierschenk, 1984a). Monod (1971) assumes that all behavior is "teleonomic" inasmuch as behavior is directed towards the environment. Further, there are good reasons to believe that the original language communication has been verbal and not nominal (Pribram, 1981). For example, what is implied by the nominal "understanding"? A starting point for a discussion could be Jaynes' (1976, p 199) description of a scene showing the Assyrian steward-king Hammurabi (about 1750 B.C.) hallucinating judgments from his god Marduk, the city god of Babylon. Jaynes writes:

"Hammurabi listens intently as he stands just below him ("understands")" (p 199).

First of all, a denominalization process has to be initiated, by which the abstractions can be processed to prove structural affinity. The possibility of representing observations this way prerequires the distinguishing of an agent and the object(ive) of his action. Such a schematization into the

Agent - action - Objective

(1)

schema imposes a constitutive function on the context. The formula (') will be used in an attempt to place the nominal understanding in its proper ecological context to obtain the observation, assertion, or predication to be communicated. The quotation from Jaynes indicates a West Germanic prefix construction
to the verb 'stand'. Its transitive property (from a cognitive point of view) is marked by means of the prefix 'under', meaning protection. The verb itself can be compared with the Latin verb 'stare', which is closely related to the word group 'stable'. Thus the scene gives expression to a stable or solid protection of the steward-king by his god. The understanding relationship has nothing to do with the placement of the king "below" or "before" (Old English 'forstanden') the god. It rather indicates that the king functions as the intention- and orientationless medium of the judgments of his god.

When obedience in the sense of hearing the judgments of the gods is a sole constituent, a number of predications on a series of very specific cases can be dictated (Jaynes, 1976, p 200). How do the various predications intertwine? and what are the conditions for structural stability? The means necessary consists of the control of perspective, which can only be accomplished by determining the agent's position.

According to Gibson (1979), this measure does not result in a better representation of what is objectively given. It only concerns the observer's position. Gibson writes:

"Perspective ... puts the viewer into the scene, ... that is all. It does not enhance the reality of the scene". (p 283).

Thus perspective control lies in the definition of position of the agent, which in its turn determines what objectives or viewpoints are chosen and how they may change in the abstraction of structural relations.

The structure embedded in the formula (1) has been described as a complementary arrangement of its components in a cubic space (Bierschenk, 1984 a). With respect to the complementary roles of the A and O components, five basic activities can be carried out:

1. fixating the component to which the value (-) is bound,
2. binding the value (-) right adjusted,
3. mobilizing the component to which the value (+) is bound,
4. binding the value (+) right adjusted, and
These basic activities may be carried out by fixating or mobilizing the action component of the formula (1). All pairings possible are (--), (→+), (+→), and (++) and the change of information can be studied except for the first combination of symbols. A fixation of both intention and orientation, namely, means zero processing, while their mobilization implies maximal information synthesis. In this respect, the relations (--; ++) and (→+, ↔) are complementary to each other. This double asymmetry gives every pair a certain control over the development of every other. The asymmetrical pairs constitute the mechanism for the developmental control over the processes of differentiation and integration.

Every novel phase starts with a twist. The manipulable factors of the first phase, as discussed in Bierschenk (1984 a) can only secondarily influence the cooperation between the manipulable factors of the second one. A discussion of the second phase has been presented in Bierschenk (1984 b). The third phase is characterized by a cooperation of the manipulable factors perspective and viewpoints. Through their complementary arrangement, novel characteristics emerge. What operates in this new phase are the variations in (1) type of viewpoints and (2) function of perspective, which can be studied differentially. The differentiation process rests on the (→+) and (↔+) combinations, while integration is represented by the (→−) and (↔−) combinations. Making "understanding" (++) the outcome of a cooperative process of perspective and viewpoints presupposes the passing on of the effect of experience to a related activity, which means a specified judgment. This transformation can be regarded as a developmental twist leading from ecological to symbolic information processing and, consequently, integrating the study of natural language.

**Understanding as Predication**

Human language is generally conceived of as a conventional code. This implies that natural language is characterized by a
definable form. Within the unit of a sentence, the rules of grammar state whether the sequencing of symbols is a true description of a certain language or not. In the development from simple phrase structure grammars over transformational grammars to more semantically oriented ones, it seems to be of fundamental importance to freeze both perspective and viewpoints. Thus they build on a frozen symbol system and on syntactic models which prerequire the processing to take place according to rules for summative combinations of simple and discrete units, that is, primitives, which are structureless and timeless.

The concept of independence is basic within this kind of approach, which has as its consequence that the processing of symbols is based on rules of association. Symbolic input, therefore, becomes organized on the basis of mathematical-logical propositions about associations between variables. In agreement with a general conviction, at least in the formal sciences, symbolic processing results in strings of symbols or sentences on which operations such as addition, deletion, and insertion can be performed according to formally defined rules. With the perspective of Simon (1981), understanding can only be defined by the tasks that are mastered by "the most central parts of the central nerve system" (p 72). Assuming that this system performs its tasks in agreement with natural laws, which, according to Simon, are exclusively founded on logical form, understanding comes about only through what can be given propositional expression. The assumption that understanding a natural phenomenon can only be fully expressed and explained through a formal language of description has led to a shift from the simulation of perceptual systems (discussed in Bierschenk, 1984 b) to the construction of understanding systems.

Several projects have been in progress with the aim at simulating understanding. In these projects, natural language is made the basis for the processing of symbolic information. The beginning of natural language processing by computer was Ross Quillian's work (1968) on a "Teachable Language Comprehender".
This aim necessarily imposes a cognitive aspect on the linguistic model, which has as its consequence that some processing device is added to the grammar and a mechanism for accumulating the kind of facts assumed to be relevant for understanding and to be extracted from natural language sentences. Quillian presents three parameters, \( S \) for the subject of a clause, \( D \) for direct object, and \( M \) for modificand, which modifies another word in the same clause. (The linguistic unit of description is not a sentence.) \( S \) and \( D \) are related in a way indicated by a relation \( R \), which can be formulated with a statement in predicate logics, \( R(S, D) \). Thus a clause gets the following representation:

\[
R \rightarrow S \quad \quad D
\]

The symbols (\( \rightarrow, \rightarrow \)) represent prepositions and are being substituted with these when prepositions may be utilized as links between verbs and their subjects and objects.

Quillian explains that the subject of a clause needs not be a subject in the linguistic sense. It turns out to be somewhat obscure, since the symbol for preposition in the representation seems to indicate at least two kinds of linguistic subject. It would have been clarifying to get the definition of a non-linguistic subject. It would also have been more relevant to explain the grounds for letting prepositions determine direct objects, because here the linguistic sense of "direct" is different from the one employed. One explanation may be the influence from a case grammar model. It is namely proposed an interpretation such as Ergative and Locative for nouns within clauses of a certain type, where Ergative replaces \( S \). This is, however, not a redefinition of subject, rather a change of model. Moreover, any syntactic model defines the linguistic meaning of nouns through the verb. Even if \( S \) and \( D \) get their syntactic markers, it is unclear which ones and how, since the
formula indicates no meaning for verbs and no rules are given. With such a non-linguistic treatment of the verb, special problems are raised in explaining the M parameter, which is particularly syntactic.

The selection process has been manually performed by coders. Quillian argues that even if the coders are trained to decide the "exact" meaning of a sentence, they are "most unreliable and unhappy about making this distinction" (p 250-251). Therefore, Quillian discusses the need for a model by which ambiguity can be solved. The sentence

\[ \text{I through the man in the ring} \]

(3)

is rephrased in various ways in order to discuss possible solutions for an unambiguous coding. Regardless of whether the rephrasing can be logically represented by his graph or not, Quillian decides for this alternative:

\[ \text{THREW} \quad \text{IN} \quad \text{MAN} \quad \text{RING} \]

(4)

Even though the rephrasing "While in the ring I through the man" could be inferred from the sentence, the example is not the corresponding notation. The link indicates the modification of a relation that seems unfounded. If, instead, it represents a modification of the entire clause, this means that the formula is ambiguous and that the meaning of the rephrasing cannot be correctly understood by the coders. Otherwise it would not have been necessary to reformulate the coding instruction.

The reason for focussing on this model is that it layed the groundwork to a representation of propositional knowledge in the form of semantic nets, which means that frozen symbols could be studied with respect to variations in pattern. That is all. There is no indication how to create a world view out
of the various predications possible in Quillian's device. A modelling of a cognitive mechanism to simulate understanding presupposes, on the contrary, a dynamic relationship between the understander and that which shall be understood. No understanding can come about under the condition where perspective and viewpoints are frozen. The viewpoints must vary or be viewed from different angles, and the understander must be able to shift position, because his perspective determines what is perceived and how it is understood. Identification of invariant structure constitutes the prerequisite of the direct perception of information carried by symbols that characterize language understanding processes.

Understanding as State Change

The fundamentals of information processing are change of environment and organism. To acquire the invariant structures of symbolic information, variations of type and function must be understood. Winograd's (1972, 1973) system represents a progress in experimentation with language understanding compared to Quillian's model. In Winograd's system, a robot makes the changes of its perspective through tasks. The aim is to represent a robot's dialogue about objects and its answers to questions about its environment. With reference in ecological assumptions, it may be said that the geometrical configurations are used to indicate changes in the perspective of the robot. To describe the robot's perspective and viewpoints within this configuration the space metaphor will be used. The microworld staged, namely pictures a formal representation of a toy robot within an invisible room. This setting cannot be thought to represent a careful and controlled decrease of complex interrelationships as would have been required for a microworld. Because the scene is a fiction, it lacks both physical and conceptual anchorage, which means that the robot's response behaviour cannot be assigned the value meaningful. For the robot to be considered giving meaningful answers, its behaviour must be empirically testable.
The space metaphor implies a person sitting in front of a spacecraft window giving the robot different exploration tasks. What it sees and how it behaves are indicated through moving points and output shown on a TV screen. The metaphor implies, moreover, that the robot's environment is built up in correspondence with the construction of space, that is, it is composed of abstract and concrete objects. The abstract ones tell something about the configuration of the environment, while the concrete ones (data) correspond to the data base containing the information specifying the attributes of certain blocks (e.g., green, big, pyramid).

But let us examine in somewhat more detail the processes and procedures involved in computing understanding in Winograd's sense. The interaction with the robot is governed by the relatively uninteresting hypothesis that the sense organs can register single events and by the assumption that logical deduction mechanisms are sufficient for a simulation of the robot's understanding, although they necessarily imply problems of selection.

Understanding is represented through elementarization of actions into primitive events, as is shown by the definition of the PICKUP procedure:

\[
\begin{align*}
\text{(COLOR : BLOCK : GREEN)} \\
\text{EVENT1 (TYPE : PICKUP)} \\
\text{(OBJECT : BLOCK : GREEN)} \\
\text{(TIMEX : Y : 2)} \\
\text{(HOW (EVENT2, EVENT3)} \\
\text{(WHY : COMMAND)}
\end{align*}
\]

In the beginning, the robot in its world of blocks, probably sees nothing but a pattern of geometric figures, the abstract configuration of blocks. However, if it shall be able to orient itself in its world, it must be able to choose a viewpoint, which means it must apply a perspective. The person initiates a choice by, for example, giving the robot the command: at the green block! " When it answers that it now
is looking at the green block, the next following command would mean that it changes the perspective by focussing on certain attributes, specified in the command. Choice and change of the robot's perspective involve that "the path of perception" is a priori defined, just like the type of information and the contents related to the changes have to be specified in advance. With respect to the data base, the restriction implies a special order, whereas in the program, it has been built in semantic restrictions which make possible that certain events are excluded and others are equalized or treated differentially.

The crucial question now is: What does the system express? Apart from the logical inconsistency in the choice of features for the blocks (abstractions have no colour or size), the language used is indeed very adapted to the viewpoint studied, because not many words may be said to be ambiguous in relation to what they are to symbolize, like logical set relations, size, and other measurable features. This correspondence between language and "real" world expressions is particularly deceptive in this case, because, as a consequence of the viewpoints being of one single logical type, the world is contextless. The procedure model which also pertains to the definition of the objects is of such a nature that a category cannot be created and used over time. The same applies to the actions. These are well-defined steps in a preprogrammed sequence, that is, a totally known ritual, which, therefore, functions as a frozen symbol unit and is called upon again and again. In the robot's world, an object or event is out of existence as soon as it is out of sight. Thus the establishment of object permanence (Piaget & Inhelder, 1956) cannot be acquired. This mnemonic way of processing can never proceed to the symbolic invariance we call understanding. To speak with Gibson, what would the shapeless invariants be? Can the robot name the logics of the sequence or propose an alternative solution? With those prerequisites, what can the robot actually do? As I see it, it is capable of finding the value of the function. In this res. .t the experimental variable is under control. But as a typical representative of a non-
living system, the robot has no potential to understand what the function symbolizes.

**Understanding as Classification**

With reference to the scene analogy made by Schank (Schank & Abelson, 1977), it is clear that he, like the director of a play, imposes his own conception of the world and uses actors as instruments for his own ideas. His cognitive instruments are the images of those actors the way he sees them through the picture of language. He introduces Picture Producers and Picture Aiders as well as ACTs and Action Aiders (Schank, 1972). These are the experimental viewpoints and in this role they are guiding the present discussion of Schank's approach to understanding.

The influence from image theories has led Schank to believe that the picture or image is produced by a word or clause, instantiating information processing by the activation of the links describing the image. The arousal of the image is supposed to lead to the recognition of conceptual information represented in a sentence. His model for Conceptual Dependency may be conceived that way. The theoretical crucial point is, however, that the supposition made may be the case for concrete words (although cf. Rosch, 1975) but in far lesser degree for abstract words. As a philosophical-logical model it builds on semantic classes and must, therefore, be regarded as an elaborated way of describing relationships among concrete objects (humans, animals, things). But contrary to Winograd's approach, which varies only one type, Schank's objects have different features and thus belong to varying logical types. This means that the prerequisite for meaningful information processing per se is provided for through the "multiple typing" (Bateson, 1979, p 116). In what sense then do the semantic classes stand for reality? Or, conversely, in what sense can context provide meaning to the Picture Producers?

In general, context is essential to meaning. (Bateson, 1979; Gibson, 1979). When Schank and Abelson (1977) introduced the
script as a predetermined stereotyped sequence of actions defining a well-known situation, they took the same direction as Minsky (1975), who proposed a frame supplying the computer with "context". This "contextual" approach should be regarded as a prescription of a way of looking at objects and events in the real world, a way of manipulating the understanding process. But unlike newspaper journalists who frame a story (script) their way by attending certain (political) values to it allowing for varying versions of the same story, Schank composes a normative frame by means of philosophical-semantic values, which per definitionem cannot vary. Actors with a semantic perspective will be prescribed roles that suit the model. This implies that psychologically significant roles and strategies in the social context are pre-attributed and pre-ordered. Shifts are non-logical. Thus Schank's picture of language frames reality with the effect that images are artefacts (frozen symbols). Since the model is static (the features are well-defined), there is no attention possible in the understanding process.

Schank's (1972, 1973) goal is to create a theory of natural language understanding with the premise that the basis of natural language is "conceptual". The largest unit on the conceptual level is termed conceptualization, which requires a definition of relationships between Picture Producers. This kind of dependency will be illustrated with the verb 'eat':

\[
\text{vt } x \leftrightarrow \text{INGEST } \overset{o}{\rightarrow} y \xrightarrow{R} x \xleftarrow{\text{animate}} x \quad \text{and} \quad y
\]

(6)

R = Recipient case
\(\overset{o}{\rightarrow}\) = Objective case
\(\text{INGEST } = \text{TRANS(ition)}\)
\(\xrightarrow{\text{in}}\) = state change
\(\leftrightarrow\) = dependency relation
\(x = \text{Actor}\)
\(\text{vt } = \text{transitive verb}\)
Let us first make clear that the image analogy presupposes a state condition in the same way as a nominal and its modifiers (attributes) may be arranged as to their dependency relationship in a logical class, e.g., set inclusion. The nominal - nominal relationship may be classified with respect to the logic of a case frame. Since this is highly dependent on lexical entries for verbs, the nominal - nominal relationship could be further typified into a state relation, resulting from classes of instances. Just like any other classification system, language is here treated as a predetermined symbolic-logic hierarchy. With this position, changes can only take place between classes which serve as units for identification purposes. A class-based or, from an ecological point of view, type-based model has serious consequences when proposed for understanding processes. There is no intentionality in the model, i.e., no action producer, which means that the perspective is frozen.

A second question to be posed is: What is the cue function for the identification procedure? Schank's purpose has been to model the mind, that is to come up with a theory of cognitive prerequisites for understanding natural sentences, and not primarily to study understanding as performance, as was the purpose of Winograd. While he used rules for logical deduction and theorem proving to measure physical change (know how), Schank has tried to deal with immediate inference rules by the use of a set of generalized ACTs, labelled TRANS(itions) to measure mental or cognitive states (know that). This means that from a symbolic expression of an event, it should be possible to generate both the pre and the post states. These ACTs are used as building blocks for understanding (representing) every-day actions. Compared to Winograd's mechanism by which an action is represented by the inherent physically defined steps or movements, a building block is a universal serving as the link between entire networks of objects and attributes involved in the every-day actions.

The idea of the ACTs is similar to that of propositional knowledge proposed to represent abstract relationships between
entities in which a bridge or link is implicit or unnecessary once the relationship has been established. It is, therefore, confusing that the general term is not event. The term "primitive ACT" denotes that they are treated as semantic primitives suitable for computer manipulation without further explanation. They collect the Picture Producers under greater and greater classes. In this way the function of the ACTs has not really been as was expected by the model. It has no abstracting power, which means that no higher order functions can operate on the Picture Producer classes. Seen in this light, the number of viewpoints has shrunked into one single predefined giant semantic class, the script, consisting of one logical type.

**Understanding as Self-Reference**

Despite that it is obvious that an action or any other signal constitutes an intent (Howe & von Foerster, 1975), the developers of the metalanguages outlined have tried to avoid the self-reference in the classical way, although actions like pick-up, eat, take, or give signalize a distinction and prerequire an "intender" or agent. Unfreezing the symbol, namely implies making its affordance accessible.

When symbols are considered carriers of information abstracted from expressive behaviour, it is implied that they signify "regularized meaning". Moreover, it should be clear that experience and practice are underlying the development of symbols tying the symbol directly to the individual's own predisposition. The development of the symbols thus depends on continuous transformations and, therefore, is empirically founded. Making self-reference the starting point for the development of a theory that takes reality into account leads to a possibility to allow for intentionality and a formalism for answering the question Who does what? Preconditions for studying understanding are then the production of natural language texts, that is, running text and environmental descriptions, such as the producer of the text, the circumstances under which it was produced, and the topic(s). This is the ecological or empirical context. Thus on the one
hand text is an expression of the understander's (A) cooperation (a) with some objectives (O). His conceptions made by function of perspective and viewpoint are in some way or the other incorporated in the text.

When the text shall be understood by someone else, on the other hand, its function shifts from being expression of conceptualized experience to being context for making experience, and so the text becomes the objective for the understander's cooperation throughout the reading.

It can be assumed that, in communication, certain viewpoints are chosen according to a particular orientation of the text producer. These are intentionally expressed, which implies that the perspective of the producer is in the verbal flow (Bierschenk & Bierschenk, 1984). As has been presented in Bierschenk & Bierschenk (1986 a), linguistic variables may get different functions depending on perspective. The AaO mechanism developed is capable of controlling not only the organizational layout of the objectives and their change in complexity, but also the perspective chosen by means of verifying the agent function. To be complete, each observation must express an AaO. On the observational level, however, this relation does not always get its manifestation in language. With the help of context the observer may choose not to make explicit the full relationship and still be understood. On the other hand, the processing of an observation is only possible if the complete conceptualization of the observation can be made explicit. This explicitness is obtained with reference to the scope of action and the supplementation rule 5 (see page 4 above). The scope of action may be exemplified with the verb 'land' from a study with reference to the Visual Cliff (Bierschenk & Bierschenk, 1985).

\[
\text{A} \quad \text{a} \quad \text{O} \\
\text{Agent} \quad \text{land} \quad \text{Ground}
\]
The conceptualization follows closely a pure linguistic approach, but the viewpoint involved in the action is lifted up onto the observational level in case it has been left out by the observer. The phenomenon can be illustrated with the sentence connected to the previous example:

When they came close to the shallow side they extended their legs and singers as if they expected to land (on the surface of the shallow side)

(8)

The point of orientation, "land" or rather "ground" is conceptualized into the action, marking a generalization, which requires an explicit point in the single case. It is obvious that "ground" plays an important role and that the action involving this point of orientation requires a movement towards it. The English generalized verb through the expression chosen is literally translated with "touch ground". This example has been chosen to point at the difference between a description founded on pure linguistic components and one which presupposes conceptual components. The rules of grammar allow the verb land and similar verbs to be expressed without a point of orientation. But in the empirical observation, it is there of course, and is supplemented in from the context.

Textual transformations and change in the representation of observations are abstract phenomena but they are still the only prerequisite for the discovery of understanding. The relation of the structural nature of the observations to the entire text, on the other hand, is of a multivariate kind and thus too complex to illustrate other than by means of some graphical device. As has been demonstrated in Bierschenk & Bierschenk (1986, b, c), the depiction of this structure brings out information about the teleonomic component of the text.
References


Bierschenk, B. (1984). The split between meaning and being. *Kognitionsvetenskaplig forskning* (3). (b)


