ON THE INTERDEPENDENCE OF LANGUAGE AND PERCEPTION*

David L. Waltz
Coordinated Science Laboratory
University of Illinois at Urbana/Champaign

ABSTRACT

It is argued that without a connection to the real world via perception, a language system cannot know what it is talking about. Similarly, a perceptual system must have ways of expressing its outputs via a language (spoken, written, gestural or other). The relationship between perception and language is explored, with special attention to the implications of results in language research for our models of vision systems, and vice-versa. It is suggested that early language learning is an especially fertile area for this exploration. Within this area, we argue that perceptual data is conceptualized prior to language acquisition according to largely innate strategies, that this conceptualization is in terms of an internal, non-ambiguous "language," that language production from its beginnings to adulthood is a projection of the internal language which selects and highlights the most important portions of internal concepts, and that schemata produced in the sensory/motor world are evolved into schemata to describe abstract worlds. Examples are provided which stress the important of "gestalt" (figure-ground) relationships and projection (3-D to 2-1/2 or 2-D, conceptual to linguistic, and linguistic to conceptual); finally mechanisms for an integrated vision-language system are proposed, and some preliminary results are described.

Introduction

perception 1. obs.: CONSCIOUSNESS
2a: a result of perceiving: OBSERVATION
b: a mental image: CONCEPT
3a: awareness of the elements of environment through physical sensation ⟨color ~⟩
b: physical sensation interpreted in the light of experience
4a: direct or intuitive cognition: INSIGHT
b: a capacity for comprehension
syn see DISCERNMENT
(Webster's Seventh New Collegiate Dictionary)

Language understanding in its deepest sense is not possible without direct experience of its real world correlates. I believe that it is no accident that the same word can refer both to sensory awareness and to comprehension. Nearly all efforts in language processing, both in artificial intelligence and linguistics, have concentrated on transforming strings of words into trees or other structures of words (sometimes surface words, sometimes "primitive" words) or conversely, on producing strings of words from these structures. Few researchers have even recognized the importance of interfacing language and vision systems,† let alone uniting the two lines of research. (Exceptions include [Minsky 1975], [Woods 1978], [Miller & Johnson-Laird [1976], [Schank & Selfridge 1977], [Pylyshyn 1977 a & b], [Clark 1973]). At this time in history, AI vision and natural language researchers have little to say to each other; most of the work which treats language and perception together would I think be considered to lie in the realms of philosophy or psychology.

Moreover, the areas of language processing which could have a bearing on perception have been largely ignored. Very little work has been done on programs to understand language about space, spatial relations, or object descriptions. (Some exceptions are [Simmons 1975] and [Novak 1976], [Kuipers 1975], and [Goguen 1973].)

By the same token, current computer vision systems are not able to describe what they "see" in natural language; in fact very few programs can even identify objects within a scene (except for programs which operate in very constrained universes). Most vision systems produce scene segmentations, labellings or 3-D interpretations of scene portions, etc. Very little progress has been made toward the goal of having programs which could describe a general scence, let alone describe the most salient features of a scene. (Some exceptions are [Preparata and Ray 1972], [Yakimovsky 1973], [Bajcsy and Joshi 1978], [Zucker, Rosenfeld and Davis 1975], and [Tenenbaum and Weyl 1975].) Similarly, no programs I know are able to locate or "point to" scene items, given a natural language description of scene items or their whereabouts.

*This work was supported by the Office of Naval Research under Contract ONR-N00014-75-C-0612.

†While I intend perception to refer in the human examples to all the senses: vision, hearing, touch, smell, taste, and motor sense, in the case of computers, only vision has been explored in more than a cursory manner.

The need for an internal representation separate natural language

It is now reasonably well-established that people use large structures like "scripts" [Schank and Abelson 1977] or "frames-systems" [Minsky 1975] prevasively for reasoning and language and that a large script can be invoked by referring to a single salient aspect of the script. Thus we can answer a question like "How did you get here?" by saying "I borrowed my brother's car," and this answer can only be understood if we are able to use it to reliably retrieve a larger structure which answers the question more directly. (Example from George Lakoff [1978].) To understand language at the level of an adult human will certainly require a huge number of such scripts, with rich interconnections and powerful, flexible matching procedures as in Bobrow and Winograd [1977]. For scripts that refer to the physical world directly, what language can be used to construct the scripts? How can we construct scripts for abstract worlds (e.g. economics, psychology, politics)? What language should be used for abstract worlds? Are all these scripts to be hand coded?

Consider also sentences like "A man was bitten by a dog". If we were to be asked "Where could the man have been bitten, we would probably in the absence of more information guess the ankle, leg or arm. However if we are also told that the dog was a doberman or that it was a dashshund or that the man was lying down or that the dog was standing at the time, we would give somewhat different answers. It seems to me that natural structures for representing and answering questions about such language will be very different from those used in all programs today - a prototypical dog which can be scaled, representations of a person in various canonical positions, sizes of mouth openings and limbs, etc. would be the most appropriate, economical representations.

There is also a great deal of prima facie evidence of close ties between perception and the language used by adults to describe abstract processes such as thinking, learning, and communicating, and to describe abstract fields like economics, diplomacy, and psychology. Witness the wide use of basically perceptual words like: start, stop, attract, repel, divide, separate, join, connect, shatter, scratch, smash, touch, lean, flow, support, hang, sink, slide, scrape, appear, disappear, emerge, submerge, deflect, rise, fall, grow, shrink, waver, shake, spread, congeal, dissolve, precipitate, roll, bend, warp, wear, chip, break, tear, etc., etc. While we obviously do not always (or even usually) experience perceptual images when we use or hear such words, I suggest that much of the machinery used during perception is used during the processing of language about space and is also used during the processing of abstract descriptions. I do not find it plausible that words like these have two or more completely different meanings which simply share the same lexical entry.

There are significant linguistic generalizations to be found in language about perception. As an example, Clark [1973] demonstrates beautifully the structural regularities underlying prepositions which express spatial relations and the metaphorical transfer of spatial prepositions to describe time.

Finally, language plays an important role in guiding or directing attention and in providing explanations via analogy or via connections which are not directly accessible to sensory perception.

I contend that (1) we should strive to understand and to learn to represent the sensory/motor world; (2) we should study the relationship of language to the representations of the world, being aware that language does not itself represent the sensory/motor world, but instead points to the representations of this world via a set of word and structure conventions.

The development of perception and language

What we learn to name and describe in our experience must in some sense exist prior to and separate from the words associated with the experience. I believe that an infant develops very early a kind of "language of perception," i.e. a natural, innate segmentation of experience and ordering of the importance or interest of segmented items. Moreover, before an infant ever learns (or can learn) the name of an object, the infant must (1) be able to perceive that object as a unitary concept, and (2) must in fact perceive the unitary concept of the object as the most salient characteristic of that object. Thus, we assume that when we point to a telephone and say "telephone," the infant prefers to attach the name to be entire object and not to some property (e.g. color or size) of the object.

I will use the word "gestalt" to refer to such a unitary concept, because the words "concept" or "percept" may be misleading, and because I would not want to coin an entirely new word. By "object," I will mean not only visual objects, but also auditory "objects," having figure/ground relationships, such as a clap of thunder or a word spoken in isolation, and of course "objects" from other sensory and motor domains as well.

As I will discuss later, I believe that we can get around having to postulate perceptual primitives by viewing gestalts as the result of information theoretic types of processes, e.g. processes which select and attach importance to points with highly improbable surroundings (for example, points of symmetry).

How much is innate?

There has recently been a good deal of discussion about the "language" of thought or "mentalese" ([Fodor 1975]), [Pylyshyn 1978], [Woods 1978], [Johnson-Laird 1978]). The central issues discussed in these accounts are: (1) the innate "vocabulary" of such a language (innate concepts); (2) ways in which new concepts are added to mentalese; and (3) the relationship of mentalese items to words.

I would like to focus on one aspect of these discussions: innate concepts. To quote Pylyshyn [1978] at some length:

"There is no explanation, not even the beginnings of an approach, for dealing with the accommodation of schemata or conceptual structures into ones not expressible as definitional composites of existing ones. There is, in other words, no inkling as to how a completely new non-eliminable concept can come into being."

and later,

"The first approach [to this dilemma] is to simple accept what seems an inevitable conclusion and see what it entails. This is the approach taken by Fodor [1975] who simply accepts that mentalese is innate..."

"This approach to the innateness dilemma places the puzzle of conceptual development on a different mechanism from the usual one of concept learning. Now the problem becomes: given that most of the concepts are innate why do they only emerge as effective after certain perceptual and cognitive experience and at various levels of maturation?"

Pylyshyn goes on to sketch another solution in which mentalese is viewed as a sort of machine language for use with the fixed hardware architecture of the nervous system; new concepts could then arise if we allow the hardwired connections or architecture to change. As he points out, this merely buries the problem in hardware, and does not really provide a solution, but a different locus for the problem; at least it does get beyond the limitations inherent if the only formal concept development mechanism available is symbolic composition.

I find the notion of an innate language to be unsatisfying, and offer below a different sort of solution to the problem of the source of novel concepts.

A sketch of the development of perception and language

In this section I sketch what I feel is a plausible account of the development of perception and language. This account is heavily based on intuition and on my observation of my two children (Vanessa, now 5 and Jeremy, now 3); it thus represents an extreme case of inductive generalization. However, I have attempted to also cite ties with and supportive evidence from other work of which I am aware - I will be grateful to readers of this paper who supply relevant supporting or conflicting references which I do not acknowledge.

The basic positions I would like to argue on these issues are as follows:

(1) mentalese arises out of perceptual experience, and is not per se innate (i.e. present at birth);
(2) the development of mentalese depends instead on certain innate abilities and reflexes, plus perceptual experience. The innate abilities* are (at least):

a) the ability to form "figure/ground" perceptual relationships, where figures have distinguishing properties like local coherence on a homogeneous background ("objectness"), symmetry, repetition, local movement against a fixed background, etc. I will assume that the gestalts each have a certain salience or measure of "interestingness" to the infant which is a function of the inherent perceptual characteristics of each gestalt, the order and timing of attention to various gestalts (in turn these are eventually related to the current goals and hypotheses of the infant) and the current degree of pleasure or pain being experienced by the infant - at the extremes of pleasure or pain, gestalts have high salience, and could become independent goals to be pursued or items to be avoided.

b) the ability to remember quite literally one or more figures ("gestalts") from a figure/ground relationship for a short period of time (on the order of 10 seconds):

c) the ability to form associations between gestalts, where by association I mean that the experience of one gestalt can lead to the experience of an associated gestalt;

d) infants also have reflexes and certain innate behaviors, such as crying when hungry or in pain. Throughout this article, I will assume that these reflexes and behaviors - physical, mental, emotional, etc. - are also portion of an infant's perceptual experience.

(3) The primary goal of an infant is to maximize its pleasure and minimize its pain, and this goal drives the infant to attempt to understand its perceptual experience;

(4) The primary mechanism of understanding its experience is the organization of gestalts; this organization involves:

a) the formation of a taxonomy of the gestalts of experience, where the taxonomy is generated by successively subdividing existing categories into two (usually) or more new categories, and

b) the formation of associations between two or more gestalts to form new gestalts.

Reorganization occurs when previous taxonomic decisions appear to be deficient (e.g. are not leading to the achievement of pleasure or the cessation of pain), and the particular form chosen for reorganization depends on which gestalts are currently available, and of these which are most salient. The formation of gestalts by association is only possible initially between gestalts which both fall within the time period during which gestalts can be remembered. Associations initially are (probably) merely links; these links are themselves later sub-categorized into temporal sequence (elementary source of cause-effect relationships and "scripts"), constant copresence (elementary source of notions of identity or inherent connectedness), and eventually semantic relatedness (e.g. the link between the gestalt of a perceived visual object and the auditory gestalt of a word) as well as other connections.

*It is a bit strange to call these "abilities" since I do not believe that it is possible for us to experience the world at all except through the action of these "abilities," so that they might better be called "processes" or "properties of perception".

(5) Once associations are formed, items can become available as gestalts even if they are not at the time directly available to the external senses; this allows escape from the narrow bounds of the initial time window associated with externally perceived gestalts, since each gestalt can continue to reactivate others associated with it for indefinite periods (though "habituation" and competing external gestalts will soon interfere in general).

Taken together, any gestalt and the associates it can evoke form something like a "frame" [Minsky 1975]; every non-isolated concept thus has a frame. Default values for slots correspond to gestalts evoked in the absence of definite perceptual input. Language, then, is a sort of projection, where only some of the items to be communicated need to actually be mentioned directly. Syntax can be viewed as a means of constructing a perspective toward the gestalts selected by words and context; specific structures and words select specific connections between gestalts, as in [Fillmore 1977].

Early language is an extreme projection: a child beginning to speak can only output one word per sentence, later two words (this is the limit for a long time), etc. Thus "ball" when uttered by a one-year old might mean "I want the ball," "That's a ball," "Where is my ball?", "I was just hit by a ball," etc. There is striking recent evidence from the study of deaf children deliberately deprived of language* [Feldman, Goldin-Meadow and Gleitman 1977] that these children develop language independently, and that the length and the contents of their "sentences" are extremely similar to "sentences" of hearing children, in which certain types of sentences (e.g. verb + patient case) predominate and certain case roles (e.g. agent) are usually omitted. I suggest that their language development is similar because their perceptual experiences (and needs to communicate) are similar, and that the items chosen to appear in sentences are the ones with the highest salience.

In order to understand projected language, one must understand the context in which it is occurring. (For example, at age 2 years 8 months Jeremy Waltz held a new toy train up to the telephone receiver and said "look at the present I got, grammy.") Later language development can be viewed as learning to communicate in the absence of a shared physical context.

In the direction of language comprehension, we must then postulate a reconstruction process. Schank [1973] supplies evidence that by the age of one year, children understand the concepts of the primitive ACTs of conceptual dependancy;† Schank and Selfridge [1977] have demonstrated that children's responses to sentences at one year can be mimicked by a program by assuming the child has a single word input buffer, that (s)he selects only one word from a given input sentence, and finally picks and executes an ACT which plausibly could involve the concept associated with the word selected. Thus, when told "Take the ball to Roger" a child might simply get the ball, or take the ball to someone else (if ball were selected) or run to Roger (if Roger were selected).

I would finally like to emphasize the idea that language at all ages (not just for children) involves the complementary processes of projection from and reconstruction into mentalese. (See also Marcus [1978] for more evidence on input butter restrictions in adults.)

(6) New gestalts can probably be integrated into the infant's taxonomy only one at a time, i.e., new items must be associated with items which are already part of the organized taxonomy. Thus words would usually be learned for items which are already organized conceptually, although novel words could be used to point out the need for new categories (e.g. by pointing out that a dog and sheep are different). The net result is the likelihood of many more total concepts (individual concepts, associated individual concepts, and associations of associations) than there are concepts with words attached to them ([Woods 1978] comes to a similar conclusion).

Properties can be selectively named by (a) presenting two or more quite different objects which share a single property, say color, or (b) contrasting objects which differ in only a single property (big/small), or (c) having names firmly enough in place so that items pointed to can be understood as details or properties, not the name per se. ((a) and (b) are like Winston's [1975] "near-misses"). I would like to point out the analogy given above and the use of metaphor to select and highlight relationships for which we do not already have names.

Concepts are at least potentially completely unambiguous, with the exception of auditory gestalts corresponding to words.§ Clearly some auditory gestalts corresponding to words can be associated with two or more different gestalts (e.g. fair (carnival), fair (clear or beautiful), fare (travel fee), fare (menu items)); I suggest that in order to be understood unambiguously, such words must occur in a context where one underlying concept is associated much more closely with concepts in the current context (verbal or other perceptual). This idea is related to work in spreading activation for semantic networks [Collins and Loftus 1972], as well as to "focussing as in Grosz [1978].

---

*Because the Philadlphia school system believes that lip-reading and vocal speech are best, and that learning sign language destroys the willingness of children to learn to lip read and speak.

†E.g MOVE (a body part), INGEST, EXPEL, PTRANS (transfer a physical object), ATRANS (transfer an abstract relationship, e.g. possession). MTRANS (transfer information between or within animals), PROPEL (apply force to), GRASP, SPEAK (make a noise, and ATTEND (Focus a sense organ on an object [Schank 1975].

§Of course, visual or other sensory input can be ambiguous at times, but if a unique mentalese item is selected for a sensory item, the item is then uniquely understood.

(7) Jackendoff [1975] and Gruber [1965] have pointed out evidence that linguistic schemata we develop to describe GO, BE and STAY events in the sensory/motor ("position") world are later transferred via a broad metaphor to describe events in abstract worlds (possession, "identification" and "circumstantial"). Thus we learn to use parallel surface structures for conceptually very different sentences like:

(1a) The dishes stayed in the sink (position)
(1b) The business stayed in the family (possession).
(2a) His puppy went home (position).
(2b) His face went white (identification).
(3a) She got into her car and went to work (position).
(3b) She sat down at her desk and went to work (circumstantial).

Along these same lines, there are striking parallels in the structures of Schank's [1975] conceptual dependency diagrams for PTRANS, ATRANS, and MTRANS (see earlier footnote). Reddy [1977] has described what he calls the "conduit metaphor" for linguistic communication in which we typically speak of ideas and information as though they were objects which could be given or shipped to others who need only to look at the "objects" to understand them. Thus we say "You aren't getting your message across," "She gave me some good ideas," "He kept his thoughts to himself," "Let me give you a piece of advice," etc. (Reddy has compiled a very long list of examples.)

These examples suggest many deep and fascinating questions. It seems clear that the same words and similar syntactic structures can be transferred to describe quite different phenomena. What internal structures (if any) are also transferred in such cases? What perceptual criteria are used to classify events to begin with? Ultimately? How does a child transfer observation to imitation? How are memories of specific events generalized to form event types, and how are the representations of event types related to memories of specific events?

To answer one portion of these questions, it seems clear from an economic point of view that if syntax and words are conventional and not innate, we would want to include only enough distinct syntactic structures and words to make distinctions that are necessary and to unambiguously select mentalese representations. We would thus predict that words and syntactic structures would be heavily shared (see also [Woods 1978]).

I suggest that internal mentalese structures are not transferred, but that, just as single words can point to more than one concept, these parallel structures for verbs can point to more than one mentalese structure. However, there are limitations: the structures pointed to must share some properties, e.g. the number of case roles must be the same, and selection restrictions on case role slots should be sufficient to choose the appropriate concept unambiguously.

Another interesting question involves the status of inferential knowledge - is it attached to mentalese concepts or to words? Surprisingly, there may be some evidence that inferential knowledge is attached to words. In the position world we know that an object can only be in one place at a time, that two objects cannot occupy the same place at the the same time, etc. Some of these same inferences may be carried over inappropriately to the possession world: for example my children appeared to have some difficulty fully understanding concepts like "joint ownership". If we assume that in the conceptual transfer a child creates an imaginary "possession basket" for each person, and that the interiors of two such baskets cannot intersect, then objects must be in one basket or another, and sentences like "Real [our dog] belongs to all of us but he's really mine" (Vanessa, about 4-1/2) become more intelligible. (There are of course other plausible explanations for this sentence.) Reddy [1977] has also pointed out ways in which the "conduit metaphor" for communication minimizes the constructive role of the listener, and leads to the notion that failure of communication is due primarily to the speaker. Whorf's [1956] ideas and data may be relevant here also.

The role of aesthetics

I feel that it is important to keep our central attention on the functional roles of perception and language for the survival of the infant, which I take to be the primary goal in evolution, and the place where we must look ultimately for explanations about innate abilities and early development. I accept Pugh's [1977] suggestion that all our values (pleasure, pain, good, bad, happy, unhappy, etc.) serve as "secondary values," i.e. as surrogate values for the primary value of survival. We have these secondary values because they allow us to distinguish situations which have significant positive or negative survival value. Woods [1978] has pointed out the survival value of language in allowing the transmission of knowledge in the absence of genetically "wired" behavior. (See [Dennett 1974].)

I suggest that the values like goodness, economy, aesthetics, and interestingness are pervasive in our perceptual systems and in the mechanisms which evaluate hypothesized taxonomies of experience. We attend to sensory items which interest us, store descriptions in ways that are aesthetically satisfying (e.g. have good symmetry properties, divide phenomena into balanced categories, help avoid dangling, unexplained phenomena, etc.), in addition to evaluating whether our hypotheses are helping us get what we want.

Development of a taxonomy of experience

Let us assume that we start with a unitary concept of the world, and examine a plausible development of distinctions in the visual world.* The first sort of distinction likely is moving/not moving, where "moving" refers to a figure on a ground. The "moving" category is soon divided into categories for moving items over which the infant has some control and moving items where (s)he does not (random motions). Later, this category is separated into items where the infant has direct control (e.g. parts of the body), and others (e.g.

*It is likely that some distinctions, e.g. kinesthetic moving/not moving, are made in utero.

parents who sometimes come when the child cries, objects nearby which can sometimes be hit or touched by body movements, etc.).

Out of this process eventually, come basic distinctions like self/other, mind/body, near (reachable)/far (unreachable); also, categories from various sense modalities can be merged (objects from the tactile and visual worlds, mother from the visual, auditory and tactile worlds, etc.). I have a wealth of observations on the development of these distinctions from watching my children which cannot be expounded further in the space here. I would like to suggest in passing that the development of this taxonomy can have deep psychological significance - to point out one example, consider the following contrasting situations: (1) parents are attentive to an infant's cries and thus are thus initially within the category of moving items controlled by the infant vs. (2) the parents are inattentive to cries, and thus initially are classified in the "random motion" category. See Wilber [1976] for an extension exploration of the development of fundamental dualities.

A computer model of gestalt formation

My recent work in vision [Waltz 1978] has explored computational methods for finding points in scenes which have high information content, which I suggest as the primary basic of the definition of "interestingness," which in turn should drive attention.

Because we (George Hadden and I) have been working with static scenes, our programs do not sperate moving figures on grounds (which I take to be important, as should be obvious from earlier discussions).* We have concentrated instead on methods for funding symmetry axes, points with high curvature, edges and edge completions, isolated objects on backgrounds, spatially repeated patterns, and characteristics texture elements. In each case we are assuming that processes that be bottom-up and task-independent (although I would be willing to include some special preferences for things like vertical or horizontal directions).

This work is based on the notion that shape is the best "property" with which to sort items into categories. Our programs attempt to locate unique points of high information (e.g. the center of a circle) and to store at that point sufficient information to "unfold" a shape envelope of an object (the shape envelope is the same for a solid line rectangle, dotted line rectangle, rectangle with a notch removed from the side, etc.).† The notion here is that shape should be represented hierarchically, with the shape envelope typically at the top of the hierarchy, and deviations from the shape envelope located lower, along with other properties like color, size, etc.

However, in the long run visual objects should be described in a more flexible structure which draws on a list of properties; my current favorite list of properties comes from Pylyshyn [1977b] who in turn got the list from Basso [1968]. Basso identified the items through the analysis of classificatory morphemes in diverse languages. He identifies semantic dimensions: animal/non-animal, enclosed/non-enclosed; solid/plastic/liquid; one/two/more that two; rigid/nonrigid; horizontal length > 3 times width or height/"equidimensional"; portable/nonportable. These can be combined to form categories which recur commonly in other cultures, e.g. "rigid and extended in one dimension" (pencil, knife, cigarette); "rigid and equidimensional" (pail, light bulb, egg, box, coin, book); "flexible and extended in two dimensions" (paper, blanket, shirt); "flexible and extended in one dimension" (rope, belt, chain).

Of importance in all these cases is that the descriptions be hierarchical, with meaningful generalizations at the top of the hierarchy (see Preparata and Ray [1972] for other ideas along these lines that we have adopted), and the description be capable of being generated bottom up.

Visual imagery

My position may be acceptable to both Pylyshyn [1973] and Kosslyn [1978]. With Pylyshyn, I believe that visual descriptions are propositional, and that the descriptions are organized hierarchically. However, as argued in the last section, shape seems to be the primary distinguishing property of objects, and we have reason to believe that shape can in general be represented rather compactly with respect to some point (e.g. of symmetry or a centroid). I suggest that shape representations may actually be capable of being "run backwards" or "unfolded," and that the result may be activation of portions of our brains (visual cortex?) which are also activated when an item is directly perceived.

In this view, visually imagery could provide useful clues about the nature of shape representation. However visual imagery does not seem to be generally experiences or used - based on informal questioning of my classes, fewer than half of engineering students (who might be expected to visualize more frequently than average) report other than occasional use of imagery. (As a person who does use visual imagery extensively, I found this result surprising.) Perhaps imagery is a latent talent which can be developed; once developed I believe it has significant value for problem solving, organization of material, and memorization.

---

*Moving figures are however trivial to compute by subtraction of successive frames of a moving scene.

†As discussed in Bajcsy & Joshi [1978], in adult speech shape is described verbally by referring to other familiar (or canonical) objects. However, in order to note the similarity of objects, we must have neutral descriptions of each, e.g. the kinds of descriptions I am discussing here. Also of interest is the observed fact that we have very few verbal items to describe shape in a non-relational manner, except for highly regular objects (sphere, cube, etc.).

In a related vein, I am intrigued by (and intend to follow up further) the ides that we may organize memory in such a way that we can use perceptual strategies for understanding its contents. Two particularly suggestive phenomena (other than visual imagery):

(1) The striking similarity of some memories to sensory phenomena: in order to retrieve the punchline of a joke or content of a story, I sometimes have to go through the whole joke or story; I can "play back" music; etc.,

(2) recent work by Fillmore [1977] and Grosz [1978] which suggests that language may guide an analog of the attention process by suggesting a perspective from which to view memory structure(s) as they are retrieved.

Can we dispense with the idea of innate ideas?

In order to show that we can account for mentalese without requiring innate ideas, I must show (1) that the mechanisms proposed are capable of generating all the primitive concepts of mentalese, and (2) that I have not simply buried innate ideas somewhere in the mechanisms. Let me say immediately, relative to point (2) that there are some innate ideas in my account; one set of ideas are related to the values (good/bad, symmetrical/nonsymmetrical, etc.) discussed earlier. There must also be ideas relating to generating hypotheses on which the values can operate, and the idea of objectness (if this can be called an idea) must be present. Hypothesis generation might seem a candidate for further search for embedded ideas; however, as I have described it, hypothesis generation is primarily a categorizing operation, where it acts on the "raw material" of perception. On the whole I do not believe that it is difficult to accept the sorts of innate ideas which remain in my account.

It is much more difficult to make a convincing case for the sufficiency of these mechanisms to explain mentalese. (The situation is not aided by the fact that there are few suggestions concerning the nature of mentalese, let alone general agreement on its nature.) I have dealt at least briefly here with physical objects (from the points of view of all senses), properties, actions (to a slight degree - I do have what I feel is a reasonable account), cause-effect relationships, aspects of the mind-body problem, as well as a number of other concepts. What is missing? The two major areas I am aware of are (1) quantification (I suggest this could be handled by assuming that its origins are in operations on finite sets); and (2) logical operations (probably these also need to be innate).

Afterthoughts and acknowledgements

It has been a long time since I read Koffka [1935] and Piaget's works (e.g. [1967] and [Piaget & Inhalder 1967]), but clearly many of the ideas in this paper can be traced to those two sources. I had not read Jacendoff's [1978] paper in this volume before writing this paper, but I wish I had been able to.

I would especially like to acknowledge the ideas and criticisms I have received in conversations with Bill Woods, Phil Johnson-Laird, Harry Klopf, Lois Boggess, and Jeff Gibbons.

References

Bajcsy, R. and Joshi, A. (1978), The problem of naming shapes: vision-language interface. In TINLAP-2.

Basso, K. H. (1968), The western apache classificatory verb system: a formal analysis. Southwestern Journal of Anthropology 24, 252-266.

Bobrow, D. G. and Winograd, T. (1977), An overview of KRL, a knowledge representation language. Cognitive Science 1, 1, 1977.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (ed.) Cognitive Development and the Acquisition of Language, Academic Press, N.Y., 27-63.

Collins, A. and Loftus, E. (1975), A spreading activation theory of semantic processing. Psychological Review 82, (5).

Dennett, D. (1974), Cited by Woods [1978] as source of many ideas; I could not locate citation.

Feldman, H., Goldin-Meadow, S. and Gleitman, L. (1977), Beyond Herodotus: The creation of language by linguistically deprived deaf childred. In A. Lock (ed.) Action, Gesture and Symbol: The Emergence of Language, Academic Press.

Fillmore, C. (1977). The case for case reopened. Draft of paper to appear in Syntax and Semantics series.

Fodor, J. (1975), The language of thought. New York: Crowell.

Goguen N. (1973), A procedural description of spatial prepositions. (M.S. thesis) University of Pennsylvania, Dept. of Computer

Grosz, B. J. (1978), Focussing in dialog. In TINLAP-2.

Gruber, J. S. (1965), Studies in lexical relations. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Jackendoff, R. (1975), A system of semantic primitives. In R. Schank and B. Nash-Webber (eds.) Theoretical Issues in Natural Language Processing, ACL, Arlington, VA.

Jackendoff, R. (1978), An argument about the composition of conceptual structure. In TINLAP-2.

Johnson-Laird (1978), Mental models of meaning. Paper presented at the Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, University of Pennsylvania, May 1978.

Koffka, K. (1935), Principles of gestalt psychology, Harcourt Brace, New York.

Kosslyn, S. M. (1978), On the ontological status of visual mental images. In TINLAP-2.

Kuipers, B. J. (1975), A frame for frames: representing knowledge for recognition. In D. Bobrow & A. Collins, Representation and Understanding, Academic, New York, 151-184.

Lakoff, G. (1978), Comments during colloquium at Dept. of Linguistics, Univ. of Illinois, April 1978.

Marcus, M. (1978), A computational account of some constraints on language. In TINLAP-2.

Miller, G. A. and Johnson-Laird, P. (1976), Language and Perception, Harvard University Press, Cambridge, MA.

Minsky, M. L. (1975), A framework for representing knowledge. In Winston (ed.) The Psychology of Computer Vision, McGraw-Hill, N.Y.

Novak, G. S. (1976), Computer understanding of physics problems stated in natural language. Tech. Report NL-30, Dept. of Computer Science, Univ. of Texas, Austin.

Piaget, J. (1967), Six Psychological Studies. Vintage, New York.

Piaget, J. and Inhelder, B. (1967], The Child's Conception of Space. Norton, New York.

Preparata, F P. and Ray, S. R. (1972), An approach to artificial nonsymbolic cognition. Information Sciences 4, 65-86.

Pugh, G. E. (1977), The Biological Origin of Human Values, Basic Books, New York.

Pylyshyn, Z. W. (1973), What the mind's ege tells the mind's brain: a critique of mental imagery. Psychological Bulletin 80, 1, 1-24.

Pylyshyn, Z. W. (1977a). What does it take to bookstrap a language? In Language Learning and Thought, Academic Press, N.Y., 37-45.

Pylyshyn, Z. W. (1977b), Children's internal descriptions. In Language, Learning, and Thought, Academic Press, N.Y., 169-176.

Pylyshyn, Z. W. (1978), What has language to do with perception? Some speculations on the Linguamentis. In TINLAP-2.

Reddy, M. (1977), Remarks delivered at the Conference on Metaphor and Thought, University of Illinois, Urbana, Sept. 1977.

Schank, R. C. (1973), The development of conceptual structures in children. Memo AIM-203, Stanford AI Lab., Stanford, CA.

Schank, R. C. (1975), The primitive ACTs of conceptual dependency. In R. Schank & B. Nash-Webber, Theoretical Issues in Natural Language Processing, ACL, Arlington, VA, 34-7.

Schank, R. C. and Abelson, R. P. (1977), Scripts, Plans, Goals, and Understanding, Lawrence Erlbaum, N.J.

Schank, R. C. and Selfridge, M. (1977), How to learn/what to learn. Proceedings of the 5th Int'l Joint Conf. on Artificial Intelligence, MIT, Cambridge, MA, 8-14.

Simmons, R. F. (1975), The Clowns Microworld. In R. Schank and B. Nash-Webber (eds.) Theoretical Issues in Natural Language Processing, ACL, Arlington, VA.

Tenenbaum, M. and Wayl, S. (1975), A region analysis subsystem for interactive scene analysis. Advance Papers of the 4th Int'l. Joint Conf. on Artificial Intelligence, Tbilisi, USSR, 682-7.

Waltz, D. L. (1978), A model for low level vision. In A. Hanson & E. Riseman (eds.) Machine Vision, Academic Press, N.Y. (to appear).

Whorf, B. L. (1956), Language, Thought and Reality, MIT Press, Cambridge, MA.

Wilber, K. (1977), The Spectrum of Consciousness, Quest.

Winston, P. H. (1975), Learning structural descriptions from examples. In Winston (ed.) The Psychology of Computer Vision, McGraw-Hill, NY.

Woods, W. A. (1978), Procedural semantics as a theory of meaning. Draft of paper presented at Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, Univ. of Pennsylvania, May 1978.

Yakimovsky, Y. (1973), A semantics - based decision theory region analyzer. Advance papers of the 3rd Int'l Joint Conf. on Artificial Intelligence, Stanford, CA, 580-8.

Zucker, S., Rosenfeld, A., and Davis, (1975), General purpose models: expectations about the unexpected. Advance Papers of the 4th Int'l Joint Conf. on Artificial Intelligence, Tbilisi USSR, 716-21.