Schank: Conceptual Dependency

Wednesday, March 3, 2021 8:36 PM

Safari

Scripts, Plans, Goals and Understanding

An Inquiry into Human Knowledge Structures

Roger C Schank
Robert P Abelson
Yale University

LEA
1977
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Hillsdale, New Jersey

DISTRIBUTED BY THE HALSTED PRESS DIVISION OF
JOHN WILEY & SONS
New York Toronto London Sydney

bated breath. The geometry of bouncing balls, the 'feel' of dough texture, and many other aspects of human activities involve knowledge falling outside of our present boundaries. This is because (among other reasons) visual and kinesthetic processes cannot readily be represented in verbal form. However, a great deal of the human scene can be represented verbally, and we have no lack of things to work on.

1.3 Traditional Points of View

We have mentioned that our task lies at the intersection of psychology (more specifically, cognitive psychology and cognitive social psychology) and artificial intelligence. Since we are concerned with verbally expressible knowledge, there is also an overlap with linguistics. When one tries to work in a disciplinary intersection, one inevitably comes into conflict with the traditional standards, habits, and orientations of the parent disciplines. This is especially true when the disciplines correspond to university departments, breeding suspicion of out-groups (cf. Campbell, 1969). Here we briefly sketch some of these conflicts, which we have resolved somewhat differently from others working at the same intersection.

Psychology is a heterogeneous discipline. The major subdivisions are developmental, clinical, cognitive and social psychology, and psychobiology. It is surprising to the non-psychologist but familiar to all but the youngest generation of psychologists that cognitive psychology is a relatively new branch of study. American experimental psychology was dominated for so long by behaviorism – roughly, from 1935 to 1960 – that the study of mental processes lay almost entirely dormant while other branches of psychology were developing rapidly. Since mental events could not be observed directly, there was scientific resistance toward relying on them to explain anything, whatever the scientist's common sense might tell him. Introspective evidence was not regarded as objectively trustworthy.

Since 1960, there has been an enormous surge of careful experimental work on mental phenomena. Skinner notwithstanding, human psychology could not seem to do without cognitive processes. Nevertheless, the methodological caution of the behaviorists was carried over into this resurgence. Acceptable scientific procedure called for quantitative response measurements such as accuracy of recall or choice reaction time when subjects were confronted with well-controlled stimulus tasks. In the verbal domain, stimulus control usually entailed repetitive trials on isolated verbal materials, deliberately avoiding meaningful connotations in the experimental situation. While recent experimental materials have not been as trivial as the old-fashioned nonsense syllables, neither have they been genuinely meaningful or even necessarily plausible. Experimental tasks are often unusual and/or unnatural in relation to tasks encountered daily by people in using language. For example, in a well-known experiment by Foss and Jenkins (1973), subjects listened to 48 sentences such as 'The farmer placed the straw beside the wagon', with instructions to press a key the instant they first heard the phoneme 'b'. In another well-known series of experiments by Anderson and Bower (1973), subjects heard 32 unrelated sentences such as 'In the park, the hippie kissed the debutante', 'In the bank, the tailor tackled the lawyer', etc., and an hour later were asked to recall as many of them as they could. The artificiality of tasks such as the latter led Spiro (1975) to remark tartly,

> *Why should a research subject integrate the to-be-remembered information with his or her other knowledge? The role the information will play in his or her life can be summarized as follows: take in the information, hold it for some period of time, give it back to the experimenter in as close to the original form as possible, and then forget it forever. The information cannot be perceived as anything but useless to the subject in his or her life (given the common employment of esoteric or clearly fictional topics as stimulus materials). The information, even when not clearly fictional, is probably not true. In any case, the subject knows that the relative truth of the information has nothing to do with the purpose of the experiment. (p.11)*

In complaining about the lack of meaningful context in experiments such as these, it is no doubt unfair to present them out of their context. The experimenters had serious purposes, and the data were of some interest. But since our needs are for a set of interrelated constructs to explain the process of natural understanding of connected discourse, this style of experimentation is both too unnatural and too slow. There has been a gradual increase in research with connected discourse as stimulus material (e.g., Bransford and Johnson, 1972; Kintsch, 1974; Frederiksen, 1975; Thorndyke, 1977) but the field is still marked with a very cautious theoretical attitude. We are willing to theorize far in advance of the usual kind of experimental validation because we need a large theory whereas experimental validation comes by tiny bits and pieces. Our approach, in the artificial intelligence tradition, is discussed in Section 1.6.

If the research properties of experimental cognitive psychology are often unduly restrictive, the traditions in the field of linguistics are even more restrictive. Linguistics has concerned itself with the problem of how to map deep representations into surface representations (see Chomsky, 1965). After a long obsession with syntactically dominated deep representations, recent work in linguistics has oriented deep representations much more towards considerations of meaning (Lakoff, 1971; Clark, 1974). Despite this reorientation linguists have managed to miss the central problems.

Two fundamental problems stand out: How do people map natural language strings into a representation of their meaning? How do people encode thoughts into natural language strings? Because of a purported interest in the purely formal properties of language, linguists have consciously avoided both of these naturalistic problems. The second question seems, on the surface, to be closer to a linguist's heart. But linguists treat generation as a problem of determining whether a string is grammatical, i.e., whether it can be generated by the grammar they have set up. A grammar that generates natural language strings would be interesting and useful of course, if, and this is a big 'if', it started at the right place. Linguists tend to start their grammars at the node S (for sentence). People, on the other hand, start with an already well-formed idea (or the beginnings of an idea) that they want to express. Linguists thus wind up concerning themselves with considerations of semantics at the level of 'Can I say this string? Will it mean something'? People already know what they want to say and that it is meaningful.

Two remedies for this linguistic notion of semantics come to mind. For the generation problem the obvious solution is to start the process earlier. How do people get thoughts to express? Linguists explicitly consign this question to other disciplines; yet it is an important part of the generation process, and one which when treated as a linguistic question completely changes the process under investigation. The other remedy is to apply such semantic considerations as 'Does this string mean something?' to the problem of understanding what someone else has said. Questions of how strings can be meaningfully interpreted belong to the domain of understanding, not generation, where Chomsky (1965, 1971) has repeatedly put them. (Actually Chomsky would deny that he works on generation. Transformationalists prefer to think of themselves as working on an abstract formalism with no process notions present at all.)

Linguists have almost totally ignored the question of how human understanding works. Since human understanding is dependent on the ability to decode language this seems odd at best. Some 'computational linguists', (e.g., Friedman, 1969 and Kay, 1973) have attacked the problem. However, they have followed linguistic tradition and consequently have maintained one of the fundamental flaws of linguistics in their work. They have divided the problem into linguistic and non-linguistic parts, a division that holds up no better for understanding than it does for generation.

Artificial intelligence has a somewhat more congenial recent history. The field is relatively new, and its early efforts were predominantly oriented toward getting computers to solve logical and mathematical problems (e.g., Newell, Shaw and Simon, 1957; Minsky, 1961; Feigenbaum and Feldman, 1963; Nilsson, 1971), and to play games such as checkers (Samuel, 1963) and chess (Bernstein et al, 1958; Newell, Shaw and Simon, 1958) intelligently. Early efforts to have computers deal with natural language were marked either by drastic failure (as in the case of mechanical translation from one language to another) or drastic oversimplification in the admissible vocabulary (Green et al, 1961) and grammar (Abelson, 1963; Colby and Gilbert, 1964), or by programming tricks producing smooth locutions which made the computer seem smarter than it actually was (Weizenbaum, 1966).

It has nevertheless been consistently regarded as important that computers deal well with natural language. In practical terms, such a development would mean that anyone could interact with a computer without learning a programming language or some special code to communicate about a special problem, whether it be library or consumer information, travel and ticket reservations, suggestions about home repairs, crop protection, first aid, etc. Computerized teaching programs would not have to be restricted to giving multiple-choice tests of the student's knowledge, but could interpret and respond intelligently to free-form answers and questions from the student. None of these high-sounding things are possible, of course, unless the computer really 'understands' the input. And that is the theoretical significance of these practical questions – to solve them requires no less than articulating the detailed nature of 'understanding'. If we understood how a human understands, then we might know how to make a computer understand, and vice versa.

In the last several years there have been two clusters of developments in artificial intelligence which are miles ahead of previous efforts. First, there is a new generation of programs for 'parsing' sentences (in English and other languages) – that is, for deciding the proper features (such as what part of speech) to assign to each word in a sentence. The approaches of Woods (1970), Winograd (1972), Riesbeck (1975) and Marcus (1975) differ in the relative priority they give to syntactic or semantic features in parsing, but all agree that semantic features are considerably more important than linguists had generally been willing to acknowledge. Second, there has been increasing recognition that context is of overwhelming importance in the interpretation of text. Implicit real-world knowledge is very often applied by the understander, and this knowledge can be very highly structured. The appropriate ingredients for extracting the meaning of a sentence, therefore, are often nowhere to be found within the sentence.

There are several famous illustrations of this latter point. Collins and Quillian's (1972) is:

1 The policeman held up his hand and stopped the car.

Somehow in understanding this sentence we effortlessly create a driver who steps on a brake in response to seeing the policeman's hand. None of these intermediate links are mentioned in sentence (1). Another example, (from Abelson, 1969) is:

2 I went to three drugstores this morning.

Very innocently, the concept that the person must not have found what he wanted in the first two drugstores is implied, otherwise why would he have gone to three? This kind of implicit inference is very common – and of course can be wrong, but it is intrinsic to natural understanding that useful, fallible presumptions creep in.

Perhaps the simplest example of implicit inferences can be seen in a simple sentence such as (from Schank, 1972):

3 I like apples.

The speaker is talking about 'eating' but this is not explicitly mentioned. And why should it be? The speaker, unless he is deliberately trying to fool his listener, knows that the listener knows what action is being implicitly referenced. These examples were constructed with a point in mind, but are not really unusual. In all of them, and in many, many other examples to be found in this book, more is at issue than 'semantics'. It is 'pragmatics', the way things usually work – not how they might conceivably work – which most often impels the reader toward an interpretation. The reader brings a large repertoire of knowledge structures to the understanding task. Elsewhere these structures have been called 'frames' (Minsky, 1975) and 'schemata' (Rumelhart, 1976). Rumelhart puts the matter very well when he says, 'The process of understanding a passage consists in finding a schema which will account for it.'

Interestingly, the idea of the schema in the interpretation of human events has a long tradition in social psychology. American social psychology had its roots in Gestalt psychology, and therefore did not succumb to the excesses of behaviorism the way human experimental psychology did. The phenomenology of mental life maintained a central role, largely through the towering influence of Kurt Lewin in the 1940's. Lewin (1936) wrote about human goal strivings in terms of internal images people had of their 'life spaces'. Since then a long succession of social psychologists have appealed to structured ideational kernels of the way people supposed the world to be organized: Heider's (1946, 1958) 'balance principle' and 'naive psychology'; Festinger's (1957) 'cognitive dissonance theory'; Abelson and Rosenberg's (1958) 'psycho-logic'; Kelley's (1967) and Jones and Davis' (1966) 'attribution theory', and many more. The terminology of the 'schema' is very much active in the 1970's (cf. Kelley, 1971; Tesser, 1977), even in areas well beyond social psychology (Rumelhart, 1975; Bobrow and Norman, 1975; Rumelhart and Ortony, 1976). The second author's orientations in the present book can be traced back to earlier excursions into 'hot cognition' (Abelson, 1963), 'individual belief systems' (Abelson and Carroll, 1965), and 'implicational molecules' (Abelson and Reich, 1969).

There is a very long theoretical stride, however, from the idea that highly structured knowledge dominates the understanding process, to the specification of the details of the most appropriate structures. It does not take one very far to say that schemas are important: one must know the content of the schemas. To be eclectic here is to say nothing. If one falls back on the abstract position that only form is important, that the human mind is capable of developing knowledge structures of infinitely varied content, then one sacrifices the essence of the structure concept, namely the strong expectations which make reality understandable. In other words, a knowledge structure theory must make a commitment to particular content schemas.

The commitment to particular content is a policy we follow consistently throughout the book. Whether we are talking of scripts, plans, goals, themes, etc., we try whenever feasible to lay out the particulars of members of these conceptual categories. This is the same policy as was followed by the first author in developing Conceptual Dependency theory (Schank, 1972) to describe individual actions.

There has been much debate over whether the conceptual primitives of CD theory are the 'right' primitives, and some criticism that the theory is ad hoc. For many purposes, however, the important criterion is whether the theory is useful. Further, we would argue that any theory proposed as a replacement will have to come to grips with the same content issues as CD theory, and will more than likely end up with much the same primitives (as did Norman and Rumelhart (1975) for example). Indeed, the systematic linguistic exploration by Jackendoff (1976) of candidates for primitives seems to point in this direction.

We anticipate that there will be similar debate about the primitives we will propose in this book for higher-level knowledge structures. We will not be dogmatic about particular primitives, however, knowing that revisions in our scheme will no doubt be necessary as psychological validations and unanticipated theoretical considerations come along.

1.4 Conceptual Dependency Theory

In order to understand what follows in this book it is helpful to have a rudimentary exposure to Conceptual Dependency Theory. The theory has been described at length elsewhere (see especially Schank, 1975); we need not go into it in much detail here.

Conceptual Dependency (henceforth CD) is a theory of the representation of the meaning of sentences. The basic axiom of the theory is:

A For any two sentences that are identical in meaning, regardless of language, there should be only one representation.

The above axiom has an important corollary that derives from it.

B Any information in a sentence that is implicit must be made explicit in the representation of the meaning of that sentence.

These two rules have forced us to look for one economical form for representing meaning. In doing so, we have invented the initial framework:

C The meaning propositions underlying language are called conceptualizations. A conceptualization can be active or stative.

D An active conceptualization has the form:
Actor Action Object Direction (Instrument)

E A stative conceptualization has the form:
Object (is in) State (with Value)

The form that we postulate for conceptualizations has led us to the principle of primitive actions. That is, because a conceptualization is defined as an actor doing something to an object in a direction, we have had to determine just what an actor can do. Clearly, Principle A forces us to look closely at actions that seem similar to see if we can extract the essence of their similarity. Principle B forces us to make explicit whatever differences there might be between two actions and to express them accordingly. For example, two verbs in a language may share a similar primitive element (as 'give' and 'take' share the primitive element **TRANSFER of POSSESSION**) but also have differences. The best representation for our purposes for a given verb then, will be the primitive element it shares with other verbs, plus the explicitly stated concepts that make it unique. As it happens, these explicitly stated concepts also turn out to share similar elements with other verbs, so that often a verb is represented as a particular combination of primitive actions and states none of which are unique to that verb, but whose combination is entirely unique. (Many verbs are represented entirely by states with no primitive act used at all.)

The primitive acts of Conceptual Dependency are:

ATRANS The transfer of an abstract relationship such as possession, ownership or control. Thus, one sense of 'give' is: **ATRANS** something to someone else; a sense of 'take' is: **ATRANS** something to oneself. 'Buy' is made up of two conceptualizations that cause each other, one an **ATRANS** of money, the other an **ATRANS** of the object being bought.

PTRANS	The transfer of the physical location of an object. Thus, 'go' is **PTRANS** oneself to a place; 'put' is **PTRANS** of an object to a place.
PROPEL	The application of a physical force to an object. **PROPEL** is used whenever any force is applied regardless of whether a movement (**PTRANS**) took place. In English, 'push', 'pull', 'throw', 'kick', have **PROPEL** as part of them. 'John pushed the table to the wall' is a **PROPEL** that causes a **PTRANS**. 'John threw the ball' is **PROPEL** that involves an ending of a **GRASP** ACT at the same time. Often words that do not necessarily mean **PROPEL** can probably infer **PROPEL**. Thus, 'break' means to **DO** something that causes a change in physical state of a specific sort (where **DO** indicates an unknown ACT). Most of the time the ACT that fills in the **DO** is **PROPEL** although this is certainly not necessarily the case.
MOVE	The movement of a body part of an animal by that animal. **MOVE** is nearly always the ACT in an instrumental conceptualization for other ACTs. That is, in order to throw, it is necessary to **MOVE** one's arm. Likewise **MOVE** foot is the instrument of 'kick' and **MOVE** hand is often the instrument of the verb 'hand'. **MOVE** is less frequently used noninstrumentally, but 'kiss', 'raise your hand', 'scratch' are examples.
GRASP	The grasping of an object by an actor. The verbs 'hold', 'grab', 'let go', and 'throw' involve **GRASP** or the ending of a **GRASP**.
INGEST	The taking in of an object by an animal to the inside of that animal. Most commonly the semantics for the objects of **INGEST** (that is, what is usually **INGEST**ed) are food, liquid, and gas. Thus, 'eat', 'drink', 'smoke', 'breathe', are common examples in **INGEST**.
EXPEL	The expulsion of an object from the body of an animal into the physical world. Whatever is **EXPEL**ed is very likely to have been previously **INGEST**ed. Words for excretion and secretion are described by **EXPEL**, among them, 'sweat', 'spit', and 'cry'.

MTRANS	The transfer of mental information between animals or within an animal. We partition memory into two pieces: The **CP** (conscious processor where something is thought of), and the **LTM** (long term memory where things are stored). The various sense organs can also serve as the originators of an **MTRANS**. Thus, 'tell' is **MTRANS** between people, 'see' is **MTRANS** from eyes to **CP**, 'remember' is **MTRANS** from **LTM** to **CP**, 'forget' is the inability to do that, 'learn' is the **MTRANS**ing of new information to **LTM**.
MBUILD	The construction by an animal of new information from old information. Thus, 'decide', 'conclude', 'imagine', 'consider', are common examples of **MBUILD**.
SPEAK	The actions of producing sounds. Many objects can **SPEAK**, but human ones usually are **SPEAK**ing as an instrument of **MTRANS**ing. The words 'say', 'play music', 'purr', 'scream' involve **SPEAK**.
ATTEND	The action of attending or focusing a sense organ towards a stimulus. **ATTEND** ear is 'listen', **ATTEND** eye is 'see' and so on. **ATTEND** is nearly always referred to in English as the instrument of **MTRANS**. Thus, in Conceptual Dependency, 'see' is treated as **MTRANS** to **CP** from eye by instrument of **ATTEND** eye to object.

Some set of primitive ACTs is essential for representing meanings, especially if sentences that have the same meaning are going to be be represented in only one way. The ACTs presented here are not category names for verbs. They are the elements of action. An analogous situation is the formation of compounds from the elements in chemistry.

The use of such primitives severely reduces the inference problem (see Schank, 1975), since inference rules need only be written once for any ACT rather than many times for each verb that references that ACT. For example, one rule is that if you **MTRANS** something to your LTM, then it is present there (i.e., you know it). This is true whether the verb of **MTRANS**ing was 'see', 'hear', 'inform', 'memorize', or whatever. The inference comes from the ACT rather than the verb.

Conceptualizations that are attribute-value statements make use of a large number of SCALES. These scales run between boundaries which by convention are labeled -10 to 10. Scales are useful for indicating changes in state. Some of the scales we use, with their boundaries and some steps in between, are indicated below. In current applications of Conceptual Dependency Theory, it has not been necessary to undertake a serious quantitative scaling of relative points along the -10 to 10 continuum. At present, the occasional numerical references are only used suggestively.

HEALTH (dead, diseased, under the weather, tolerable, in the pink)

ANTICIPATION (terrified, nervous, hoping, confident)

MENTAL STATE (broken, depressed, all right, happy, ecstatic)

PHYSICAL STATE (end of existence, damaged, OK, perfect)

AWARENESS (dead, unconscious, asleep, awake, keen)

The symbol ⇑ denotes causality. Some example sentences and their representations are:

John killed Mary.

John **DO**

⇑

Mary **HEALTH**(-10)

John kicked Mary.

John **PROPEL** foot **to** Mary

⇑

foot(John) **BE PHYSICAL CONTACT**(Mary)

John told Mary that Bill was happy.

John **MTRANS**(Bill **BE MENT.ST**(5)) **to** Mary

John read a book.

John **MTRANS**(Information) **to LTM**(John) **from** book
inst(John **ATTEND** eyes **to** book)

In the original development of Conceptual Dependency theory, we spent most of our effort on representation of verbs and states. The bulk of Chapter 3 is one answer to the question of how to represent nouns. How does one represent a restaurant? Is it 'a place where people eat'? Or 'a place you go to eat where someone serves you and you pay'? How far do you go in such a representation? Scripts, although invented to handle a different but related problem, form the basis of the answer to the representation of certain complex nouns as well. How to represent concrete nouns is discussed briefly when we deal with memory in the next section.

Other researchers in artificial intelligence have much discussed the primitive actions that we have developed. Many of them seem to adopt one or more of them for their purposes, while usually rejecting either the rest of the set or the principle that it is necessary to represent sentences at the level of primitive actions each and every time. The most often heard suggestion is that one should only 'break down words into primitives when necessary'.

When is it necessary to break down a sentence into its minimal meaning units? The answer is simple enough: only when you need to exploit the 'meaning' itself. It is not necessary for word association tasks, for microworlds where there is little or no ambiguity or overlap in meaning, or for simple retrieval tasks where the meaning of the elements dealt with is not needed.

If you need to know the meaning of what you are dealing with, then it is necessary to look at the elements that make up that meaning. The only argument to this can be an argument based upon when you break down a sentence, not if you break down a sentence.

The 'when' question seems clear enough to us, although others differ with our position. Since memory ideally stores information in only one way, any pattern matching that needs to be done against information stored in memory requires a canonical form for the information. That is, information in memory must be stored in something like the primitive terms of Conceptual Dependency, and likewise the inference processes that are part of memory must be in those terms.

Should the breakdown into primitives occur after parsing ('when necessary'), or during parsing (assuming it is always necessary)? A good parser should exploit the meaning of a sentence. In understanding it seems doubtful that people first do a syntactic analysis without recourse to meaning and than look at the meaning. People understand as they go. Our parser (Riesbeck, 1975) has been quite successful using predictions that it generates based upon the kinds of meanings that it expects. Since it is hard to find a case when such breakdown is not necessary (in a real and complex system), we see little choice but to 'break down the words' every time.

One exception to this has occurred as a result of this book. In Chapters 4-7, various Knowledge Structures are introduced as an adjunct to Conceptual Dependency. We are beginning to find that it is sometimes better to parse directly into our Knowledge Structure representation rather than going by way of Conceptual Dependency. Thus, for example, the word 'want', which seemed primitive enough but was not so treated in Conceptual Dependency, is primitive in Knowledge Structures. It is reasonable with such words to go directly to where we want to be, thus bypassing Conceptual Dependency. This is, in fact, a complaint sometimes made about our work, namely that at the highest memory levels it will be necessary to reorganize information at places other than the primitive actions and thus we will have to 'unbreak down' again. The Knowledge Structure representation that we develop should answer this complaint.

1.5 Memory

Before we get into the substance of this book, it is worthwhile to introduce one more issue, namely memory. For a long time, the problems of natural language processing seemed to be separate from the problems of memory. Recently, Quillian (1968), Anderson and Bower (1973), Rieger (1975), Norman and Rumelhart (1975), and others have made it quite clear that memory and language are inextricably bound together. However, while the importance of dealing with memory has been generally agreed upon, the form that memory takes is still at issue. This book is, in a sense, entirely about memory. We are arguing here for certain theoretical entities that must form the basis of human memory organization.

The form of memory organization upon which our arguments are based is the notion of episodic memory. An episodic view of memory claims that memory is organized around personal experiences or episodes rather than around abstract semantic categories. If