American Journal of Computational Linguistics Microfiche 79

CONTEXTS OF LANGUAGE

DAVID L. WALTZ, EDITOR.

Coordinated Science Laboratory
University of Illinois
Urbana 61801

Papers presented in two sessions of TINLAP-2, the 1978 Meeting of the Association for Computational Linguistics, held with joint sponsorship by the Association for Computing Machinery and its Special Interest Group in Artificial Intelligence.

TABLE OF CONTENTS

FOCUSING IN DIALOG[1]

Barbara J. Grosz
Artificial Intelligence Center
SRI International, Menlo Park, California 94025

A. Introduction

When two people talk, they focus their attention on only a small portion of what each of them knows or believes. Not only do they concentrate on particular entities (objects or relationships), but they do so using particular perspectives on those entities. In choosing a particular set of words with which to describe an entity, a speaker indicates a perspective on that entity. The hearer is led, then, to see the entity more as one kind of thing than as another. For example, a single building may be viewed as an architectural wonder, a house, or a home, and a single event may be viewed at one time as a selling, another as a buying, and still another as a trading. Some entities are central to the dialog at a certain point and hence are focused on more sharply than others. More importantly, much of what each participant knows is not clearly in view at all; it is not considered by the speaker in choosing what to say or how to say it, or by the hearer in interpreting an utterance.

Focusing is an active process.[2] As a dialog progresses, the participants shift their focus to new entities or to new perspectives on entities previously highlighted by the dialog. Furthermore, an actor is involved in focusing (as the term is used in this paper): if an entity is in focus, it is the object of someone's focusing; it cannot be impersonally in focus. When I use the constructions "highlighted", "focused on", or "in focus", there is always an implicit actor doing the highlighting or focusing. Finally, the entities that the speaker and hearer focus on are entities in their (external) shared reality. Focusing, then, is the active process, engaged in by the participants in a dialog, of concentrating attention on, or highlighting, a subset of their shared reality.

The relationship between language and focusing is two-way: what is said influences focusing; what is focused on influences what is said. The speaker provides clues for the hearer both to what s/he is currently focused on and to what s/he wants to focus on next. These clues may be linguistic or may derive from shared linguistic or nonlinguistic knowledge. The hearer depends on shared beliefs about what entities are highlighted to interpret such things as the appropriate sense of a particular word and the object or event corresponding to a definite description. The link between the entities discussed in an utterance and the entities focused on when the utterance is spoken is thus an important aspect both of producing and of interpreting that utterance.

The use and interpretation of definite descriptions in dialog demonstrate the importance of focusing to dialog participants.[3] This paper examines the relationship between focusing and definite description and the implications of this relationship for computer systems for dialog understanding. Section B presents an example that illustrates this relationship. Section C discusses definite descriptions from both the speaker's and the hearer's perspectives and presents problems that arise for both participants whose solutions are influenced by how the participants are focused. Section D addresses some problems that arise in computationally capturing the notion of focusing and discusses other aspects of dialog with which focusing mechanisms must be coordinated in a natural language processing system, in order to handle the problems introduced in the preceding sections.

B. An Example

To begin, I want to examine a sample dialog between two people, an expert and an apprentice, cooperating to complete a task. It illustrates several important aspects of the role of focusing in communication. The sample comes from a corpus of task-oriented dialogs collected in situations simulating direct interaction between a person and a computer (Grosz, 1977; Deutsch, 1974). The particular task being performed is disassembly of an air compressor.

(1) E: First you have to remove the flywheel.
(2) A: How do I remove the flywheel?
(3) E: First, loosen the two allen head setscrews holding it to the shaft, then pull it off.
(4) A: OK.
(5) I can only find one screw. Where's the other one?
(6) E: On the hub of the flywheel.
(7) A: That's the one I found. Where's the other one?
(8) E: About ninety degrees around the hub from the first one.
(9) A: I don't understand. I can only find one. Oh wait, yes I think I was on the wrong wheel.

---

[1] The work reported herein was supported by the National Science Foundation under Grant No. MCS 76-22004 and by the Advanced Research Projects Agency of the Department of Defense under Contract No. N00039-78-C-0060. I would like to thank Jerry Hobbs, David Levy, Ann Robinson, Jane Robinson, Candy Sidner, and Brian Smith for discussing the ideas in this paper and commenting on various drafts of it.

[2] This is the reason the verb "focusing" rather than the noun "focus" is used most often in this paper.

[3] Although I will concentrate on dialog, much of what I have to say carries over to other forms of discourse.

[4] For most of these dialogs the expert and apprentice had only limited visual contact.

(10) E: Show me what you are doing.
(11) A: I was on the wrong wheel and I can find them both now.
(12) The tool I have is awkward. Is there another tool that I could use instead?
(13) E: Show me the tool you are using.
(14) A: OK.
(15) E: Are you sure you are using the right size key?
(16) A: I'll try some others.
(17) I found an angle I can get at it.
(18) The two screws are loose, but I'm having trouble getting the wheel off.
(19) E: Use the wheelpuller. Do you know how to use it?
(20) A: No.
(21) E: Do you know what it looks like?
(22) A: Yes.
(23) E: Show it to me please.
(24) A: OK
(25) E: Good. Loosen the screw in the center and place the jaws around the hub of the wheel, then tighten the screw onto the center of the shaft. The wheel should slide off.

First, consider the use of the phrase "the two screws" in (18) to refer to the two setscrews holding the pulley on its shaft and the use of the phrases "the screw in the center" and "the screw" in (25) to refer to a part of the wheelpuller.[5] Since most objects do not have proper names, definite descriptions are a primary means of identifying objects. However, as in this dialog, the same description may be used to identify different objects at different times. When (25) was uttered, the two screws mentioned in (3) through (18) were the most recently mentioned objects that could be referred to by a phrase such as "the screw", but they were no longer focused on by the dialog participants -- they were no longer relevant to either the dialog or the task -- and hence were not considered as possible referents for either "the screw in the center" or "the screw" in (25).

One can see in this example that the most recently mentioned object that satisfies a description may not be the object identified by that description. What entities a speaker and hearer are focused on influences both the kinds of descriptions they use and how their descriptions are interpreted. In utterance (3), the expert indicates that he is focused on, and concurrently gets the apprentice to focus on, the two subtasks involved in removing the pulley. In particular, the two allen head setscrews involved in the first task are brought into focus; they continue to be in focus through the first part of (18). The initial clause of (18) indicates the completion of the task involving the screws and hence suggests that the apprentice will shift her attention to some new task (she might not -- she could still say something more about the screws). She does make such a shift in the second clause of (18) ("but I'm having trouble getting the wheel off"). In (19), the expert indicates that he has followed this shift (note that he might have asked a question about the screws -- e.g., "How loose are they?" -- and thereby continued to focus on them and the associated task) and narrows focusing from the task of removing the flywheel to a particular tool involved in that task. In this context, it is clear that the phrase "the screw" cannot refer to either of the setscrews, but must refer to something else.[6]

This dialog also indicates some of the ways in which focusing is manipulated in a dialog. In particular, it illustrates how the structure of the entities being discussed (the 'domain') influences focusing and hence the structure of the discourse. The dialog concerns the performance of a task; its topic is that task. As a result, the way in which the apprentice and expert focus, and hence the structure of the dialog, are closely linked to the structure of the task. Information about the structure of entities in the domain provides one kind of clue to how focusing can change. What about general linguistic clues to focusing? What information in words themselves or in sentence structure can influence focusing? The use of "but" in (18) illustrates one kind of linguistic clue to focus. The indication of contrast suggests a shifting of focus to the entities described in the clause following the "but". In fact, this shift does occur and the remainder of the fragment concerns things involved with "getting the wheel off"[8]

The final point I want to make with respect to this fragment concerns the relationship between how the speaker and hearer are focused and how differences in focusing affect understanding. It is clearly crucial for speaker and hearer to be able to distinguish their own beliefs from each other's. What about focus? I am concerned here not with the consistent difference in focusing

---

[5] The modifying phrase "in the center" does not distinguish the main wheelpuller screw from the setscrews, but from other screws that are part of the wheelpuller.

[6] It is interesting that some people who are not familiar with the compressor or wheelpuller find this sequence confusing: (18) seems to end any concern with screws and hence (25) is unintelligible. One must know -- or infer -- that the wheelpuller has a screw for the statement to make sense.

[7] The concept of structure used here is similar to that in Levy (1977), but different from that in work on story and text grammars (cf. vanDijk 1972; Rumelhart 1975). In particular, I am not interested in such things as generating or recognizing a valid dialog (the analogy to sentence grammars), but rather in those dynamic aspects of intersentential relationships such as focusing that influence the interpretation and generation of utterances in a dialog.

[8] One of the key open problems for incorporating focusing mechanisms in natural language processing systems is identifying the different kinds of clues to focusing and how they interact. Some aspects of this problem are discussed in Section D.

that results from the speaker being one step ahead of the hearer (closing this gap is one goal of an utterance), but rather with whether speaker and hearer purposely maintain differences in focusing over several interactions (as they do with beliefs). An analysis of the dialogs we collected indicates that, in most cases, whether or not a speaker and hearer are focused similarly, they speak as though they were. Speaker and hearer assume a common focus; they usually do not have distinct models of each other's focus. That is, the speaker assumes that the hearer in understanding an utterance has followed any shift in focus indicated by that utterance and is, to the extent it matters, focused on the entities the speaker intended (from the perspective the speaker intended). It is only when a difference in focusing results in some fairly major incompatibility that a problem is detected. The interchange in (5) through (11) illustrates what happens when the two participants in a dialog believe erroneously that they are focused on the same entity. Initially, the apprentice is focused on the motor pulley, which she thinks is the flywheel. Because the expert is not aware of this (he probably doesn't even consider the possibility), his responses are not very helpful.

C. Descriptions

One of the key ways in which the influence of focusing on dialog is manifest is in the definite descriptions used. There is a two-way interaction between definite descriptions and focusing: what entities a speaker and hearer concentrate on (and from what perspectives) influences how they describe entities, and how entities are described influences how the speaker and hearer continue to focus their attention. Two specific problems relating to descriptions are strongly influenced by focusing. From the speaker's perspective, there is the problem of what to include in a description. From the hearer's perspective, there is the problem of what to do when a description doesn't correspond to any known entity, when it doesn't "match" anything.

1. Generating Descriptions

Three factors that influence the production of a description are: the information speaker and hearer share about the entity being described, the perspectives they have on it, and the use of redundancy. The following fragment of dialog illustrates the first two of these factors.[9]

E: OK. Now we need to attach the conduit to the motor. The conduit is the covering around the wires that you . . . were working with earlier. There is a small part . . . oh brother
A: Now wait a s . . . the conduit is the cover to the wires?
E: Yes and . . .
A: Oh I see, there's a part that . . .a part that's supposed to go over it.
E: Yes.
A: I see . . . it looks just the right shape too. Ah hah! Yes.
E: Wonderful, since I did not know how to describe the part.

The problem that arises here is that there is no simple shape-based description for the object the expert needs to identify, so he must find some other shared information on which to base his description (cf. Downing, 1977; Chafe, 1977). The problem is complicated because the expert and apprentice do not share a visual field. If they did, the expert could point (if they and the object being pointed at were all in the same location) or use relative location (e.g., "it's next to the red-handled screwdriver").[10] The expert's solution in this case is to anchor the description on the basis of a past action the apprentice performed and then to describe the object functionally (i.e., to describe its function rather than its shape). Functional descriptions often enable bypassing other more complex descriptions. The statement "it is used for doing x" or "it has the right shape for doing x" may be used to communicate complex shapes and structures. As always, the success of such descriptions depends on the hearer's ability to determine what such an object is like, or to pick out the object from a set.

The fragment also illustrates the problems that arise when two participants in a dialog have different perspectives on what is being described. The expert's orientation is basically functional; he has a model of what is going on, of how the compressor works, and of how it goes together. His descriptions are based on this model. The apprentice's orientation is basically visual or shape-based. He can see the parts and can tell by trying whether they fit. This discrepancy is even clearer in the following fragment, where from the functional perspective of the expert we get the descriptions "pump" and "cooling fins", while from the shape-based perspective of the apprentice, the same objects are described as "thing with flanges" and "little ribby things":

E: Remove the pump and the belt.
A: Is this thing with flanges on it the pump?
E: Point at "the thing with flanges on it" please.
A: I'm pointing at the thing with flanges on it. These little ribby things are flanges.
E: Yes, the thing you are pointing at is the pump. The little ribby things are cooling fins.

In this fragment, one can see the expert and apprentice working toward a shared view, trying to

---

[9] This segment also illustrates the cooperative nature of task-oriented dialogs: the two participants work together to achieve a shared goal of identifying the object the expert wants the apprentice to locate.

[10] Rubin (1978) describes spatial and temporal commonality between speaker and hearer as two dimensions along which language experiences may differ and considers how these dimensions affect the interpretation of deictic expressions.

establish, or check that they have established, a common referent and hence a common focus.[11] An implicit goal in a dialog is to establish this commonality -- the effort this requires is very clear here. One of the ways in which misunderstandings arise is when the participants in a dialog fail to establish this commonality but think they have (this happened with the flywheel and motor pulley in the initial dialog fragment). Not only do such mismatches occur, they are difficult to detect and often go unnoticed until a fairly major problem arises.

A further problem that arises in producing a description is deciding how much information to include in it. The linguistic description of an object must distinguish it from all others currently focused on by the speaker and hearer.[12] But the situation is more complicated than this. It is clear from an analysis of the task-oriented dialogs and from other data (Freedle, 1972) that the description of an object seldom contains only the minimal amount of information necessary to distinguish it. Descriptions, like the rest of language, are often redundant.[13] What appears to be the case for physical objects is that the speaker describes an object not in the minimum number of 'bits' of information, but rather in a manner that will enable the hearer to locate the object as quickly as possible. Clear distinguishing features (e.g., color, size, and shape) are part of a description precisely because they eliminate large numbers of wrong objects and hence help the hearer to isolate the correct object more quickly.

The use of redundant information (and not just distinguishing information) to speed up the search for a referent can be seen easily from an example. If someone asks What tool should I use?" the response "The red-handled one." may not be satisfactory even if there is only one red-handled tool, because processing such a description requires considering too many alternatives. The phrase the red-handled screwdriver is more helpful, because it limits the search to screwdrivers. In giving a description that minimizes the time it takes the hearer to identify the referent of a referring expression, a balance must be reached. Too much information is as harmful as too little, since all parts of the description must be processed to make sure the object is the correct one. Furthermore, the hearer may wonder whether he is mistaken if he thinks he has determined the referent but there is more description to process. (cf Grice, 1975). Using the phrase, "the red-handled screwdriver with the small chip on the bottom and a loose handle" to identify the only red-handled screwdriver will probably both increase the hearer's search time and confuse him. Rather than minimize either the communication time (including processing of the description) or the search time alone, the combination of communication time and search time must be minimized. A speaker should be redundant only to the degree that redundancy reduces the total time involved in identifying the referent.

---

[11] There is a clear indication at the end of the previous fragment that the expert realizes the importance of shape in the apprentice's orientation: he says he didn't know how to describe the part, apparently meaning that he didn't have a description of its shape (he did describe it functionally and in fact that seems to have worked just fine).

[12] Olson (1970) has shown that the description of an object changes depending on the surrounding objects from which it must be distinguished. For example, the same flat, round, white object was described as "the round one" when a flat, square object of similar size and material was present but as "the white one" when a similarly shaped but black object was present. The importance of contrast for distinguishing objects is well established in vision research (e.g., Gregory, 1966). Comparison of differences has also played a crucial role in computer programs that reason analogically (Evans, 1963; similar strategies are used in Winston, 1970).

[13] Olson, 1970, p.266, comments on this phenomenon and on the need for further investigation of it.

Matching a Description

As the preceding discussion illustrates, a major role of descriptions is to point; the speaker is directing the hearer's attention to some entity. For the hearer, focusing is crucial in providing a small set of items from which to choose that entity. Being able to so restrict attention is necessary both for identifying the correct referent (as the interpretation of the phrase "the screw" in the initial dialog fragment illustrates) and constraining search time (see Grosz 1977).

One problem that arises for a hearer, especially a computer system in the role of hearer, is what to do when a reference does not correspond to (or match) any known entity. If the description suffices to distinguish the entity being pointed at from others that are currently focused on, then the mismatch does not matter. But, what does "suffice to distinguish" mean? The question of what kind of mismatch is significant depends on more than the entities in focus. For example, the difference between yellow and green may not matter when a yellow-green shirt is being distinguished from a red one; it does matter when picking lemons.

In addition, the hearer must decide whether or not an inexact match should even be considered. In the usual use of definite descriptions, to identify some entity in the domain of discourse, inexact matches are always acceptable. Donellan (1966) distinguishes this referential use from an attributive use for which an inexact match is not possible: "In the attributive use, the attribute of being the so-and-so is all important, while it is not in the referential use" (p.102). But the distinction in the terms that Donnellan makes it poses a problem for a hearer, since it is the speaker's intent and not the speaker's beliefs[14] that distinguishes

attributive from referential uses of a description. This means that the hearer (whether a person or a computer system) must be able to detect this intent. In certain cases (for example descriptions of entities that do not yet exist), the attributive use is usually clear. In using the phrase, "the winner of the 1979 Nobel Peace Prize", a speaker is describing a person whose identity is not yet known; there is no other way to describe that person (yet).[15] There are other instances in which the distinction relies on knowledge outside the dialog in which the reference occurs (in particular, what the hearer believes the speaker wants). It seems that for this problem the dialog participants must rely on the potential for clarification available in further dialog. If a hearer misinterprets an attributive use of a description, the speaker can explicitly indicate the need for an exact match.[16]

To summarize, the importance of focusing to both the interpretation and the generation of definite descriptions comes from the highlighting function it serves. By separating those items currently highlighted from those that aren't, focusing provides a boundary around the entities from which the entity being either described or identified must be distinguished. For generation purposes, this boundary circumscribes those items from which the entity being described must be distinguished, and thus provides some means of determining when a description is complete enough. It is useful for interpretation in providing a small set of items from which to choose. If an exact match cannot be found in focus, it is reasonable to ask if any of the items in focus comes close to matching the definite description and if so, which is the closest.

D. Focus in Discourse: Prospects and Problems

The major implication of the role of focusing in dialog for a natural language processing system is that such a system needs mechanisms for focusing. In particular, suppose the system has a knowledge base which encodes the portion of the world the system knows about, and that this knowledge base contains formal elements which stand for entities in that world. Then the system needs a means of highlighting those elements in its knowledge base that correspond to the entities currently focused on and must be able both to use this highlighting (for example, to interpret and generate descriptions) and to change it appropriately as the dialog progresses. This section presents several issues that arise in constructing such a computational model and for each discusses what structures and procedures are needed and what research issues must be resolved.

Grosz (1977) describes focusing mechanisms incorporated in a computer system for understanding task-oriented dialogs. These include structures for highlighting elements of a knowledge base, operations on those structures, procedures that use them for interpreting definite noun phrases, and procedures for updating them. The implementation provides for two kinds of highlighting, explicit and implicit and uses task information to determine shifts in focus. An explicit focus data structure contains those elements that are relevant to the interpretation of an utterance because they have been discussed in the preceding discourse. In addition, the focusing mechanisms provide for differential access to certain information associated with these elements. In particular, the subactions and objects involved in a task are implicitly highlighted whenever that task is highlighted. That is, implicit focus consists of those elements that are relevant to the interpretation of an utterance because they are closely connected to task-related elements in explicit focus.[17]

There are several directions in which these mechanisms must be extended for a system to be able to handle the general problems posed by focusing and definite descriptions in dialog. First, the only clues to how focusing changes that have been incorporated in the system are clues based on shared knowledge about the structure of entities in the domain (in particular, the structure of the task); linguistic clues and the interaction between different kinds of clues remain to be examined. Second, the highlighting of explicit and implicit focus are used in interpreting definite descriptions, but an exact match is required; the question of what constitutes an inexact match has not yet been faced. Third, although the highlighting structures provide for focusing on different aspects of an entity, the deduction routines do not use this information in accessing information about an entity in focus. Finally, the question of how the focusing mechanisms interact with representations of belief has not been addressed. The following sections examine the problems posed by each of these extensions in more detail.

---

[14] "A definite description can be used attributively even when the speaker believes that some particular person fits the description, and it can be used referentially in the absence of this belief."(p. 111)

[15] There is, of course, the possibility that the speaker meant to say 1977, in which case s/he is referring (wrongly) to an existing entity, but then we are back with the referential case.

[16] I have ignored a third issue that arises when considering a computer system for natural language processing: the formalism used for encoding knowledge in the system must be adequate for handling attributive descriptions. For a discussion of this issue, see Cohen, 1978 and Webber, 1978.

[17] Elements in implicit focus are separated from those in explicit focus for two reasons. First, there are numerous entities implicitly focused on in a dialog, many of which are never referenced. Including the elements corresponding to such entities in the explicit focus data structure would clutter it, weakening its highlighting function. Second, references to implicitly focused entities may indicate a shift of focus to those entities, making it useful to distinguish such references from others.

Fillmore, 1977[20]).

The perspective from which an entity is viewed influences how further information about that entity is accessed. The representation of focus presented in Grosz (1977) allows for differential access to properties of an entity, but this addresses only one part of the problem.[21] Using the initial perspective from which an entity is viewed for differential access does not rule out considering a concept differently from the way it has already been portrayed. Instead, it orders the way in which aspects of the concept are to be examined. One of the problems this raises is deciding when to consider a switch in perspective, when to abandon deriving properties or searching items implicitly focused by an initial perspective and examine other aspects of the entity.

Another problem that relates to perspective is how perspective influences the particular description a speaker chooses. Does global focus give an indication to a speaker of which properties to choose? The preceding fragments of dialog contained several examples that illustrated the effect of differences in how a speaker and hearer were focused on communication. This suggests that focusing, though often quite useful, can cause problems for people; similar problems may be unavoidable in a natural language processing system.

4. Focusing and Beliefs

An additional aspect of focus that has not yet been addressed is its interaction with a representation of beliefs. The dialog fragments in the section on description pointed out some of the problems that arise when the two participants know different things about the entity being described It is important, then, for a speaker to be able to separate his own beliefs from what he believes his hearer knows or believes. It seems equally clear from the dialogs, however, that focusing is not one of the things that is separate for the two participants. There is a pervasive assumption by speaker and hearer that they share a common focus (this is. in fact, an important part of how and why focusing works). The extension that seems to be needed here is to have the focusing mechanisms interact with an encoding of knowledge that distinguishes beliefs (e.g., Cohen 1978) rather than, as is now the case, with some uniform encoding of knowledge that does not distinguish between speaker and hearer.

E. Summary

Focusing is the active process, engaged in by the participants in a dialog, of concentrating attention on or highlighting, a subset of their shared reality. Not only does it make communication more efficient, it makes communication possible. Speaker and hearer can concentrate on a small portion of what they know and ignore the rest. The importance of focusing to communication is clearly demonstrated by the definite descriptions that are used in dialog. For a natural language processing system to carry on a dialog with a person it must include mechanisms that computationally capture this focusing process. This paper has examined the requirements definite descriptions impose on such mechanisms, discussed focusing mechanisms included in a computer system for understanding task-oriented dialog, and indicated future research problems entailed in modeling the focusing process more generally.

---

[20] Fillmore says,

> The point is that whenever we pick a word or phrase, we automatically drag along with it the larger context or framework in terms of which the word or phrase we have chosen has an interpretation. It is as if descriptions of the meanings of elements must identify simultaneously "figure" and "ground".
>
> To say it again, whenever we understand a linguistic expression of whatever sort, we have simultaneously a background scene and a perspective on that scene.

[21] Consequently, the reference resolution mechanisms did not use this feature.

REFERENCES

Chafe, Wallace L. The Flow of Thought and the Flow of Language. In Proceedings of the Symposium on Discourse and Syntax, Los Angeles, California, November, 1977 In press.

Cohen, Philip R. On Knowing What to Say: Planning Speech Acts. Ph. D. thesis, University of Toronto, Canada. 1978.

Deutsch[Grosz], Barbara G. Typescripts of Task Oriented Dialogs. SUR Note 146, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, August 20, 1974.

Donnellan, Keith. Reference and Definite Description. The Philosophical Review, vol. 75, 1966. Reprinted in: Semantics, Danny P. Steinberg and Leon A. Jakobovits, Eds. pp. 100-114. The University Press, Cambridge. 1971.

Downing Pamela A. On "Basic Levels" and the Categorization of Objects in English Discourse. Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society, Berkeley, California, February 1977.

Evans, Thomas G. A Heuristic Program to Solve Geometric-Analogy Problems. Ph. D. thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, May, 1963.

Fillmore, Charles J. The Case for Case Reopened. In: Syntax and Semantics, John P. Kimball, Ed. Academic Press. New York. In press.

Freedle, Roy O. Language Users as Fallible Information-Processors: Implications for Measuring and Modeling Comprehension. In: Language Comprehension and the Acquisition of

Fillmore, 1977[20])

The perspective from which an entity is viewed influences how further information about that entity is accessed. The representation of focus presented in Grosz (1977) allows for differential access to properties of an entity,[21] but this addresses only one part of the problem. Using the initial perspective from which an entity is viewed for differential access does not rule out considering a concept differently from the way it has already been portrayed. Instead, it orders the way in which aspects of the concept are to be examined. One of the problems this raises is deciding when to consider a switch in perspective, when to abandon deriving properties or searching items implicitly focused by an initial perspective and examine other aspects of the entity.

Another problem that relates to perspective is how perspective influences the particular description a speaker chooses. Does global focus give an indication to a speaker of which properties to choose? The preceding fragments of dialog contained several examples that illustrated the effect of differences in how a speaker and hearer were focused on communication. This suggests that focusing, though often quite useful, can cause problems for people; similar problems may be unavoidable in a natural language processing system.

4. Focusing and Beliefs

An additional aspect of focus that has not yet been addressed is its interaction with a representation of beliefs. The dialog fragments in the section on description pointed out some of the problems that arise when the two participants know different things about the entity being described. It is important, then, for a speaker to be able to separate his own beliefs from what he believes his hearer knows or believes. It seems equally clear from the dialogs, however, that focusing is not one of the things that is separate for the two participants. There is a pervasive assumption by speaker and hearer that they share a common focus (this is, in fact, an important part of how and why focusing works). The extension that seems to be needed here is to have the focusing mechanisms interact with an encoding of knowledge that distinguishes beliefs (e.g., Cohen 1978) rather than, as is now the case, with some uniform encoding of knowledge that does not distinguish between speaker and hearer.

E. Summary

Focusing is the active process, engaged in by the participants in a dialog, of concentrating attention on, or highlighting, a subset of their shared reality. Not only does it make communication more efficient, it makes communication possible. Speaker and hearer can concentrate on a small portion of what they know and ignore the rest. The importance of focusing to communication is clearly demonstrated by the definite descriptions that are used in dialog. For a natural language processing system to carry on a dialog with a person it must include mechanisms that computationally capture this focusing process. This paper has examined the requirements definite descriptions impose on such mechanisms, discussed focusing mechanisms included in a computer system for understanding task-oriented dialog, and indicated future research problems entailed in modeling the focusing process more generally.

---

[20] Fillmore says,

> The point is that whenever we pick a word or phrase, we automatically drag along with it the larger context or framework in terms of which the word or phrase we have chosen has an interpretation. It is as if descriptions of the meanings of elements must identify simultaneously "figure" and "ground".
>
> To say it again, whenever we understand a linguistic expression of whatever sort, we have simultaneously a background scene and a perspective on that scene.

[21] Consequently, the reference resolution mechanisms did not use this feature.

REFERENCES

Chafe, Wallace L. The Flow of Thought and the Flow of Language. In Proceedings of the Symposium on Discourse and Syntax, Los Angeles, California, November, 1977. In press.

Cohen, Philip R. On Knowing What to Say: Planning Speech Acts. Ph. D. thesis, University of Toronto, Canada. 1978.

Deutsch[Grosz], Barbara G. Typescripts of Task Oriented Dialogs. SUR Note 146, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, August 20, 1974.

Donnellan, Keith. Reference and Definite Description. The Philosophical Review, vol. 75, 1966. Reprinted in: Semantics, Danny P. Steinberg and Leon A. Jakobovits, Eds. pp. 100-114. The University Press, Cambridge. 1971.

Downing Pamela A. On Basic Levels" and the Categorization of Objects in English Discourse. Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society, Berkeley, California, February 1977.

Evans, Thomas G. A Heuristic Program to Solve Geometric-Analogy Problems. Ph. D. thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, May, 1963.

Fillmore, Charles J. The Case for Case Reopened. In: Syntax and Semantics, John P. Kimball, Ed. Academic Press. New York. In press.

Freedle, Roy O. Language Users as Fallible Information-Processors: Implications for Measuring and Modeling Comprehension. In: Language Comprehension and the Acquisition of

Knowledge, John B. Carroll and Roy O. Freedle, Eds., pp. 169-209. Winston, Washington, D.C., 1972.

Gregory, R. L. Eye and Brain: The Psychology of Seeing, McGraw Hill, New York, 1966.

Grice, H. Logic and Conversation. In: Syntax and Semantics, P. Cole and J. Morgan, Eds. Vol. 3, pp. 41-58. Academic Press, New York, 1975.

Grimes, Joseph E. The Thread of Discourse. The Hague, Mouton, 1975.

Grosz, Barbara J. The Representation and Use of Focus in Dialogue Understanding. Ph. D. thesis, University of California, Berkeley, California; also Technical Note No. 151, SRI International, Menlo Park, California, 1977.

Halliday, Michael A. Notes on Transitivity and Theme in English. Part 2. Journal of Linguistics, 31, 177-274, 1967.

Halliday, Michael A. Language as Code and Language as Behaviour: A Systemic-functional interpretation of the nature and ontogenesis of dialogue. In: Semiotics of Culture and Language, Sydney M. Lamb and Adam Makkai, Eds. 1977. In press.

Halliday, Michael A., and Hasan, Ruqaiya. Cohesion in English. London, Longman, 1976.

Hobbs Jerry R. A Computational Approach to Discourse Analysis. Research Report 76-2, Department of Computer Sciences, City College, CUNY, December 1976.

Levy, David M. Communicative Goals and Strategies: Between Discourse and Syntax. In Proceedings of the Symposium on Discourse and Syntax, Los Angeles, California, November, 1977. In press.

Olson, David R. Language and Thought: Aspects of a Cognitive Theory of Semantics. Psychological Review, 77, 257-273, 1970.

Rubin, A.D. A Theoretical Taxonomy of the Differences Between Oral and Written Language, In: Theoretical Issues in Reading Comprehension, R. Sprio, B. Bruce and W. Brewer, Eds., Lawrence Erlbaum, Hillsdale, N.J., 1978. Also as Center for the Study of Reading Technical Report No. 35, January 1978.

Rumelhart, David E. Notes on a Schema for Stories. In: Representation and Understanding: Studies in Cognitive Science, Daniel R. Bobrow and Alan Collins, Eds. Academic Press, New York, 1975.

Sidner, Candace L. A Computational Model of Co-reference Comprehension in English. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, forthcoming.

van Dijk, Teun A. Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics. Mouton, The Hague, 1972

Walker, Donald E. (Ed.). Understanding Spoken Language. Elsevier North-Holland, Inc., New York, 1978.

Webber, B.L. A Formal Approach to Discourse Anaphora. BBN Report No. 3761, Bolt Beranek and Newman Inc., Cambridge, Massachusetts, May 1978.

Winston, Patrick H. Learning Structural Descriptions From Examples. MAC TR-76. M.I.T. Artificial Intelligence Laboratory, 1970.

Topic Levels

Joseph E. Grimes
Cornell University and
Summer Institute of Linguistics

Now that the sentence is no longer the edge of our world we see more clearly than ever how totally responsive speech is to the situation that calls it forth and to the people involved in it. Bare content is shaped and packaged to meet many requirements at once.

I have tried to sort out two broad categories of these requirements. The first, cohesion, is hearer oriented. It is the influence on form of the speaker's own assumptions about what the hearer knows at each instant of the communication process. The second, staging, is speaker oriented. It reflects how the speaker calibrates the importance of different parts of what he himself intends to say. I find it helpful to tie down the discussion of both cohesion and staging to differences in linguistic form; there may be other psychological or philosophical overtones to cohesion and staging that have no such direct repercussions, but getting at those overtones is another matter.

One of the areas where we are making progress in the linguistic study of discourse is in seeing how speaker and hearer always seek a common ground of reference. This area, however, is hidden in a terminological thicket. Charles Hockett (1959.201) originally identified it as topic, in which, in his terms, 'the speaker announces a topic and then says something about it.' Gundel (1974) has followed this usage, and I think it the best label for now even though it gets confused with topicalization, which may or may not be part of the same package. Much earlier some of the Prague School theorists (summarized in Danes 1974), and later Halliday (1967), used theme for a similar concept, and now Grosz (1977) has used focus for something not very different. Each of these three terms, topic, theme, and focus, has also been used for at least two other kinds of phenomena by reputable linguists, so eventually we are going to have to put together a road map to all the alternatives; but just because the terms are confused is no reason to conclude that no headway is being made.

The idea that I am going to continue to refer to as topic is this: for communication to succeed, speaker and hearer have to establish common ground. This common ground is usually a presumed agreement about the identity of certain objects in the real world. It may also be agreement about certain events or about certain relations that hold. As far as its linguistic expression goes, I think it significant that its formal makeup appears to revolve around nominals most of the time, treating things that are not necessarily objects as if they were.

One reason the common ground phenomenon seems so important is that without its narrowing effect, the hearer might not be able to manage the numerous semantic alternatives that could be developed from each expression in the text constructed by the speaker. Gundel has pointed out the utility in this regard of a formulation attempted by Searle (1969.126):

> For any speaker S, any object X and any predicate P, it is a necessary condition of S's having predicated P of X in the utterance of a sentence containing P, that X should have been successfully referred to in that utterance and all the presuppositions of P should be true of X.

Searle's X is very much like what I am calling the topic, in that unless the hearer can relate to it referentially, he can neither agree nor disagree with whatever else may be said about it.

Gundel illustrates how even an isolated sentence like George ate a plate of shrimp cannot be assimilated as part of a real communication unless the hearer has some way of knowing which of the people named George is being referred to; once he knows that, he can react with yes he did or no he didn't or oh, the sign of a new constellation of information in memory (Winograd 1972).

Grosz has made a useful distinction

between the hearer's memory for concepts, which tends to be long term and global to a text, and the hearer's memory for linguistic form, which tends to be short term and local to a segment of text. Her distinction interlocks with the one Halliday and Hasan (1976) make between reference and substitution. What they call reference identifies concepts, objects outside of language, and even pieces of language itself, as they are mentioned once they have been introduced. What they call substitution includes ellipsis; it refers to the reactivation of stretches of speech from earlier in the text in order to talk about situations that have not already been referred to, but which have enough in common with others that have been referred to that the same linguistic expression can apply to the new situation.

It is the first of these, memory for concept rather than form, from which the speaker appears to take what he hopes is common ground between himself and his hearer. This selection from the field available for global reference is what is behind Gundel's observation that the topic has to be accessible to the hearer. (The most accessible things are characteristically the standard elements of the communication situation: I, you, here, and now.)

In the course of a text, be it monologue or dialogue, the referential common ground that is used as the core of communication may change. This is true of the global topic and apparently of local topics as well. The initial core of reference may be designated very simply, for example by a single noun phrase, with no differentiation of parts or functions at the beginning. Gundel even shows that the topic of some sentences may be implicit, not mentioned in that sentence, but nevertheless to be taken into account if the sentence is to make sense.

Once the topic of a text is put into play, that topic may be developed in at least three different ways that have been described so far in the literature: it may be expanded, shifted, or split. Expansion adds things to the core of reference. Shift adds new referents to the core and leaves others off, so that what is taken as common ground at one point in the text differs from what was taken as common ground earlier. Splitting the core results in local topics being brought into play in relation to global topics; or rather, higher level topics are split into a higher level part and a lower level part, a process which if repeated may yield more than two levels in the same text.

Topic expansion is illustrated in a story from Time magazine (June 21, 1976 p. 56). The date of the issue needs to be taken as the initial topic; there is no other common ground to begin with between the writer and the reader. This is normal in news stories: consider how impossible it is to agree or disagree with anything like The Giants beat the Dodgers until one knows the occasion. The title of this piece is Teton: Eyewitness to Disaster. For a reader who knows his geography, Teton identifies a place, while for one who does not, there is at least a good chance that it is a place name or the name of a person. The idea of referential common ground gives us for starters a reasonable guess at an event that happened the week before the appearance of the magazine, and possibly a place, as a limiting field within which to place the interpretation of the rest of the message.

The text begins with a paragraph set off in italic type and quotation marks, in which the speaker is not identified: "This wet spot on the side of the dam started spurting a little water ." The noun phrase that begins the sentence is definite, as is the dam contained within it. The definiteness here suggests that the writer expects the reader to be able to find the reference because it is accessible within the limits already set. If he follows that suggestion and takes Teton as the name of the dam and accepts this as identifying something new within that field, his reference succeeds. (The side is legitimately definite once we identify the dam by what Halliday and Hasan call lexical cohesion; dams have sides.) As far as pinning down a core of reference is concerned, the text so far has its topic built up as clearly as if the article had begun much more dully, as for example Last week at a place called Teton where there is a dam, a wet spot appeared on its side.

The text goes on "... and I asked my mother, 'Do you think we should notify the authorities?' She said: 'I don't think ...'" Here the person who is making the report is mentioned explicitly for the first time, as is the mother to whom the question is addressed. This complex of the observers, the wet spot on the dam, the location, and the time persists as the referential core or topic through the course of over a column. It is built up by small references to give an expanded topic.

Topic shift differs from expansion in that some referential elements appear to be dropped from the topic as the text progresses; some things that were treated as part of the common ground in the earlier part of the text are not so treated later. Schank (1977) focuses on the intersection of the referential field of one sentence and that of its successor, and tries to define some (but not all) ways in which that intersection relates to the referential field of the next

sentence. His initial definition of topic as 'any object, person, location, action, state, or time that is mentioned in the sentence to be responded to' is probably too inclusive, because it does not take into account Searle's factor of successful reference; but once a text is begun, Searle's boundary condition is no longer needed, because the reference has been established by the text itself.

For Schank a new topic is 'derived from the original input but is not identical to it', in that reference may shift from a specific element mentioned in the earlier sentence to the class of which it is a part in the later sentence, or vice versa. The element in the later sentence may also be a different conceptualization that is like the first in kind; Schank calls this supertopic. It may also be 'a comment that can be inferred from the interaction of two conceptualizations', or metatopic. Schank suggests more specific rules that he hopes will characterize the way topics shift in the course of a text.

The key concept, however, is his observation that a sentence out of context cannot be said to have a topic, because for him the topic arises only out of the interaction of adjacent sentences by the process of intersection. If he is right, or close to right, it is reasonable that some of the things that are treated as topic earlier in the text be given different treatment later in the same text, because they are no longer taken as part of the referential common ground between the speaker and hearer.

The idea of splitting up the referential field embraced by the topic into higher and lower level topics has been treated in two different ways, each of which may be valid in its place. Grosz recognizes the phenomenon, and gives an intriguing example (1977.23) of a pronoun it which refers back to a global topic last mentioned half an hour earlier, even though a whole series of local topics has come between the pronoun and its antecedent. In her discussion of global and local topics, however, she tends to equate the first with memory for concepts and the second with memory for forms, chiefly because she finds the domain of ellipsis to be restricted to a local segment of text, and to always involve memory for forms.

Meyer (1974) and Clements (1976), however, find the global-local phenomenon operating independently of ellipsis or other substitute-like memory for form. They construct a topical hierarchy consisting of a global topic, whatever local topics are talked about as part of the discussion of the global topic, and whatever lower level topics are talked about as part of the discussion of those local topics, in what apparently gives a recursively definable topic tree of unrestricted depth. In this model, psychological tests of recall show that subordinate position in the topic tree regularly gives worse recall than superordinate position.

The definition of topic used by Meyer and Clements is earlier than Gundel's, so that one could expect the variance in their results to be reduced by attention to her principles for recognizing topics. Topic for Clements is more like Halliday's theme, ordinarily the first thing in the sentence. His topic hierarchy comes from three rules:

(i) Topic rule:
Identify the topic of each clause and simple sentence.

(ii) Old/new rule:
Decide whether the topic is new (never previously mentioned) or old (mentioned in an earlier topic or comment). If new -- assign it one level below the previous topic. If old -- assign it the same level as its first mention.

(iii) Coordination rule:
If a topic is coordinated with an earlier topic or comment, assign it the same level as that earlier topic or comment.

The work of Clements and Meyer lends credence to the idea that there may be a hierarchy of topics in a text, all referential in Halliday and Hasan's sense, rather than dependent only on short term memory for form. Some observations from Koine Greek, the vernacular Greek of the first century before Christ to the third century after, appear to bear this out.

Word order is used much less in Koine than it is in English to specify grammatical relations, because the case system of nouns carries that load. One of the functions which word order expresses seems to be that of identifying shifts in topic (Grimes 1975). Noun phrases in the nominative case that precede the main verb of a clause regularly make that nominative the topic; that is, they signal the reader to take it as part of the referential core that is to be the common ground between him and the writer.

The conjunctions of Koine Greek also play a part in the topical structure. There are three kinds of conjunctions: coordinating, subordinating, and resumptive. The coordinating conjunctions include most of the ones translatable as 'and' or 'but'. The subordinating conjunctions include the 'because' and 'if' varieties. The resumptive ara, dio, and occasionally idou mean something like

now back to the main point . They reset the topic level from wherever it was to the global topic.

Working along with the conjunction system is the system of grammatical subordination. It uses relativizers, complementizers, participles, and verbal nouns to signal that certain propositions are peripheral in the author's perspective on what he has to say. All these subordinating grammatical mechanisms are used constantly in the ancient Koine documents, even though it would be perfectly possible to express what they express in strings of independent clauses with no subordination.

The result is texts with a richly elaborated grammatical structure, with some clauses at four or even five levels of subordination. The general distinction between global and local applies, but recursively: at any level of subordination, it appears possible to have yet another level of subordination attached.

When we turn, however, to languages that make distinctions of topic level explicit in their pronominal systems, we find no greater elaboration than a distinction between global and local topics. Bacairi of Brazil (Wheatley 1973) is such a language. The terms Wheatley uses in his description are 'thematic - athematic' and 'focal - nonfocal'; but it seems clear from perusal of his paper that we would now want to call the thematic category the topic, and the focal category global. Someone who has been identified as part of the common ground between speaker and hearer for a text as a whole is referred to by the pronoun maca 'thematic focal animate'; someone who is topic for a local segment but is not the global topic at the same time is referred to by auaca 'thematic nonfocal animate'; someone who is global topic, but is referred to within a stretch that has a local topic active, is mauanca or maunca athematic focal animate' and those who are topic neither at the global level nor at the local level are referred to by uanca 'athematic nonfocal animate' There are inanimate counterparts for all four pronouns.

On the thematic side -- that is, in the pronouns used for topics -- there is a situational or deictic use that confuses the picture. The pronoun inara 'thematic nondeictic animate' is only textually defined, with an anaphoric antecedent that is taken always from the preceding sentence, not from the topical structure of the text as a whole. On the other hand there is a pronoun mira 'thematic deictic animate' which applies only to animate things that can be seen and are near the speaker. In between the nondeictic and the deictic are the situational uses of maca and auaca. The first denotes someone far away but in sight, and the second denotes someone nearer to the speaker but not as near as mira.

There are situations in which these two pronouns appear to flip their reference; what they actually do is change the basis of the reference from the situation to the text. Thus I can begin a text by identifying a boy over there as maca because he is relatively far away, and a woman standing closer as auaca because she is not so far away. But if what I have to say revolves around the woman rather than the boy, I will switch after a few sentences to the textual definition of the pronouns and use maca for the woman because she is more central to what I have to say, and auaca to the boy when I treat him as a local topic. However, we have no information about the use of these pronouns to topicalize at more than two levels at a time.

Longuda of Nigeria has a less elaborate pronoun system with respect to topics than that of Bacairi (Newman, in press), but nevertheless it distinguishes topic from nontopic. There is actually a series of pronoun-aspect particles that appear in sequence in texts to identify actions of the central character and distinguish them from the actions of others. In some parts of a text the character singled out by the pronouns is the one the text is about; he is treated as the global topic. Where local topics are introduced, however, the topic pronoun set switches over to the local topic, and the referent of the global topic is referred to by the same nontopic pronoun as all the other characters.

For example, in a story about a rabbit, most of the story uses the topic pronoun series for the rabbit and a nontopic pronoun for everybody else, including an elephant who interacts with the rabbit. One section, however, is about what the elephant did In that section it is the elephant who gets the topic pronoun series, and the rabbit gets the nontopic series, even if it is the rabbit whom the story is about globally. When the section about the elephant ends, the topic pronoun series reverts to the global topic, the rabbit.

Evidence from languages like these proves the linguistic realism of a distinction between one kind of topic and another. It also agrees in the main with the kinds of analysis we have been making for English; apparently we are fairly close to the right track, in terms of those languages that put this kind of thing right out on the surface What is not yet clear is the number of levels of topic we may deal with in a text: Clements's analysis of English and my analysis of Greek point toward more than two simultaneous levels of topic, while

the languages that distinguish levels of topic in their pronoun systems seem to support two levels.

Bibliography

Clements, Paul. 1976. The effects of staging on recall from prose. Cornell University Ph.D. thesis.

Danes, Frantisek, ed. 1974. Papers on functional sentence perspective (Janua Linguarum series minor no. 147). The Hague: Mouton.

Grimes, Joseph E. 1975. 'Signals of discourse structure in Koine'. In George MacRae, ed. Society for Biblical Literature 1975 Seminar Papers, Vol. 1. Missoula, Montana: Scholars Press, 151-164.

Grosz, Barbara. 1977. The representation and use of focus in dialogue understanding (Technical Note 151). Menlo Park: Stanford Research Institute.

Gundel, Jeannette. 1974. The role of topic and comment in linguistic theory. University of Texas Ph.D. thesis.

Halliday, Michael A. K. 1967. 'Notes on transitivity and theme in English, Part 2 . Journal of Linguistics 3:199-244.

and Ruqaiya Hasan. 1976. Cohesion in English. London: Longman Group Limited.

Hockett, Charles F. 1959. A course in modern linguistics. New York: The Macmillan Company.

Meyer, Bonnie J. F. 1974. The organization of prose and its effects on recall. Cornell University Ph.D. thesis.

Newman, John F. in press. 'Participant orientation in Longuda folktales'. In Joseph E. Grimes, ed. Papers on Discourse. Dallás: Summer Institute of Linguistics.

Schank, Roger. 1977. 'Rules and topics in conversation'. Cognitive Science 1:4.421-442.

Searle, John. 1969. Speech acts. Cambridge: Cambridge University Press.

Wheatley, James. 1973. 'Pronouns and nominal elements in Bacairi discourse'. Linguistics 104:105-115.

Winograd, Terry. 1972. Understanding natural language. New York: Academic Press.

Toward a Rational Model of Discourse Comprehension

Jerry L. Morgan
Center for the Study of Reading
University of Illinois
at Urbana-Champaign

1 Introduction

I begin my tale with the moral: a quotation from the greatest English grammarian, Otto Jespersen (1965)

> The essence of language is human activity--activity on the part of one individual to make himself understood by another, and activity on the part of that other to understand what was in the mind of the first. These two individuals, the producer and the recipient of language, or as we may more conveniently call them, the speaker and the hearer, and their relations to one another, should never be lost sight of if we want to understand the nature of language and of that part of language which is dealt with in grammar. But in former times this was often overlooked, and words and forms were often treated as if they were things or natural objects with an existence of their own--a conception which may have been to a great extent fostered through a too exclusive preoccupation with written or printed words, but which is fundamentally false, as will easily be seen with a little reflexion. (p. 17)

But the temptation to think of language as pure form is great, Jespersen himself slips into this metaphor a few pages later.

> . . . we always find that there is one word of supreme importance to which the others are joined as subordinates This chief word is defined (qualified, modified) by another word, which in its turn may be defined (qualified, modified) by a third word, etc. (p. 96)

But words do not define, modify, or qualify other words. Speakers define, qualify, and modify This confusion is so tempting that it is pervasive in every field that studies language, at any level. It is almost universal in linguistics. We find it, for example, in the following from Halliday and Hasan (1976), who probably know better

> Let us start with a simple and trivial example Suppose we find the following instructions in the cookery book·
>
> > [1 1] Wash and core six cooking apples. Put them into a fireproof dish.
>
> It is clear that them in the second sentence refers back to (is ANAPHORIC to) the six cooking apples in the first sentence. This ANAPHORIC function of them gives cohesion to the two sentences, so that we interpret them as a whole; the two sentences together constitute a text. Or rather, they form part of the same text, there may be more of it to follow.
>
> The texture is provided by the cohesive RELATION that exists between them and six cooking apples. It is important to make this point, because we shall be constantly focusing attention on the items, such as them, which typically refer back to something that has gone before; but the cohesion is effected not by the presence of the referring item alone but by the presence of both the referring item and the item it refers to (p. 2).

There are two serious confusions here. First, words do not refer; speakers refer to things by using words. The word them does not refer to anything at all, obviously so since it can be used to refer to any set one wants to refer to. There is no particular set of entities that one can say the word them refers to. But one can use it to refer to sets of things, when one's intended referent will be recoverable in some way by the hearer.

The second confusion is the idea that words "refer back" to other words. The muddle here is obvious Whether it is people or words that refer, it is things, not (usually) other words, that they refer to. Thus in Halliday and Hasan's example [1:1], it is not the words six cooking apples that them is used to refer to, one is not being instructed to put three words in a fireproof dish. The word them is used to refer to certain apples that were previously referred to by use of the words six cooking apples. My objection to such descriptions is not based merely on a niggling concern with sloppy language. If it were, one might respond that it's clear what Halliday and Hasan mean here, so my complaint is beside the point. Rather, I think the pervasive confusion on just this point is a symptom of a serious comceptual confusion that renders a lot of the related work useless. This is the case with the passage from Halliday and Hasan. They say that it is some relation between sentences in a text that gives it "cohesion", that renders it coherent,

"so that we interpret them as a whole; the two sentences together constitute a text." The relation that gives rise to this cohesion is that them in one sentence "refers back" to the six cooking apples in a previous sentence. If we interpret this phrase charitably, then the question arises, how do we know what them refers to? How do we know that it refers to the apples, and not to two or the writer's bachelor uncles? We can't know such a thing. We can only assume that the writer is rational, and that the recipe is coherent. If it is coherent, we are justified in assuming that it is the apples that are referred to by them. But there is a vicious circularity here. The recipe has cohesion, is a coherent text, just in case them refers to the apples. But we are only justified in inferring that them refers to the apples if we assume that the text is coherent. Thus, in spite of Halliday and Hasan's claim, it is not the anaphoric facts that give rise to cohesion; rather, the assumption that the text is coherent gives rise to the inference that them refers to the apples.

This kind of confusion, it seems to me, arises from the linguist's habit of looking at every aspect of language in terms of linguistic forms and relations between them. Thus in this case the mistaken characterization of reference as a relation between words, and of coherence as a property of an abstract linguistic object called a text. In the rest of this brief paper I want to sketch an opposing view, and to claim that notions like "reference," "text structure," "relevance" and "coherence" are best treated, at least in part, in terms of communicative acts and the plans and goals of speakers/writers who perform such acts.

II. Three Ways of Looking at a Text

Assume for the moment that we know what a text (oral or written) is, and can tell a coherent text from a random transcription of English sentences (I will return to what counts as a coherent text later). Then there are (at least) three kinds of things and relations involved in a text.

1. Sentences. First, conventional wisdom in linguistics has it that texts consist of sentences. I shall accept this for the moment, though a bit later I will show cause to modify it.

But what kind of "thing" is a sentence? It is, if anything is, an abstract linguistic object, a unit of form. It is not a proposition, nor a fact, though it is a means by which such things are asserted, denied, questioned, etc. Nor is a sentence a speech act, though a speech act will usually be performed by means of the utterance of a sentence. But a sentence and an utterance of a sentence are different kinds of things.

A sentence is not the kind of thing that is true or false; "facts," or "propositions," that sentences can be used to express, are true or false. Or perhaps it would be more appropriate to speak of assertions as being true or false. At any rate, it is quite clear that it is nonsense to speak of sentences as true or false, as evidenced by the familiar problem of indexical expressions.

A sentence, then, may be used to assert that something is true, or false, or has occurred, but the sentence itself is not true or false, and does not occur. Thus relations like causation, order in time, entailment, and so on, do not hold between sentences. It is not clear what kind of relation can accurately be said to hold between the sentences of a text.

2. "Facts." The second kind of "thing" involved in a text is what I shall call "facts." (Notice that I do not say texts consist of or contain facts, merely that they somehow involve facts.) The term "fact" is a bit misleading--though I can think of no better term--in that I wish to include as facts events, states, and so forth that do not actually hold in the real world; "propositions," more or less.

Relations among the "facts" involved in a text, then, consist of two classes: first, the same relations that hold between facts in the real world--causation, relations of temporal order, motivation, and so forth; second, those relations that have to do with logic and hypothetical facts, like entailment and contradiction. It may be necessary to distinguish facts on the one hand and propositions on the other, on grounds that relations between facts are of a kind different from relations between propositions, but I will ignore the problem here.

3. Speech acts. The third kind of thing involved in texts is the "speech act" (by this term I mean to include as a sub-case acts of linguistic communication by writing). Speech acts are not sentences, nor "logical forms," nor propositions, *in spite of occasional attempts to define them* in these terms. They are acts, just as the term implies.

Relations between the speech acts involved in a text are just those that can hold between acts in general. First, since an act is a subtype of event, the relations that can hold between events can, in general, hold between acts, thus between speech acts: relations of temporal order, for example. Second, since a speech act is a sub-type of act, relations that can hold between acts can, in general, hold between speech acts. The most important relation in this regard is the relation of purpose: one does such-and-such in order that such-and-such; or one performs a certain act in order thereby to perform a second act. Long chains of these relations can hold between acts. I may throw a switch in order to turn on a light in order to frighten away a burglar in order to save the family jewels. I may tell my friend that there is a charging bull behind him in order that he realize that he is in danger, in order that he get out of the way. It is a mistake to ask whether my speech act was an assertion or a warning, since this presupposes that the two are mutually exclusive. It was both; I asserted something and thereby warned somebody, just as I threw the switch and thereby turned on the lights. I may make a certain mark on a piece of paper, thereby marking my ballot for Smith, thereby casting a vote for Smith. I may

assert that I will do the dishes, thereby volunteering to do the dishes. And so on.

It is commonly the case that acts are linked by complex relations of purpose and goal, including the case where one act is performed by means of performing another act. This is especially true of communicative acts.

There are several subvarieties of speech acts, for which several taxonomies have been proposed; Austin (1962), McCawley (1977), for example. One important distinction in kind is the distinction between the act of saying a sentence, and the act one thereby performs. In performing an act of saying the English sentence "Your hair is in my yogurt" I may, in the right circumstances, thereby inform someone that their hair is in my yogurt. The first kind of act, the act of saying, includes making sounds in a way that conforms to the conventions for what counts as a saying of a sentence, or making visible marks in a way that counts as a saying of a sentence. Texts, then, do not really consist of sentences, but of sayings ("uses") or sentences; or in the case of written texts, of a permanent kind of record of uses of sentences.

III. The Interpretation of Texts

A. Speech acts. The interpretation of a text, then, consists of the interpretation of this record of sayings of sentences. Each saying is interpreted in terms of some speech act(s) performed by saying a given sentence. (Henceforth by "speech act" I will mean the communicative act one performs by saying a sentence, as opposed to the act of saying itself.) There are three aspects to the interpretation of speech acts. the interpretation of what speech acts are performed--assertion, promising, denial, questioning, warning, etc.--by the saying of the sentence; the interpretation of what "facts" are asserted, denied, etc., and the interpretation of the speaker's purpose and goal in performing the speech act.

As an aside I should mention the special instance where nothing is directly asserted, denied, etc.: the case of speech acts of reference. An act of asserting, etc. (for brevity I will henceforth use assertion as representative of all speech acts types), will usually include an act of referring as a subpart; a reference to the entity of which something is asserted. But acts of referring can occur independently. For example, I might say "The door!" to someone under a number of circumstances, to get them to open it, close it, shoot the bad guy standing in it, or merely observe what beautiful hardwood it is made of. It would be a mistake to say that "The door!" means any of these things, or that I have performed (directly) any kind of speech act beyond merely referring. I have only referred to the door, thereby to call my hearer's attention to it, with the expectation that when he turns his attention to the door he will realize what it is I want him to do about it.

The typical immediate goal associated with speech acts of all kinds is the same: that the hearer modify his model of a certain "world" (in the sense of "possible worlds") in a way that involves the "facts" that are asserted, etc. in the speech act. The world involved may be the real world, or, in the case of story-telling, for example, some imaginary world. The modification may include the construction ex nihilo of some hypothetical or imaginary world. The relation between the "facts" of the speech act and the intended modification vary with the nature of the speech act; but in all cases some modification is involved. The simplest case is that of assertion; normally the immediate goal of an assertion is that the hearer modify the world under discussion in a fashion that makes the asserted fact true in that world. In the case of yes-no questions, the goal is that the hearer modify his model of the world such that in that world the speaker wants the hearer to tell him whether the fact questioned is true. In the case of imperatives, the goal is that the hearer modify his model such that in that world the speaker wants the hearer to bring about the truth of the ordered fact, and that certain social consequences will follow from non-compliance.

The raw datum of comprehension, then, is not the sentence or the proposition, but the fact that a certain speech act has occurred. In comprehension, people do not process sentences as abstract formulae; they observe that someone has said something to them, and attempt to interpret that act and its consequences, which may include modification of their model of the world. The process of modifying the model according to what is said is not direct, but the result of several steps of evaluation. Interpretation of an assertion might go roughly like this, from the viewpoint of the hearer (where S is the speaker, A the addressee, addressee and hearer may be identical):

> S has said x to A. Saying x counts as asserting p. S knows that saying x counts as asserting p. S knows that therefore his saying x is likely to be interpreted by A as an assertion of p. S has done nothing to prevent A from making this conclusion. Therefore S's intention is that his saying x be taken by A as an assertion of p. Then if S is sincere S believes that P is true. A must conclude that S has asserted p because he wants A to take p as true and modify his model of the world accordingly.

But the decision to believe p, i.e. modify his model of the world to include p, is a matter of choice on H's part, not an automatic consequence of processing the "sentence." The steps involved in making this decision are equally complex, involving the ability to construct a hypothetical world just like the real one except that p is true, to evaluate the consistency and plausibility of that world, and so on. Some of the facts that are asserted will relate to this decision-making process. For example, in saying (1) my goal is most likely that the hearer come to believe that both facts asserted are true.

(1) John is here. He has a dog with him.

But in the case of (2), I am not so much concerned with the second asserted fact in itself, but with the goal that from concluding that it is true, the hearer will be more likely to believe the first, since I intend that he take the second fact as evidence that my source is reliable.

(2) The world is flat. It says so in the Encyclopedia.

Matters that are sometimes construed as rhetorical relations between sentences fall into this category. Some fact is asserted not because it is important in itself, but because it bears on H's evaluation of some other asserted fact. Thus the relation is not one between sentences, but between speech acts. One speech act is performed in order to influence the interpretation and evaluation of another. At any rate, my point here is that in comprehending a text in the serious sense, comprehension proceeds not from some disembodied abstract object called a "sentence," nor from a "proposition," but from the perceived fact that S has said such-and-such, and that so saying counts as a speech act of a certain type.

There is another way in which modification of the world model is not a direct function of the asserted fact: the widely studied problem of inference. Given the hearer's acceptance of what the speaker has asserted, incorporation of the facts into the model of the world may involve more than merely adding the asserted facts. There is, for example, a general principle of ceteris paribus that comes into play in consideration of alternative worlds. Roughly, when constructing a model of a world alternative to some point-of-reference world (usually the real one), the hearer will assume, lacking evidence (from assertion or inference) to the contrary, that the alternative world is consistent with the point-of-reference world in all relevant respects. To take an extreme example, if someone is telling me about life on Arcturus, I will assume that the laws of physics are the same there as on earth, unless something the speaker says leads me to believe otherwise. In the same way, hearers will assume, lacking counter-evidence, that what is typical in the point-of-reference (e.g. real) world is also typical in the alternative world. They will also assume that things of a given type have the properties typical of things of that type. Gricean rules of conversation support these inferential strategies in the following way: The hearer knows that the speaker knows the hearer is likely to make inferences according to these and other strategies. The speaker has done nothing to prevent the hearer from making them. Therefore the hearer is justified in inferring that the speaker intends for the inference to be made.

Using these and other strategies, then, the hearer modifies his model of one or more worlds, based not on detached sentences or propositions floating in some abstract semantic space, but on his observation that a certain person has performed a certain speech act.

B. Relations among speech acts. But there is more to the interpretation of a text than just the interpretation of individual speech acts. A speech act is performed for some purpose, with some goal in mind. And complete understanding of a text involves the ability to infer such goals and purposes at every level, from inferring the purpose of referring expressions to inferring the speaker's overall goal in constructing the text. One can understand every sentence in a text, yet come away puzzled at what it was the speaker was trying to say, or what the parts of the text had to do with each other. To understand the purpose of a speech act is to understand how it relates to a goal, how it is a step toward the achievement of that goal. The most appropriate kind of theory for this aspect of a text is a theory of plans, in which purposes, goals, acts, and intentions play a crucial role.

There are a large number of goals a speaker can have in constructing a text, including many that are irrelevant to comprehension: to derive royalties, for example, or to confuse an enemy by furnishing misinformation. A proper theory of text comprehension must distinguish goals like these from those that are central to communication and comprehension, probably by means of conditions like those Grice (1957) proposes as criteria for meaning.

C. What can go wrong. Then we can sketch the task of text comprehension as follows

1. From the sounds or markings, H must recover what sayings are recorded in the text, in what order.

2. From this H must recover what speech acts have been performed, in what order.

3. From each speech act H must recover what facts are being asserted, denied, promised, etc.

4. From this H must infer what modifications he is intended to make in his model of the world, and how to make them in the most consistent way; this is not a direct function of the facts, as discussed earlier.

5. For each speech act H must infer a purpose that is consistent with the purposes he inferred for earlier speech acts; or he must revise earlier hypotheses about purposes accordingly. Questions H must infer answers to are, "Why did the speaker perform this particular speech act, at this particular point in the text?" and "Why does he want me to have this particular fact just now?"

6. From speech acts and their purposes taken jointly, he must construct a hypothesis of the speaker's goal in the text, and of the plan that the speaker is following in advancing toward that goal. At each step the purpose of a given speech act must somehow be construed as consistent

> with, and actually advancing that plan, or the plan hypothesis must be modified so that it can.
>
> From hypotheses about the speaker's plans and goals in the text, the hearer will form expectations: hypotheses about what the speaker is likely to do next in advancing toward the goal of the text.

These matters do not proceed in separate compartments, of course, but feed each other. The plan one has constructed so far can influence decisions about what speech act is performed in a given utterance, for example, and the interpretation of pronouns can be influenced by hypotheses about the speaker's goals, just as a decision about what a referring expression is being used to refer to can influence the process of inferring a plan, and expectations about what the speaker will do next can influence the interpretation of what he actually does.

From this sketch we can derive a picture of where things can go wrong in comprehension, giving some insight perhaps into notions like "text structure," "relevance," and "coherence."

The hearer can have difficulty in tasks through 3, of course, but the matter seems straightforward, so I will not discuss it. Difficulties can arise in task 4 in at least two ways. First, the world described may be so factually or logically bizarre, or so inconsistent with the hearer's beliefs (a description of ping pong in a black hole, for example), that the hearer is unable to construct a consistent model with any degree of detail. The term "incoherent" might be applied to such cases, but I think this is not what linguists mean by "textual coherence," which I will discuss below.

A second kind of difficulty with task 4 arises when the facts are consistent, but the hearer lacks the knowledge necessary to figure out how to construct a consistent model that incorporates those facts. For example, if I describe in detail a walk through the South Side of Chicago, a person who has been there before will be able to construct a much more richly detailed model of my walk than a person who has not.

Difficulties can arise with task 5, insofar as the hearer is able to understand clearly what's being asserted, but unable to determine the speaker's purpose in asserting it. Here is the place to look for an adequate definition of relevance. Actually there are two senses of the word in ordinary usage. One can speak of relevance as a relation between facts. One fact is relevant to another when the truth of one depends in some way on the truth of the other. But I think more often, linguists who speak of "relevance" as a problem of text comprehension have in mind a problem that is best treated in terms of purposes behind speech acts. Given a hypothesis about the goal and plans of a speaker in a text, a given "sentence" (i.e. speech act) is taken to be irrelevant when the hearer is unable to see how it functions within the plan to advance toward the goal. Relevance under this interpretation, then, is a relation between an act and a goal, not a relation between sentences. if in recounting my recipe for Wienerschnitzel I describe my new driveway, it's not that the sentences are irrelevant; rather, I have done something irrelevant. The same passage may count as full of irrelevancies, relative to one goal, but uniformly relevant, relative to another goal.

Task 6 is probably the most complex and difficult, and the one we know least about. But I suspect that it is a likely source of progress in understanding such important but elusive notions as "coherence," "text structure," and "topic." In understanding a text, the hearer unconsciously searches out a primary goal behind the text, and tries to construe every part of the text as a purposeful step toward that goal, according to some plan. If the hearer is unable to reconstruct the goal or plan, or indeed decides there is none, the text will be judged "incoherent." Coherence is not a formal property of texts, nor of "logical structures" of texts, but a function of the hearer's ability to relate parts of the text to a plan for achieving some goal. If it should turn out that the coherence of texts correlates with the number of pronouns, it would be a mistake to conclude that lots of pronouns makes a text coherent. Rather, it would show that coherent texts tend to be ones where the speaker says a lot about one or two topics, rather than saying one thing about 32 topics. It is the coherence of what the speaker is doing in the text that gives rise to the abundance of pronouns; the formal property of having a lot of pronouns does not give rise to coherence.

At least some aspects of "text structure" can also be treated in these terms. An ideal unified paragraph, for example, is a unit of function, not of form; the speaker formulates a subgoal as a step toward the primary goal of the text and sets about to achieve that goal in a series of speech acts. Insofar as the hearer is able to discover this, the series of speech acts will be judged to be a unit; but a unit of function, not of form, defined not in terms of sentences or propositions, but communicative acts of some person, who uses those sentences to convey those propositions.

It is likely that an understanding of task 6 will lead to an understanding of "topic" as well. At present, there are nearly as many definitions of "topic" as there are linguists, and none of the definitions is clear enough to be usable. For some linguists the topic is a certain NP in a sentence; for others a topic is something a sentence has, though the NP may not be present in the sentences. For some every sentence has a topic; for others, only some sentences have topics. But I suspect that all of these attempts miss by a wide mark. First, it is not NP's that are topics, but the things in the world they refer to. Second, I suspect that such definitions can never be made sense of in that it is speakers, not sentences or even texts, that have topics. If so, then the proper theoretical treatment of "topic" would be framed

in terms of a theory of complex communicative acts, not formal linguistic properties.

IV. Conclusion

In this speculative paper I have proposed a way of looking at the comprehension of connected text that is counter to the linguist's usual way of looking at language. My main point is that certain notions are more likely to receive adequate treatment in a theory that incorporates a theory of speech acts, a theory of plans and goals, and a theory of inference, in place of a theory that looks for answers in terms of formal properties of texts. It remains, of course, to develop such theories to a level where my claims can be rigorously tested. The construction of such theories should be a prime goal of theoretical linguistics.

References

Austin, J. How to do things with words. Oxford: Oxford University Press, 1962.

Grice, H. P. Meaning. Philosophical Review, 1957, 66, 377-388.

Halliday, M. A. K., & Hasan, R. Cohesion in English. London. Longman, 1976.

Jespersen, O. The philosophy of grammar. New York: Norton, 1965.

McCawley, J. Remarks on the lexicography of performative verbs. In A. Rogers, R. Wall, and J. Murphy (Eds.), Proceedings of the Austin Conference on Performatives, Presuppositions, and Implicatures. Arlington, Va.: Center for Applied Linguistics, 1977.

Footnote

This research was supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116.

22

ASSENT AND COMPLIANCE IN CHILDREN'S LANGUAGE COMPREHENSION

David R. Olson
Ontario Institute for Studies in Education
Toronto, Canada

Abstract

It is conventional to treat the meaning of an utterance in a discourse in terms of two components, the propositional and the pragmatic or speech act component, the first indicating the meaning of the sentence, the second indicating its intended use by the speaker. Some arguments and evidence are presented to show that these two systems are interdependent. Roughly, it appears that social considerations, primarily status, determine which aspects of a proposition are lexicalized in the utterance. Thus, a child with high status relative to his interlocutor may use a command, "Give me a block", while if he has low status relative to his interlocutor he may use a request, "May I have a block?" If he is an equal, a peer, (and perhaps only then) he will use an explicit true proposition such as, "You have two more than me". Only in this third case is the propositional meaning explicit in the sentence per se, and only in this case is an affirmative or negative response dependent strictly upon truth conditions (on assent rather than compliance).

This concept of the social aspects of meaning is examined through an analysis of what is said vs what is meant in some child-child and teacher-child conversations.

. . . . . .

Theories of human cognition have gradually adjusted their accounts of perception and of knowledge to the fact that neither perception nor knowledge are simple copies of the environmental events that occasion them. Bruner's (1957) "New Look" in perception, which was devoted to showing the role of hypotheses, expectancies and set on the processes of perceptual recognition along with Bartlett's (1977) analyses of the role of "schema" in remembering helped to relativize the accounts of the relation between stimulus and perception or between reality and knowledge.

Now, however, we are asked to make our accounts of human cognition even more relative as not only to innate categories, and to expectancies based on prior experience, but rather to the social relations in which those knowledge structures are constructed. Theories advanced under the banner of the "sociology of knowledge" have claimed that the structures of knowledge and perception reflect the organizing properties of the social system in which the experiences occur and are assimilated.

This line of argument is usually attributed to Durkheim. According to his biographer Steven Lukes (1975) the cognitive processes reflect directly the social and political structures of a society. "Conceptual thought, was," Durkheim claimed, "social and nothing but an extension of society" (Lukes, 1973, pp. 23-24) and again, "Logical life has its first source in society" (Lukes, 1973, p. 441).

In an admirable collection of essays and monographs, Mary Douglas (1975) takes up and extends the Durkheimian view. With Durkheim, she claims that: "...ideas rest on classification. Ultimately any form of knowledge depend on principles of classification. But these principles arise out of social experience, sustain a given social pattern and themselves are sustained by it. If this guideline and base is grossly disturbed, knowledge itself is at risk" (p. 245). Specifically, she argues that the discriminating principles as to what is clean and what is polluted and what generally is "against nature" is derived from social structure. Nature is classified in such a way as to uphold the social order--thus in a social order in which men are status-dominant over women and children it seems only natural that women and children be assigned low-status duties such as dish-washing.

Taken in their boldest form, these theories argue that knowledge is socially constructed; there are close ties between the social order and the conceptual order. But how does this social order affect, come to be affected, or otherwise interact with the conceptual order?

Most of the theories which attempt to relate social structures to cognitive ones have postulated symbols as the mediating link. Symbols are culturally designed and they are acquired by children for use in communication and for the interpretation of experience. But where in a symbol system, such as language, shall we look for evidence of their relation to social structures? The best known of these theories of this relation such as those

---

Paper prepared for Theoretical Issues in Natural Language Processing (TINLAP-2). Urbana, University of Illinois, July 25-27, 1978. I am indebted to Ed Sullivan and Frank Smith for our lively discussions of these issues, and to Canada Council for support.

of Vygotsky (1962), Bruner (1966) or Whorf (1956) have tried to isolate the social constraints on the semantic, propositional side of language rather than the social, interpersonal or pragmatic side of language. In so doing they followed the then-predominant focus of linguistic and semiotic theory.

Theories of symbols have tended to emphasize the semantic, denotative or referential side of meaning at the expense of the social, pragmatic or interpersonal side of meaning. In his classical treatment of symbol systems, Nelson Goodman (1968) focused exclusively on the objective or informative or semantic aspects of symbols, that is, the relation that exists between the symbol and the event it represents, denotes, expresses or exemplifies.

Chomsky (1972, p. 24) too, focuses on the logical or semantic aspects of meaning--that aspect of meaning which is invariant across the various functions to which a sentence may be put, in his theory of language. Studies of language and cognition have similarly focused on labels and on grammatical relations in their attempts to find the relations between language and thinking. Hence Durkheim looked at classification systems, Vygotsky looked at the superordinate categories, and "scientific concepts" Bruner and Greenfield looked in the structure of adjectives, Cole and Scribner looked in the structure of logical inferences and so on for evidence of social-cultural differences. And indeed some indications of social differences have shown up in those studies although they tend to reflect simply degrees of familiarity with a certain approach to problems primarily, "learning to confine interpretation to the text" (Olson and Nickerson, 1978).

There is however a second dimension of language which has been brought to the fore by recent analysis of the functions of language under the rubric of speech act theory. Sentences are now considered and analyzed both in terms of the propositional content of the sentence, traditionally the semantics or meanings of the language, and the pragmatic function, what one is attempting to do by means of that sentence--promising, convincing, commanding, requesting, stating and the like. Searle (1969) following Austin (1962) differentiates propositional acts of referring and predicating from illocationary acts or stating, questioning, commanding, promising and so on. The same reference and predication can occur in the performance of completely different speech acts as in Searle's example:

1. Sam smokes habitually.
2. Does Sam smoke habitually?
3. Sam, smoke habitually!
4. Would that Sam smoked habitually. (p. 22).

Halliday (1970, 1973) too, in his analysis of the functions of language, differentiated three primary functions of language--the ideational function, the interpersonal function and the textual function--which jointly specify the structure of any particular utterance or series of utterances. Clark and Clark (1977) suggest that speakers and listeners of the language use two general principles in comprehending utterances, the "reality principle" and the "cooperative principle". Again note that the reality principle has to do with the relation between symbol and referent while the latter has to do with the relation between speakers.

Let us try to simply this relationship by means of a diagram. Figure 1 shows the set of relations sustained by any symbol.

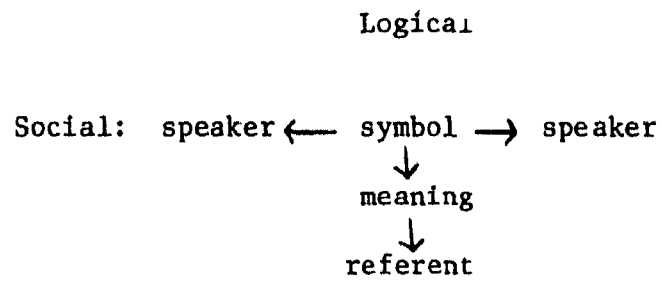


Figure 1.

Note first that a symbol simultaneously serves two sorts of relations represented in the diagram by two dimensions of language--one dimension, represented by the horizontal axis in Figure 1, is that specifying the social relations between speaker/writer and listener/audience; the symbol expresses a social relation between speaker and listener. And the second dimension, represented by the vertical axis, is that specifying a relation between symbol and meaning or between signifier and signified. The first we may call the social or interpersonal meaning of a symbol and the second the logical meaning of that symbol, linguistic or otherwise.

It is this horizontal dimension, the social one, which until recently has not been given its due. Most studies of language for example have been devoted primarily to the syntactic/semantic meaning of words and sentences and the veridical relationships between sentences and the situations they represent. This is the aspect of language which most weakly reflects the social relations. And yet, as I mentioned above, this is the dimension of meaning that most hypotheses of the Durkheimian sort have focused upon.

It is the horizontal relation, the relations between persons that promises to illuminate the relations between social and cognitive structures. Let us consider what some of these social constraints on the meaning of a symbol may be. First, even the aspect of meaning represented by the vertical axis has an important social component. To be represented in the language, symbols must be socially shared. What gets into the language depends upon what others will agree to. Indeed, the semantic aspect of language may be important partly because it constitutes a relatively undisputed ground for the reaching of social consensus, what is coming to be called "intersubjectivity "

More recent work on the pragmatics of language, the horizontal dimension, and the social relations assumed and maintained in the expression of any sentence, goes far to put the social dimension of language back onto center stage. In the simplest case the declaratives "The door is closed" is logically related to the command "Open the door" and the request "Could you open the door." As Searle (1975) would say, the three utterances represent the same propositional content yet they

perform different speech acts: the first, an assertion, the second, a command and the third, a question serving as a request. While this speech act analysis contributes importantly to the view that the meaning of a sentence is more than the proposition it expresses, that is, it explicates the function or pragmatics of language, it seems not to go far enough.

It may be argued that in the above three utterances both the propositional content and the pragmatic function of the sentences are similar--they all represent the proposition: not (door(open)) and they all serve the pragmatic function of A attempting to get B to open the door. These sentences which have different illocutionary forces as part of their "sentence meanings" are being used indirectly to perform the same illocutionary act (Searle, 1975, p. 71). They vary in their indirectness, and politeness.

However, the important point for us to notice is that the social relations between the speakers assumed by the three sentences is entirely different. The command assumes that the speaker has superior status to the listener, the request assumes a differential inferiority of the speaker and the declarative assumes, perhaps, the equality of the participants. The point I wish to emphasize is that an utterance is simultaneously doing two things--it is specifying the logical relation between symbol and referent, the vertical dimension of Figure 1, and it is specifying a social relation between the interlocutors. Together, they contribute to the meaning of the symbol, utterance or expression. However, Searle's analysis is above all a theory of language. It is less a theory of the social structures that those linguistic structures construct and/or sustain. My concern here is to relate these two sets of structures.

One of the most notable attempts to construct a bridge between sociological theory and psychological theory of language and the cognitive processes is that of Bernstein and his collaborators (1971). Bernstein isolated two patterns of speech which, he claimed, mapped on to two social classes. The language of working class children he characterized as a restricted code which limited grammatical and lexical options, while that of the middle class children he characterized as an elaborated code with an expanded set of lexical and syntactic options. More fundamentally however, he argued that the root cause of these linguistic differences could be found in the patterns of social relations which held within each of the family types of the differing social classes. Working class familial structures he characterized as "positional"--a fixed hierarchical structure in which authority, responsibility, accountibility and priviledges were assigned on the basis of one's position in the family. Middle class families he described as person-oriented. Duties and priviledges were assigned to various roles, but a person was not permanently assigned to a fixed role but rather the role was contracted or negotiated with the other members of the social group. Duties, priviledges and responsibilities were assigned by a discussion and contract rather than being permanently assigned to particular individuals. Social theorists would recognize these patterns as essentially "monarchist" versus "social contract" political structures. In the first case the parent with the highest status has the right to decree what family members are to do, he is awarded respect and he is assigned high status duties. Low status individuals are to be obedient and to accept responsibility for low status duties. In the latter case, such assignments are negotiated.

these social differences translate directly into the linguistic patterns mentioned above. The former, relying primarily upon position and status require the close observance of status relations and a minimum of negotiation. For the middle class, the social contract requires a high degree of linguistic competence in negotiating roles, rights, agreements and priviledges. Hence the latter can be expected to generate an elaborated code. Evidence for this theory has generally outweighed evidence against it but several writers have attributed the observed differences not to linguistic incompetence but to the social environments in which the data was collected. Labov (1972) for example found that speakers of Black English Vernacular were essentially as adept with the language when they conversed with their peers as were white children. Black children did especially poorly only when they were interviewed by a high-status white teacher/experimenter.

Bernstein's theory is primarily important for its assertion of a direct link between social structures and linguistic structures. However the differences should be expected less in the lexical and syntactic structures then in the social meanings--in the pragmatics of language. Working class children do know the pragmatic options as Mitchell-Kernan and Kernan's (1977) interviews make abundantly clear. However, working class children (more generally children from families with a positional, heirarchical structure) may be expected to assign themselves a <u>low</u> status relative to various authorities and assign interpretations on the basis of those status differentials more so than will middle class children. Thus, it may be predicted that if sentences which are ambiguous representations of speech acts were presented to working class and middle class children, the former would tend to <u>comply</u>, that is to treat the statement as an indirect request--while middle class children would tend to <u>assent</u>, that is to treat the statement as ability or information questions. To illustrate if asked, "Can you tell me what this is?", working class children may <u>comply</u> by responding "A pencil" while middle class children may tend to <u>assent</u> by responding "Yes, I can!"

Generally speaking then, the linguistic options are presumably much the same for all speakers. The social relations assumed and maintained differ--if you have low status you can expect to be given commands--if high, to give them. Many of our institutional practices can be seen as keeping people in their assigned status positions and language may be seen as one such means.

Nice illustrations of these social-linguistic games have been presented by Gumperz (Note 2), Ervin-Tripp (1977) and Mitchell-Kernan and Kernan (1977). Gumperz reports a conversation between a husband and wife in which the husband asks "Where's the paper?" and the wife, with some annoyance, replies

"I'll get it." The first statement may be interpreted as either a request for locative information or an indirect request that she get him the paper. Her recognition of that second possibility, that he, even if only in the back of his mind, wanted her to get him the paper, was the source of annoyance.

Mitchell-Kernan and Kernan's (1977) example is even more apposite. In the course of a group discussion between an experimenter-teacher and a group of children one child gave an indirect command to the adult who had her foot on a chair: "I want that chair!" with which the adult complied. Some of the other children gasped and said "O-o-o-o Claudia, you gon' let her talk to you like that?" (p. 205). The social message is obvious. You have to have status to give commands and the child did not have this status. The children were well aware of that aspect of meaning. Mitchell-Kernan cites further examples of precisely this social game. Some of the children would try commands on other children simply to see if the listener would comply. If they complied, the child issuing the command would have gained proof of his greater social status. When they tabulated the frequency of various kinds of "directives", Mitchell-Kernan and Kernan found that addressees who were lower in rank than the speaker received over five times as many directives as those higher in rank (p. 203). They conclude: "Directives and reactions to them were constantly used to define, reaffirm, challenge, manipulate, and redefine status and rank. At times the directives involved actually served the ordinary function of directives--that of requesting goods or services--while at the same time, because of their frequency of occurrence or the particular form they took, served to test the addressee's view of the statuses involved" (p. 201). Ervin-Tripp (1977) too found a particular social distribution of various forms of directives in her studies of how children learn to honor various factors such as age, dominance, task and familiarity in making requests (p. 186).

As these authors point out, directives are used only as instrumental means of carrying out pragmatic speech acts suitable to particular interpersonal and social conditions but also to accomplish certain interpersonal functions--primarily establishing and maintaining a social order, a status quo. In using these constructions, the child is simultaneously learning two interrelated pictures of reality--the nature of "objective reality", that is the propositional or knowledge system, and his place in the social order--that is, who has a right to make requests, to issue commands, and to make true descriptions and so on. As argued above, this is because every symbolic expression has a value on both of these dimensions. And I have suggested that these two dimensions are interdependent; descriptions are more apt to be assessed exclusively in terms of their truth in some social relations than in others.

Less direct evidence of the uses of status differentials in the speech patterns comes from studies of language in the classroom. Feldman and Wertsch (Note 1) reported that teachers in the classroom rarely used what they called "stance indicators"--verbs such as may, or should, or wish or hope, think or believe or qualifiers such as maybe in their classrooms but they frequently used them in their speech in the lunchroom. That is, teachers in the classroom act as spokesmen for the official public view and keep their feelings, aspirations and hesitancies from display. In their classic studies of the language in the classroom Bellack, Kliebard, Hyman and Smith (1966) found that a predominant form of language in the classroom was of the question-answer routines known as the "recitation method." Interestingly, it was the teachers who asked the questions and it was the children who provided the answers. It appears that the right to ask questions was a high status perrogative. Furthermore, the questions were not simply requests for information. The teacher already knew the answer--the point of the question was to see if the child knew the answer. The question serves primarily as a means of holding students accountable for the information acquired from reading the text. While the utterances specify true facts, that is, relations between symbols and referents, the form that the proposition takes again depends upon the social relations between the interlocutors.

Sinclair and Coulthard (1975) also found that many of the interrogative and declarative sentences used by teachers in fact served as imperatives. For example, the statement "Somebody's talking," "I see chewing gum" were not to be taken as true descriptions but as indirect commands. They called, as we prefer to state it, for compliance rather than assent.

Goody (Note 3) in an interesting study of the forms of questioning among the Gonja of Ghana, found that questions in that society were not merely means of securing information but were primarily reflections of the status of the interlocutors. Hence, children rarely asked questions, not because they had no need for information, but rather because asking a question of an adult would be to upset the social order. Goody comments: "The securing of information becomes secondary to considerations of status relations" (p. 42, 1975).

The surface form of an utterance depends upon both the propositional structure and on the pragmatic function. All sentences appear to do both simultaneously--as we have suggested the symbol simultaneously stands in a specifiable relation to a meaning--it represents a proposition--and in a specifiable relation to the speaker and his listener Both of these dimensions may be invariant across some set of transformations. Different sentences may assert the same proposition, as for example active and passive sentences or declaratives and questions. Presumably, as well, sentences with different propositional content may be used to construct or maintain the same social relation between interlocutors. There is presumably more than one way to be obsequious and more than one way to be insubordinate.

Most theories of the pragmatics of language would agree, roughly, to this point. However, even if they may be differentiated, it is important to notice that those two dimensions, the logical and the social are not independent. This point may be expressed by the question: What are the conditions under which an utterance can be judged simply or exclusively for its truth value or its propositional meaning? And what are the conditions under which an utterance will be judged as an order, or a request for action? Let us distinguish these two

criteria for interpretation by means of labels: the first, the judgment for truth or falsity we may say calls for assent; the second, the response to a statement as a call for action, we may say calls for compliance. My conjecture is that certain social and institutional arrangements lead any particular utterance to be regarded as a call for assent while others, as a call for compliance. Symmetrically, certain social conditions are more likely to generate assertions and information questions while others generate indirect requests and imperatives.

We have examined these conjectures in two ways. First we have examined some transcripts of children's discourse and secondly we have designed three small experiments to follow up and clarify some of the relations we seem to have isolated in the analysis of transcripts.

Nancy Nickerson has recently collected and begun to analyze a series of dialogues between pairs of children as they played with toys. We are intersted both in the quality of oral expression (in an attempt to see in what ways oral language competence is related to learning to read) and in the use of statements, questions and commands in cooperating with and controlling the behaviour of peers. Although that project is at an early stage, I shall present one analysis of a transcript in terms of the model described above. This dialogue occurred between two Nursery School children named Jamie and Lisa who had some difficulty arriving at an equitable distribution of a limited resource, namely some dominoes. Let us see how they use language to negotiate this social problem.

L: Let's make a domino house out of these
J: Okay

First, by grabs.

J: Lookit how many I got....You took a couple of mine!
L: Now you took a couple....

Then by commands.

L: Now you got to give me three back!
....
L: Now give me just one more and then we got the same

And then by requestful declaratives.

J: Now, you got more than me-e

And denials.

L: No-o we got the same

By fact collecting and inferencing.

L: (Begins to count her dominoes). One, two, three, four....twenty-eight, twenty-nine (Then counts Jamie's dominoes). One, two, three, four....eighteen, nineteen...(short pause) twenty-nine.
J: I got nineteen and you got twenty-nine.... You got more than me.
L: No-o (shouting) I COUNTED....You have the same as me....We got the same.
J: NO-O-O

And when negotiations break down and again by grasping.

(There is a shuffle of dominoes across the floor.)

And finally by appeal to authority.

L: You got much more than me now
J: No we got the same

(Paul, a volunteer teacher, enters the room.)

L: Does he have much more than me?
P: Not too many more!

Note first that almost all of these quite different utterances are attempts to alter or preserve the social arrangement of two children playing together and sharing the limited supply of dominoes. "Now you got to give me three back," a command, has the same pragmatic function as "Now, you got more than me," an assertion standing as an indirect request, spoken by the same person. And both speakers appear to be aware of the social meaning, namely; that the listener should hand over one or more of the dominoes, even if in one case it is the explicit "give me" and in the other, the implicit "you have more." Why then do they use one device rather than another?

We may see how the logical and the social meanings interact if we score each sentence for both its logical or "truth" meaning, the assent criterion, and for its social meaning, the complicance criterion. For the logical meaning, true may be marked with a "+" and false with a "-". For the social meaning, the categories are less obvious. We let "+" represent the preservation of any current social arrangement, i.e., those not requiring compliance and "-" represent the realignment of any social relationship--statements which require compliance and call for revolutionary activity, so to speak. Now let us examine some fragments of this dialogue in this framework.

			Criterion	
			Assent	Compliance
	Sentence	Gloss	Truth Value	Status-Preserving
J:	"You got more than me."	(Give me some)	+	
L:	"No, we got the same."	(I don't have to)	-	
P:	"Not too many more "	(Yes, it's true she has more but she does not have to give you any).	+	

Note that Jamie tells the truth with the hope of realigning the distribution of dominoes. Lisa, technically speaking, tells a lie. (Recall that sne was the one who counted them). But her denial was not merely one of falsehood. She knows that if she

agrees to the truth of Jamie's statement, she will have to turn over some of the blocks. That she doesn't want to do so, she denies the statement. My guess is that that is what all lies are--tampering with truth value for social or personal ends. Truth, like falsehood is motivated.

More than that. Lisa is not denying the truth of Jamie's statement simply in the service of social ends. Rather, I would guess, she does not know any means of simultaneously meeting both the social and logical criteria. Paul, the teacher does. Note his reply when Jamie appeals to him. The presupposition of his sentence is that Lisa has more. Rather than deny it, he presupposes it and uses his sentence to hold that no redistribution is required, presumably on the premise that possession is nine-tenths of the law.

What I would like to suggest from this example is that truth conditions are not separated from social utility. Claims of truth will be advanced primarily if the gaining of assent implies compliance with some social goal. Symmetrically, denials of truth will be offered if the social consequence of assent are perceived to be undesirable.

It is at this point that social relations enter into the language. Micro social orders, small scale transactions, like the one mentioned above, involve the solution of small-scale interpersonal problems which must be solved either for individual machiavellian goals or for shared social goals. The main problem is how to secure compliance, agreement or at least to prevent the loss of the status quo (An interesting expression). That may be done by several means, direct action, commanding, pleading, or hard negotiation on a common ground. Facts are one such ground, which as we have seen, are overlooked if they are embarassing; authority is another such ground, which as we have seen, tries to get involved or take sides.

Disputes in the larger social order appear to be solved on somewhat similar means. As Foucault (1971) pointed out, different social orders make use of different criteria for truth and hence different grounds for the legitimation of the social order. Authority, the father in a patriarchical order or the priest in an ecclesiastical order, has the power to decide in the case of disputes, as judges do in our own courts. Hopefully, they have adequate recourse to the truth, but poor judgments carry just as much weight as good ones. The decisions likely to gain the greatest compliance have both.

It is interesting to recall in this context the wisdom of Solomon. One may wonder if Solomon's judgments were considered so good because he was so wise or because he happened as well, to be king. More likely the stories of his wisdom and justice, helped to legitimize the authority that was socially assigned to him.

In our own society, great weight is assigned to "truth", "facts", "sense data" as objective grounds for making scientific, social and political decisions. As long as people believe that truth is objective, it serves as an important means of "legitimizing" a social order (Habermas, 1973). Furthermore the establishment of institutions like universities devoted to discovering the "truth" independent of its social utility, helps to sustain the view that there are such facts and that those facts can be used to sustain the social order. (Hence, we may count on some continued support even if we were (God forbid) unsuccessful in finding any such truths.)

A single argument between two four-year-olds may be insufficient empirical grounds to sustain a general social theory, hence, we have attempted to further examine some of these expressions and their interpretations by experimental means. Angela Hildyard recently took some of the statements from our transcripts and built them into a series of ten stories. Here is one of these:

> One Saturday morning Susie and Kevin Jones went to the movies. Their mum gave them some money to buy some popcorn. Susie bought a large box and they shared it out. When Kevin looked at Susie's share he didn't feel too happy "You've got more than me" he said.

The stories describing social predicaments of this sort were each followed by a recall test, e.g., What did they buy to eat? What did Kevin say?

The most interesting results came from Kingergarten and Grade 2 children's reply to the second question.

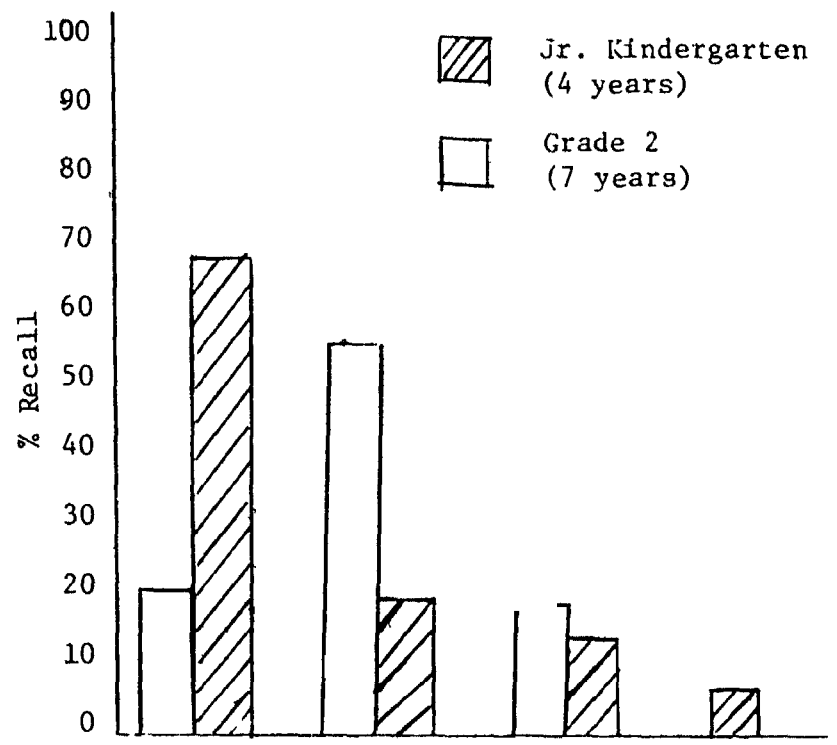


Figure 2. The recall of declarative statements as given (verbatim) and as direct and indirect requests by Junior Kindergarten and Grade 2 children.

Junior kindergarten children (Aged 4 and 5 years), when asked what Kevin had said would frequently answer with a request or a command, "Can I have some?" or "Give me some popcorn." By the second

grade (Aged 7 and 8), children tended to recall the statement verbatim: "You have more than me". When further queried as to why he said that, they replied, "Because he wanted more". These results are shown in Figure 2. The implication of these findings is that almost all of the children interpreted the statement, "You have more than me" not simply as a true statement but rather as an indirect request and that interpretation biased the recall of the younger subjects. Older subjects remembered both what was intended by the sentence and the means the speaker used, here a declarative, true statement, to achieve it. That is, the sentence was not interpreted and evaluated strictly or even primarily on truth criteria but on social ones. And the younger the children, the stronger the tendency to treat the sentence as a call for compliance and hence to report it as such.

These findings are similar to those we have obtained earlier in our studies of recall and inference from stories with children of different ages in which we showed that recall tends to be of "what was meant" rather than "what was said" and that with age (and schooling) children come to be able to differentiate the two (Hildyard and Olson, in press).

In a further study, Hildyard read the stories, excluding the last line, to Kindergarten (age 5) and Grade 2 (Age 7) children, and adults and asked them to imagine what the victim said--in the above story, for example, "What do you think Kevin said to Susie?" There were two important additional factors. First, the child may be either demanding a right--it may be that he is asking for his own sock back--or a favor--he may be asking for something that actually belongs to the other child. Further they could be asking these rights or favors of low status individuals, young children, or high status individuals, parents and teachers. There were 16 subjects at each grade level and there were three stories of each type.

Again, I shall mention only the most interesting results. First, favors were much more likely to be signalled by a conventionalized request than were rights. Over all age levels and item types, the favor items were marked by requests 77% of the time while rights were so marked only 45% of the time. How rights were signalled varied with the age of the subjects. The youngest subjects use direct commands to their peers and conventionalized requests to their parents and teachers. Thus they may say, "Give me my sock" to a peer and "May I have my gold star" (which had been promised) to the teacher. Adult subjects, while they use requests for favors, rarely use commands in attempts at obtaining their rights and tend rather to use declaratives "You have my sock", and questions, "Do you know where my sock is?" Adults use this device 33% of the time while children of both ages use it about 15% of the time. Adults, in obtaining their rights also use "legitimized requests", requests accompanied by reasons more so than do the children. These results are presented in Table 1.

Table 1

The use of various forms of directives as a function of the type of request, rights vs. favors, and the status relations involved

	Kindergarten N=16	Grade 2 N=16	Adult N=16	Total
Use of Conventionalized Requests				
Favors	86	86	58	77%
Rights	59	49	27	45%
Use of Commands				
Favors	5	2	10	6%
Rights	19	6	14	13%
Declaratives and Questions				
Favors	7	5	14	9%
Rights	13	18	33	21%
Legitimized Requests				
Favors	1	6	11	6%
Rights	5	24	19	16%
Threats, Negotiations, Appeals				
Favors	0	0	7	3%
Rights	0	3	6	3%

Note that in all of these cases, subjects aspired to the same goal--either through their direct meaning or in their indirect meanings they conveyed the same illocutionary force. Yet the utterance used to express that intention took a different form depending upon the social relations between the participants. Primary among those factors is the status relations between them--commands may be issued to lower status individuals, requests must be issued to superiors even if you are only asking for your rights. Secondly, the presumed rights, the status quo, determines the form in which that illocutionary force will be expressed. Favors are largely expressed through requests, although adults also frequently add reasons, while rights may be expressed through commands, occasionally through threats, or through the provision of reasons (Declaratives and Questions). Ervin-Tripp (1977) has recently cited similar results from an unpublished study by Sharon James.

In a third study, Beverly Wolfus gave Kindergarten and Grade 2 children a series of direct commands such as (Tell me what this is! Put the penny in the glass!) and ambiguous ability requests (Can you turn over the cup? Could you tell me the name of this?), while pointing to a cup or other objects and observed how children interpreted and responded to them. She was particularly interested in whether the children opted for the direct expressed meaning or the indirect pragmatic meaning of the ambiguous expressions. Thus "Tell me if you can put the penny in the glass" could be answered by assent, "Yes I can"--the direct meaning--or by compliance, by actually putting the penny in the glass--the indirect, or pragmatic meaning. When issued direct commands, both age groups complied extremely consistently. Told "Open the book", everyone opened the book. To "Tell me what this is" every one said, "A pen".

The differences in the responses to ambiguous questions and statements for the two age groups were striking. These differences occurred in response to the questions and statements which were ambiguous between a propositional and a pragmatic interpretation, that is ambiguous in their call for assent as opposed to compliance. To the statement "Can you turn over the cup?" older children would say "Yes" that is, assent, younger children would silently turn it over, that is, comply. To "Tell me if you can put the penny in the glass", older children would assent by saying "Yes" or "I can", younger children would comply by putting the penny in the cup. To "Do you know what this is?" while pointing at a penny or a cup, older children would assent by replying "Yes", young children would comply by replying "Penny" or "Cup". To the statement, "The book is closed", young children would, more often than older children, silently open it, older children would not but rather await further information. These results are shown in Table 2.

Table 2

Assent versus compliance in Kindergarten and Grade 2 children's interpretation of sentences

	Assent	Compliance	Both
Put the top on the pen.			
Kindergarten		100%	
Grade 2		100%	
Tell me what this is.			
Kindergarten		100%	
Grade 2		100%	
Tell me if you can X.			
Kindergarten	8%	83%	8%
Grade 2	69%	29%	2%
Tell me if you know how to X.			
Kindergarten	13%	64%	8%
Grade 2	66%	27%	8%
Do you know what this is?			
Kindergarten	5%	88%	8%
Grade 2	30%	67%	3%
Can you turn over the cup?			
Kindergarten	8%	64%	24%
Grade 2	64%	19%	18%
Do you know how to X?			
Kindergarten	25%	11%	64%
Grade 2	91%	3%	5%
The book is closed.			
Kindergarten	73%	20%	7%
Grade 2	92%	8%	

Overall, these data show that the younger children took every utterance in terms of its expressed or implied illocutionary force and complied with it, while older children tended to differentiate the direct meaning from its indirect illocutionary force and respond to the direct meaning. Ervin-Tripp (1977), observed a similar affect. When told to "Say why don't you stand up" the child said "Stand up!" and stood up--that is, he complied. Again this indicates that the meaning of the statement is, at least at the beginning, not simply its truth value, although that meaning may be calculated as part of its more pragmatic meaning. Rather, the sentences are scanned, as it were, for their implications for action, and that is what the youngest children tend to opt for. Further we can see some indication of the transition from compliance to assent in the Kindergarten children in the item "Do you know how to X?" Here these young subjects would first assent and then comply 64% of the time.

Again, with age or schooling, children begin to differentiate the propositional from the pragmatic meaning and to be able to respond to either of them. However, it appears that this propositional meaning is not primary but rather specialized out of a primary undifferentiated social pragmatic meaning.

Note that one of the factors that appears to give those simple declarative sentences a powerful illocutionary force for the youngest children is that they are spoken by an adult in a school context. The child assumes that the adult is not just stating or asserting some thing "The book is open" but indirectly requesting that something be done about it. Given a high status individual in a hierarchically structured institutional context, the child assumes that any utterance requires compliance not assent.

Let us return, in conclusion, to some of the general issues raised at the outset. I have argued that all utterances serve both objective (truth) functions and social (compliance) functions and that these are not independent. Many examples from the studies showed that the form which a directive takes depends upon the status relations between the interlocutors. This is clearly the case for imperatives and requests--requests are more likely for privileges and imperatives for rights. Requests are more likely to higher status indiv duals and imperatives to lower status individuals. Imperatives occur more frequently when children address lower status individuals and rationalized questions, declaratives and requests occur more frequently in the language of adults. Further the evidence showed that almost any form of utterance from a high status individual including a declarative sentence such as "The book is open" tended to be interpreted as an indirect imperative when spoken to a lower status individual. That is, the statement called for compliance rather than assent.

Our question at this point is whether or not status differences, differences in the social relations between speakers affect the interpretation of statements with the putative status of true descriptions. Generally, are claims of truth independent of claims of status?

As I have suggested, putative true statements call for assent (or falsification) while putative imperatives call for compliance (or defiance). But, as we have seen, if a true description is given by a high status individual, a lower status individual may respond with compliance rather than assent. Can the two criteria be specialized? Can state-

ments ever be constructed such that they call for assent and not compliance? If not why are there such things as assertions? The fact that some statements can be assented to indicate that the truth functions can be isolated at least somewhat from their more general social functions. Even in that case however, the true statements would be generated (or denied) when they have social utility, much as Lisa denied Jamie's statement that she had more.

The alternative, however, that the meaning of a sentence is purely subjective, that is whatever you can get a listener to comply with, is even more precarious. As Harre (1974) pointed out this is a territory suited only for the bravest machiavellians --to assume the status to make demands and declarations and to continue to do so until someone refuses to comply.

A more promising approach would be to argue that the meanings of utterances can gain their agreement from speakers on the two bases we have discussed and that these two bases are in continuous interaction. One may get agreement, either assent or compliance, simply because of the status relations involved. One is in a position to command or to declare that such and such is the case and the other, agreeing to that higher status, receives those commands and declarations and assents or complieswith them. The bulk of social negotiations proceed on this basis. But if there is a collapse of the social order, or a condition of general equality, no one person is in a position to demand either assent or compliance. Then the ground for the adjudication of disputesor more simply for the negotiation of meaning falls on to the objective, descriptive, or logical dimension of meaning. It is, presumably, easier to gain assent than to gain compliance; hence the importance of truth in any social order. And in the microsocial order, negotiations are carried out by any means available, but as Friere (1972) has suggested, genuine conversation is possible only between equals.

Reference Notes

Feldman, C.F. and Wertsch, J V. Context dependent properties of teachers' speech. (preprint)

Gumperz, J.J. Language, social knowledge and interpersonal relations. Language Behaviour Research Laboratory University of California, Berkeley. (mimeo.)

Goody, E. Towards a theory of questions. Draft of the Malinowski Lecture, London School of Economics, 1975.

References

Austin, J.L. How to do things with words New York: Oxford University Press, 1962.

Bartlett, F.C. Remembering. New York: Cambridge University Press, 1977.

Bellack, A.A., Kliebard, H.M., Hyman, R.T. and Smith, F.L. The language of the classroom. New York: Teacher's College Press, 1966.

Bernstein, B. Class, codes and control. London: RKP, 1971.

Bruner, J.S. On perceptual readiness. Psychological Review, 1957, 64, 123-52.

Bruner, J.S. Introduction. In J.S. Bruner, R. Olver and P.M. Greenfield (Eds.), Studies in cognitive growth. New York: John Wiley and Sons, 1966.

Chomsky, N. Problems of knowledge and freedom. London: Fontana, 1972.

Clark, H.H. and Clark, E.V. Psychology and language. New York: Harcourt, Brace and Jovanovich, 1977.

Douglas, M. Implicit meanings. London: RKP, 1975.

Ervin-Trip, S. Wait for me, Roller-Skate! In S. Ervin-Tripp and C. Mitchell-Kernan (Eds.), Child discourse. New York: Academic Press, 1977.

Foucault, M. L'ordre du discours. Paris: Gallimard, 1971.

Friere, P. Pedogogy of the oppressed. New York: Herder and Herder, 1972.

Goodman, N. Languages of art. Indianapolis: Bobbs-Merrill, 1968.

Habermas, J. Legitimation crisis. Boston: Beacon Press, 1973.

Halliday, M.A.K. Language structure and language function. In J. Lyons (Ed.) New horizons in linguistics. Harmondsworth: Penguin Books, 1970.

Halliday, M.A.K. Explorations in the functions of language. London: Edward Arnold, 1973.

Harre, R. Ethology and early socialization. In M.P.M. Richards (Ed.), The integration of the child into the social world. Cambridge: Cambridge University Press, 1974.

Hildyard, A. and Olson, D.R. Memory and inference in the comprehension of oral and written discourse. Discourse Comprehension. (in press)

Labov, W. Language in the inner city. Oxford: Blackwell, 1972

Lukes, S. Emile Durkheim. Markham, Ontario· Penguin Press, 1973.

Mitchell-Kernan, C. and Kernan, K.T. Pragmatics of directive choice among children. In S. Ervin-Tripp and C. Mitchell-Kernan (Eds.), Child discourse. New York: Academic Press, 1977

Olson, D.R. and Nickerson, N. Language development through the school years. In K.E. Nelson (Ed.), Language development. New York: Gardner Press. (in press)

Searle, J. Speech acts. Cambridge: Cambridge University Press, 1969.

Searle, J. Indirect speech acts. In P. Cole and J. Morgan (Eds.), Syntax and semantics. Volume 3: Speech Acts. New York: Academic Press, 1975.

Sinclair, J. and Coulthard, R.M. Towards an analysis of discourse: The English used by teachers and pupils. London: Oxford University Press, 1975.

Vygotsky, L.M. Thought and language. Cambridge, Mass.: M.I.T. Press. 1962.

Whorf, B.L. Language, thought and reality. New York: Wiley, 1956.

Speech Acts as a Basis for Understanding Dialogue Coherence

by

C. Raymond Perrault and James F. Allen
Dept. of Computer Science
University of Toronto
Toronto Canada

and

Philip R. Cohen
Bolt Beranek and Newman
Cambridge Mass.

1. Introduction

Webster's dictionary defines "coherence" as "the quality of being logically integrated, consistent, and intelligible". If one were asked whether a sequence of physical acts being performed by an agent was coherent, a crucial factor in the decision would be whether the acts were perceived as contributing to the achievement of an overall goal. In that case they can frequently be described briefly, by naming the goal or the procedure executed to achieve it. Once the intended goal has been conjectured, the sequence can be described as a more or less correct, more or less optimal attempt at the achievement of the goal.

One of the mainstreams of AI research has been the study of problem solving behaviour in humans and its simulation by machines. This can be considered as the task of transforming an initial state of the world into a goal state by finding an appropriate sequence of applications of operators from a given set. Each operator has two modes of execution: in the first it changes the "real world", and in the second it changes a model of the real world. Sequences of these operators we call plans. They can be constructed, simulated, executed, optimized and debugged. Operators are usually thought of as achieving certain effects and of being applicable only when certain preconditions hold.

The effects of one agent executing his plans may be observable by other agents, who, assuming that these plans were produced by the first agent's plan construction algorithms, may try to infer the plan being executed from the observed changes to the world. The fact that this inferencing may be intended by the first agent underlies human communication.

Each agent maintains a model of the world, including a model of the models of other agents. Linguistic utterances are the result of the execution of operators whose effects are mainly on the models that the speaker and hearer maintain of each other. These effects are intended by the speaker to be produced partly by the hearer's recognition of the speaker's plan.

This view of the communication process is very close in spirit to the Austin-Grice-Strawson-Searle approach to illocutionary acts, and indeed was strongly influenced by it. We are working on a theory of speech acts based on the notions of plans, world models, plan construction and plan recognition. It is intended that this theory should answer questions such as:

(1) Under what circumstances can an observer believe that a speaker has sincerely and non-defectively performed a particular illocutionary act in producing utterance for a hearer? The observer could also be the hearer or speaker.

(2) What changes does the successful execution of a speech act make to the speaker's model of the hearer, and to the hearer's model of the speaker?

(3) How is the meaning (sense/reference) of an utterance x related to the acts that can be performed in uttering x?

A theory of speech acts based on plans must specify at least the following:

(1) A Planning System: a language for describing states of the world, a language for describing operators and algorithms for plan construction and plan inference. Semantics for the languages should also be given.

(2) Definitions of speech acts as operators in the planning system. What are their effects? When are they applicable? How can they be realized in words?

---

* This research was supported in part by the National Research Council of Canada.

To make possible a first attempt at such a theory we have imposed several restrictions on the system to be modelled.

(1) Any agent A1's model of another agent A2 is defined in terms of "facts" that A1 believes A2 believes, and goals that A1 believes A2 is attempting to achieve. We are not attempting to model obligations, feelings etc.

(2) The only speech acts we try to model are some that appear to be definable in terms of beliefs and goals, namely REQUEST and INFORM. We have been taking these to be prototypical members of Searle's "directive" and "representative" classes (Searle (1976)). We represent questions as REQUESTs to INFORM. These acts are interesting for they have a wide range of syntactic realizations, and account for a large proportion of everyday utterances.

(3) We have limited ourselves so far to the study of so-called task-oriented dialogues which we interpret to be conversations between two agents cooperating in the achievement of a single high-level goal. These dialogues do not allow changes in the topic of discourse but still display a wide range of linguistic behaviour.

Much of our work so far has dealt with the problem of generating plans containing REQUEST and INFORM, as well as non-linguistic operators. Suppose that an agent is attempting to achieve some task, with incomplete knowledge of that task and of the methods to complete it, but with some knowledge of the abilities of another agent. How can the first agent make use of the abilities of the second? Under what circumstances can the first usefully produce utterances to transmit or acquire facts and goals? How can he initiate action on the part of the second?

We view the plan related aspects of language generation and recognition as indissociable, and strongly related to the process by which agents cooperate in the achievement of goals. For example, for agent2 to reply "It's closed" to agent1's query "Where's the nearest service station?" seems to require him to infer that agent1 wants to make use of the service station which he could not do if it were closed. The reply "Two blocks east" would be seen as misleading if given alone, and unnecessary if given along with "It's closed". Thus part of cooperative behaviour is the detection by one agent of obstacles in the plans he believes the other agent holds, possibly followed by an attempt to overcome them. We claim that speakers expect (and intend) hearers to operate this way and therefore that any hearer can assume that inferences that he can draw based on knowledge that is shared with the speaker are in fact intended by the speaker. These processes underlie our analysis of indirect speech acts (such as "Can you pass the salt?") - utterances which appear to result from one illocutionary act but can be used to perform another.

Section 2 of this paper outlines some requirements on the models which the various agents must have of each other. Section 3 describes the planning operators for REQUEST and INFORM, and how they can be used to generate plans which include assertions, imperatives, and several types of questions.

Section 4 discusses the relation between the operators of section 3 and the linguistic sentences which can realize them. We concentrate on the problem of identifying illocutionary force, in particular on indirect speech acts. A useful consequence of the illocutionary force identification process is that it provides a natural way to understand some elliptical utterances and utterances whose purpose is to acknowledge, correct or clarify interpretations of previous utterances.

A critical part of communication is the process by which a speaker can construct descriptions of objects involved in his plans such that the hearer can identify the intended referent. Why can someone asking "Where's the screwdriver? be answered with "In the drawer with the hammer" if it is assumed he knows where the hammer is, but maybe by "In the third drawer from the left" if he doesn't. How accurate must descriptive phrases be? Section 5 examines how the speaker and hearer's models of each other influence their references. Finally section 6 contains some ideas on future research.

Most examples in the paper are drawn from a situation in which one participant is an information clerk at a train station, whose objective is to assist passengers in boarding and meeting trains. The domain is obviously limited, but still provides a natural setting for a wide range of utterances, both in form and in intention.

2. On models of others

In this section we present criteria that one agent's model of another ought to satisfy. For convenience we dub the agents SELF and OTHER. Our research has concentrated on modelling beliefs and goals. We claim that a theory of language need not be concerned with what is actually true in the real world: it should describe language processing in terms of a person's beliefs about the world. Accordingly, SELF's model of OTHER should be based on "believe" as escribed, for example, in Hintikka(1962) and not on "know" in its sense of "true belief".

Henceforth, all uses of the words "know" and "knowledge" are to be treated as synonyms for "believe" and "beliefs". We have neglected other aspects of a model of another, such as focus of attention (but see Grosz(1977)).

Belief

Clearly, SELF ought to be able to distinguish his beliefs about the world from what he believes other believes. SELF ought to have the possibility of believing a proposition P, of believing not-P, or of being ignorant of P. Whatever his stand on P, he should also be able to believe that OTHER can hold any of these positions on P. Notice that such disagreements cannot be represented if the representation is based on "know" as in Moore(1977).

SELF's belief representation ought to allow him to represent the fact that OTHER knows whether some proposition P is true, without SELF's having to know which of P or ~P he does believe. Such information can be represented as a disjunction of beliefs (e.g., OR(OTHER BELIEVE P, OTHER BELIEVE ~P)). Such disjunctions are essential to the planning of yes no questions.

Finally, a belief representation must distinguish between situations like the following:

1. OTHER believes that the train leaves from gate 8.
2. OTHER believes that the train has a departure gate.
3. OTHER knows what the departure gate for the train is.

Case 1 can be represented by a proposition that contains no variables. Case 2 can be represented by a belief of a quantified proposition -- i.e.

OTHER BELIEVE (
    ∃ x (the y : GATE(TRAIN,y) = x))

However, case 3 is represented by a quantified belief namely,

∃ x OTHER BELIEVE
        (the y : GATE(TRAIN,y) = x)

The formal semantics such beliefs have been problematic for philosophers (cf. Quine (1956) and Hintikka (1962)). Our approach to them is discussed in Cohen (1978). In Section 3, we discuss how quantified beliefs are used during planning, and how they can be acquired during conversation.

Want

Any representation of OTHER's goals (wants) must distinguish such information from: OTHER's beliefs, SELF's beliefs and goals, and (recursively) from the other's model of someone else's beliefs and goals. The representation for WANT must also allow for different scopes of quantifiers. For example, it should distinguish between the readings of "John wants to take a train" as "There is a specific train which John wants to take" or as "John wants to take any train". Finally it should allow arbitrary embeddings with BELIEVE. Wants of beliefs (as in "SELF wants OTHER to believe P") become the reasons for telling P to OTHER, while beliefs of wants (e.g., SELF Believes SELF wants P) will be the way to represent SELF's goals P.

Levels of Embedding

A natural question to ask is how many levels of belief embedding are needed by an agent capable of participating in a dialogue. Obviously, to be able to deal with a disagreement, SELF needs two levels (SELF BELIEVE and SELF BELIEVE OTHER BELIEVE). If SELF were to lie to OTHER, he would have to be able to believe some proposition P (i.e. SELF BELIEVE (P)), while OTHER believes that SELF believes not P (i.e. SELF BELIEVE OTHER BELIEVE SELF BELIEVE (~P)), and hence he would need at least three levels.

We show in Cohen (1978) how one can represent, in a finite fashion, the unbounded number of beliefs created by any communication act or by face-to-face situations. The finite representation, which employs a circular data structure, formalizes the concept of mutual belief (cf. Schiffer (1972)). Typically, all these levels of belief embedding can be represented in three levels, but theoretically, any finite number are possible.

3. Using a Model of the Other to Decide What to Say

As an aid in evaluating speech act definitions, we have constructed a computer program, OSCAR, that plans a range of speech acts. The goal of the program is to characterize a speaker's capacity to issue speech acts by predicting for specified situations, all and only those speech acts that would be appropriately issued by a person under the circumstances. In this section, we will make reference to prototypical speakers by way of the OSCAR program, and to hearers by way of the program's user.

Specifially, the program is able to:

Plan REQUEST speech acts, for instance a speech act that could be realized by

"Please open the door", when its goal is to get the user to want to perform some action.

- Plan INFORM speech acts, such as one that could be realized by "The door is locked", when its goal is to get the user to believe some proposition.

- Combine the above to produce multiple speech acts in one plan, where one speech act may establish beliefs of the user that can then be employed in the planning of another speech act.

- Plan questions as requests that the user inform, when its goal is to believe something and when it believes that the user knows the answer.

- Plan speech acts incorporating third parties, as in "Ask Tom to tell you where the key is and then tell me."

To illustrate the planning of speech acts, consider first the following simplified definitions of REQUEST and INFORM as STRIPS-like operators (cf. Fikes and Nilsson (1971)). Let SP denote the speaker, H the hearer, ACT some action, and PROP some proposition. Due to space limitations, the intuitive English meanings of the formal terms appearing in these definitions will have to suffice as explanation.

REQUEST(SP,H,ACT)
preconditions:
SP BELIEVE H CANDO ACT
SP BELIEVE H BELIEVE H CANDO ACT
SP BELIEVE SP WANT TO REQUEST
effects:
H BELIEVE SP BELIEVE SP WANT H TO ACT

INFORM(SP,H,PROP)
preconditions:
SP BELIEVE PROP
SP BELIEVE SP WANT TO INFORM
effects:
H BELIEVE SP BELIEVE PROP

The program uses a simplistic backward-chaining algorithm that plans actions when their effects are wanted as subgoals that are not believed to be true. It is the testing of preconditions of the newly planned action before creating new subgoals that exercises the program's model of its user. We shall briefly sketch how to plan a REQUEST.

Every action has "want preconditions", which specify that before an agent does that action, he must want to do it. OSCAR plans REQUEST speech acts to achieve precisely this precondition of actions that it wants the user to perform. Similarly, the goal of the user's believing some proposition PROP becomes OSCAR's reason for planning to INFORM him of PROP.

Suppose, for example, that OSCAR is outside a room whose door is closed and that it believes that the user is inside. When planning to move itself into the room, it might REQUEST that the user open the door. However, it would only plan this speech act if it believed that the user did not already want to open the door and if it believed (and believed the user believed) that the preconditions to opening the door held. If that were not so, OSCAR could plan additional INFORM or REQUEST speech acts. For example, assume that to open a door one needs to have the key and OSCAR believes the user doesn't know where it is. Then OSCAR could plan "Please open the door. The key is in the closet". OSCAR thus employs its user model in telling him what it believes he needs to know.

Mediating Acts and Perlocutionary Effects

The effects of INFORM (and REQUEST) are modelled so that the hearer's believing P (or wanting to do ACT) is not essential to the successful completion of the speech act. Speakers, we claim, cannot influence their hearers' beliefs and goals directly. Thus, the perlocutionary effects of a speech act are not part of that act's definition. We propose, then, as a principle of communication that a speaker's purpose in sincere communication is to produce in the hearer an accurate model of his mental state.

To bridge the gap between the speech acts and their intended perlocutionary effects, we posit mediating acts, named CONVINCE and DECIDE, which model what it takes to get someone to believe something or want to do something. Our current analysis of these mediating acts trivializes the processes that they are intended to model by proposing that to convince someone of something, for example, one need only get that person to know that one believes it.

Using Quantified Beliefs -- Planning Questions

Notice that the precondition to OSCAR's getting the key -- knowing where it is -- is of the form:

$\exists$ x OSCAR BELIEVE
(the y : LOC(KEY,y) = x)

When such a quantified belief is a goal, it leads OSCAR to plan the question "Where is the key?" (i.e., REQUEST(OSCAR, USER, INFORM(USER, OSCAR, the y : LOC(KEY,y))). In creating this question, OSCAR first plans a CONVINCE and then plans the user's INFORM speech act, which it then tries to get him to perform by way of requesting.

The above definition of INFORM is inadequate for dealing with the quantified beliefs that arise in modelling someone else. This INFORM should be viewed as that version of the speech act that the planning agent (e.g., OSCAR) plans for itself to perform. A different view of INFORM, say INFORM-BY-OTHER, is necessary to represent acts of informing by agents other than the speaker. The difference between the two INFORMS is that for the first, the planner knows what he wants to say, but he obviously does not have such knowledge of the content of the second act.

The precondition for this new act is a quantified speaker-belief:

∃ x USER BELIEVE
(the y : LOC(KEY,y) = x)

where the user is to be the speaker. For the system to plan an INFORM-BY-OTHER act for the user, it must believe that the user knows where the key is, but it does not have to know that location! Similarly, the effects of the INFORM-BY-OTHER act is also a quantified belief, as in

∃ x OSCAR BELIEVE
USER BELIEVE
(the y LOC(KEY,y) = x)

Thus, OSCAR plans this INFORM-BY-OTHER act of the key's location in order to know where the user thinks the key is.

Such information has been lacking from all other formulations of ASK (or INFORM) that we have seen in the literature (e.g., Schank (1975), Mann et al. (1976), Searle (1969)). Cohen (1978) presents one approach to defining this new view of INFORM, and its associated mediating act CONVINCE

4. Recognizing Speech Acts

In the previous section we discussed the structure of plans that include instances of the operators REQUEST and INFORM without explaining the relation between these speech acts and sentences used to perform them. This section sketches our first steps in exploring this relation. We have been particularly concerned with the problem of recognizing illocutionary force and propositional content of the utterances of a speaker. Detailed algorithms which handle the examples given in this section have been designed by J. Allen and are being implemented by him. Further details can be found in (Allen and Perrault 1978) and Allen's forthcoming Ph.D. dissertation.

Certain syntactic clues in an utterance such as its mood and the use of explicit performatives indicate what act the speaker intends to perform, but, as is well known, utterances which taken literally would indicate one illocutionary force can be used to indicate another. Thus "Can you close the door?" can be a request as well as a question. These so-called indirect speech acts are the acid test of a theory of speech acts. We claim that a plan-based theory gives some insight into this phenomenon.

Searle(1975) correctly suggests that "In cases where these sentences <indirect forms of requests> are uttered as requests, they still have their literal meaning and are uttered with and as having that literal meaning'. How then can they also have their indirect meaning?

Our answer relies in part on the fact that an agent participating in a cooperative dialogue must have processes to:

(1) Achieve goals based on what he believes.
(2) Adopt goals of other agents as his own.
(3) Infer goals of other agents.
(4) Predict future behaviour of other agents.

These processes would be necessary even if all speech acts were literal to account for exchanges where the response indicates a knowledge of the speaker's plan. For example

Passenger: "When does the next train to Montreal leave?"
Clerk : "At 6:15 at Gate 7"
or
Clerk : "There won't be one until tomorrow."

Speakers expect hearers to be executing these processes and they expect hearers to know this. Inferences that a hearer can draw by executing these processes based on information he thinks the speaker believes can be taken by the hearer to be intended by the speaker. This accounts for many of the standard examples of indirect speech acts such as "Can you close the door?" and "It's cold here" For instance, even if "It's cold here" is intended literally and is recognized as such, the helpful hearer may still close the window. When the sentence is uttered as a request, the speaker intends the hearer to recognize the speaker's intention that the hearer should perform the helpful behaviour

If indirect speech acts are to be explained in terms of inferences speakers can expect of hearers, then a theory of speech acts must concern itself with how such inferences are controlled. Some heuristics are particularly helpful. If a chain of inference by the hearer has the speaker planning an action whose effects

are true before the action is executed, then the chain is likely to be wrong, or else must be continued further. This accounts for "Can you pass the salt?" as a request for the salt, not a question about salt-passing prowess. As Searle(1975) points out, a crucial part of understanding indirect speech acts is being able to recognize that they are not to be interpreted literally.

A second heuristic is that a chain of inference that leads to an action whose preconditions are known to be not easily achievable is likely to be wrong.

Inferencing can also be controlled through the use of expectations about the speaker's goals. Priority can be given to inferences which relate an observed speech act to an expected goal. Expectations enable inferencing to work top-down as well as bottom-up.

The use of expected goals to guide the inferencing has another advantage: it allows for the recognition of illocutionary force in elliptical utterances such as "The 3:15 train to Windsor?", without requiring that the syntactic and semantic analysis "reconstitute" a complete semantic representation such as "Where does the 3:15 train to Windsor leave?". For example, let the clerk assume that passengers want to either meet incoming trains or board departing ones. Then the utterance "The 3:15 train to Windsor?" is first interpreted as a REQUEST about a train to Windsor with 3:15 as either arrival or departure time. Only departing trains have destinations different from Toronto and this leads to believing that the passenger wants to board a 3:15 train to Windsor. Attempting to identify obstacles in the passenger's plan leads to finding that the passenger knows the time but probably not the place of departure. Finally, overcoming the obstacle then leads to an INFORM like "Gate 8"

Our analysis of elliptical utterances raises two questions First, what information does the illocutionary force recognition module expect from the syntax and semantics? Our approach here has been to require from the syntax and semantics a hypothesis about the literal illocutionary force and a predicate calculus-like representation of the propositional content, but where undetermined predicates and objects could be replaced by patterns on which certain restrictions can be imposed. As part of the plan inferencing process these patterns become further specified.

The second question is: what should the hearer do if more than one path between the observed utterance and the expectations is possible? He may suspend plan deduction and start planning to achieve a goal which would allow plan deduction to continue. Consider the following example.

Passenger : When is the Windsor train?
Clerk : The train to Windsor?
Passenger : Yes.
Clerk : 3:15.

After the first sentence the clerk cannot distinguish between the expectations "Passenger travel by train to Windsor" and "Passenger meets train from Windsor", so he sets up a goal : (clerk believes passenger wants to travel) or (clerk believes passenger wants to meet train). The planning for this goal produces a plan that involves asking the passenger if he wants one of the alternatives, and receiving back the answer. The execution of this plan produces the clerk response "The train to Windsor?" and recognizes the response "Yes". Once the passenger's goal is known, the clerk can continue the original deduction process with the "travel to Windsor" alternative favoured. This plan is accepted and the clerk produces the response "3:15" to overcome the obstacle "passenger knows departure time"

5. Reference and the Model of the Other

We have shown that quantified beliefs are needed in deciding to ask someone a question. They are also involved, we claim, in the representation of singular definite noun phrases and hence any natural language system will need them. According to our analysis, a hearer should represent the referring phrase in a speaker's statement "The pilot of TWA 510 is drunk" by:

$$\exists\, x \text{ SPEAKER BELIEVE } (\text{the } y \; \text{PILOT}(y,\text{TWA510}) = x \;\&\; \text{DRUNK}(x))$$

This is the reading whereby the speaker is believed to "know who the pilot of TWA 510 is" (at least partially accounting for Donnellan's (1966) referential reading). This is to be contrasted with the reading of whoever is piloting that plane is drunk (Donnellan's attributive noun phrases). In this latter case, the existential quantifier would be inside the scope of the belief.

These existential presuppositions of definite referential noun phrases give one important way for hearers to acquire quantified speaker-beliefs. Such beliefs, we have seen, can be used as the basis for planning further clarification questions.

We agree with Strawson (1950) (and many others) that hearers understand referring phrases based on what they believe speakers intend to refer to.

Undoubtedly, a hearer will understand a speaker's (reference) intentions by using a model of that speaker's beliefs. Speakers, of course, know of these interpretation strategies and thus plan their referring phrases to take the appropriate referent within the hearer's model of them. A speaker cannot use private descriptions, nor descriptions that he thinks the hearer thinks are private, for communication.

For instance, consider the following variant of an example of Donnellan's (1966): At a party, a woman is holding a martini glass which Jones believes contains water but of which he is certain everyone else believes (and believes he believes) contains a martini. Jones would understand that Smith, via question (1), but not via question (2) is referring to this woman.

(1) Who is the woman holding the martini?
(2) Who is the woman holding the water?

since Jones does not believe Smith knows about the water in her glass.

Conversely, if Jones wanted to refer to the woman in an utterance intended for Smith, he could do so using (1) but not (2) since in the latter case he would not think the hearer could pick out his intended referent.

Thus it appears that for a speaker to plan a successful singular definite referential expression requires that the speaker believe the expression he finally chooses have the right referent in the hearer's model of the speaker. Our concept of mutual belief can be used (as in Cohen (1978)) to ensure that the expression denotes appropriately in all further embedded belief models. This example is problematic for any approach to reference where a communicating party assumes that its reality is the only reality. Speakers and hearers can be "wrong" or "ignorant" and yet communication can still be meaningful and successful.

6. Further Research

We believe that speech acts provide an excellent way of explaining the relations between utterances in a dialogue, as well as relating linguistic to non-linguistic activity. Until we better understand the mechanisms by which conversants change the topic and goals of the conversation it will be difficult to extend this analysis beyond exchanges of a few utterances, in particular to non-task oriented dialogues. Fuller justification of our approach also requires its application to a much broader range of speech acts. Here the problem is mainly representational: how can we handle promises without first dealing with obligations, or warnings without the notions of danger and undesirability? We are currently considering an extension of the approach to understanding stories which report simple dialoque.

Much remains to be done on the representation of the abilities of another agent. A simple setting suggests a number of problems. Let one agent H be seated in a room in front of a table with a collection of blocks. Let another agent S be outside the room but communicating by telephone. If S believes that there is a green block on the table and wants it cleared, but knows nothing about any other blocks except that H can see them, then how can S ask H to clear the green block? The blocks S wants removed are those which are in fact there, perhaps those which he could perceive to be there if he were in the room. The goal seems to be of the form

S BELIEVE
$\forall$ x (x on the green block => S WANT
(x removed from green block))

but our planning machinery and definition of REQUEST are inadequate for generating "I request you to clear the green block"

We have not yet spent much time investigating the process of giving answers to How and Why questions, or to WH questions requiring an event description as an answer. We conjecture that because of the speech act approach answers to "What did he say?" should be found in much the same way as answers to "What did he do?" and that this parallelism should extend to other question types. The natural extension of our analysis would suggest representing "How did AGT achieve goal G?" as a REQUEST by the speaker that the hearer inform him of a plan by which AGT achieved G. We have not yet investigated the repercussions of this extension on the representation language.

Finally consider the following dialogue. Assume that S is a shady businessman A his secretary.

A : IRS is on the phone

S : I'm not here

How is A to understand S's utterance? Although its propositional content is literally false, maybe even nonsensical, the utterance's intention is unmistakable. How tolerant does the understanding system have to be to infer its way to a correct interpretation? Must "I'm not here" be treated idiomatically?

Bibliography

Allen, J.F. and Perrault, C.R. "Participating in Dialogue: Understanding via Plan Deduction", 2nd National Conference of the Canadian Society for Studies in Computational Intelligence Toronto, July, 1978.

Cohen, P.R., "On Knowing What to Say: Planning Speech Acts", TR118 Dept. of Computer Science, University of Toronto, 1978.

Donnellan, K., "Reference and Definite Description", The Philosophical Review, vol. 75, 1960, pp280-304. Reprinted in Semantics Steinberg and Jacobovits, eds., Cambridge University Press, 1970.

Fikes, R. E. and Nilsson, N. J., 1970, "STRIPS: A new approach to the application of theorem proving" Artificial Intelligence 2, 1970.

Grosz, B. J., "The Representation and Use of Focus in Natural Language Dialogues", 5IJCAI, 1977.

Hintikka, K.J., Knowledge and Belief, Cornell University Press, 1962.

Mann, W.C., Moore, J.A., Levin, J.A.; "A Comprehension Model for Human Dialogue", 5IJCAI, 1977.

Moore, R.C.; "Reasoning about Knowledge and Action", 5IJCAI, 1977.

Quine W.V., "Quantifiers and Propositional Attitudes", The Journal of Philosophy 53, (1956), 177-187.

Schiffer, S., Meaning, Oxford University Press, 1972.

Schank, R. and Abelson, R., "Scripts, Plans and Knowledge", 4IJCAI, 1975.

Searle, J. R., Speech Acts, Cambridge University Press, 1969.

Searle, J. R.; "Indirect Speech Acts" in Syntax and Semantics, Vol. 3: Speech Acts, Cole and Morgan (eds), Academic Press, 1975.

Searle, J. R., "A Taxonomy of Illocutionary Acts", Language, Mind and Knowledge, K. Gunderson (ed.) University of Minnesota Press, 1976.

Strawson, P. F., "On Referring", Mind 1950.

A Framework for Comparing Language Experiences (with particular emphasis on The Effect of Audience on Discourse Models)

Andee Rubin
Bolt Beranek and Newman Inc.
Cambridge, Massachusetts

Those of us who have been involved in the study of language have often thrown up our hands in dismay at the complexity of the problem (at least I have, almost daily) and tried somewhat desperately to find some facet of the many-faced gem we confront which appears manageable. This desire to focus - to train the flashlights we use to illuminate the problem on a well-circumscribed area - led, for example, to the early transformational grammarians' separation of syntax from the rest of language. Many researchers have since discarded that particular focus and attempted to integrate syntax, semantics and pragmatics in a single theory. Yet, even they end up focusing on a particular kind of linguistic interaction. Many study oral conversations, some look at computer-person dialogues, some study newspaper articles. These limitations in scope are often not explicit, but are reflected in the examples they choose and discuss. Very few people study more than one communicative situation and even if they do, they do not usually analyze the similarities and differences among them.

Just as the early transformational focus on syntax resulted in a model which missed many crucial insights about language, so does our current research risk formulating incomplete and even inaccurate models by focusing on certain communicative situations without adequate insight into their relationship to others. My own focus in an attempt to point out what such a narrow view might miss, and to provide a framework in which to examine a wide variety of language experiences and discover what effect the differences among them have on theories of language. This paper first presents the framework, then examines further one dimension of language experience -- the audience for an utterance -- as an example of the kind of considerations such an analysis suggests.

1. Building the Space of Language Experiences

Consider first two language experiences commonly studied by present-day investigators: face-to-face oral conversations and computer-person dialogues. These two situations differ in at least two ways: the modality of the interaction (current computer-person dialogues are written) and the lack of possibility of communicating with extra-linguistic devices such as gestures and facial expressions. The kinds of questions we want to be able to address are:

What are the effects of these two distinctions on the language used in each of these situations? What are the effects on the models of language use which we thereby formulate?

In order to capture these kinds of differences in a way which will enable us to approach these questions, I have used the metaphor of a multi-dimensional space. Each language experience lies at a point in the space defined by its position along several dimensions of the linguistic medium. The medium of a language experience is defined in contrast to its message; in as much as they can be separated, the message is what is communicated, while the medium is how it is communicated. Further, the medium here is expressed in experiential terms and does not represent simply the vehicle for the message; for example, the contrast is made between being in a conversation and watching a play, rather than between a conversation and a play. (1)

Consider as a starting point in building the space the following message: an invitation and directions to a party. One common linguistic situation in which this might be communicated is a face-to-face oral conversation. Conversations, however, do not need to be oral; it is quite possible to maintain all the aspects of a conversation while writing it down by, for example, passing notes. These two language experiences form a "minimal pair"; that is, they differ along

(1) I have omitted from this discussion any consideration of the message communicated in a language experience. In Rubin (1978) I identify three message-related dimensions - structure, function and topic - and discuss their interactions with the medium-related dimensions introduced here.

only one dimension. We might represent the situation graphically by labelling the line connecting the two experiences with the dimension along which they differ. The relevant dimension in this diagram is modality i.e. whether a language experience is oral or written. (Modality will be further dissected below.)

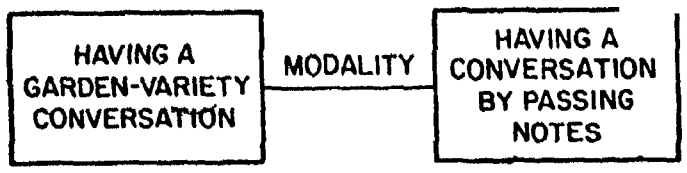


Figure 1.

Let us now look at two other pairs of language experiences which illustrate another dimension. Consider communicating the same message over the telephone compared to using a tape cassette. In the first case, it is possible for the two participants to interact, for the listener to express confusion and ask for additional information, for the speaker to monitor the listener's reactions and provide a more complete explanation. In the cassette situation, the speaker must decide once and for all how to give the directions without the benefit of intermediate feedback; any feedback which might occur would happen after the listener had heard the tape all the way through and would be temporally removed from the time the speaker composed the tape. I have termed this dimension of language experience interaction, as Figure 2 illustrates.

The other minimal pair in Figure 2, which also illustrates the interaction dimension, is communicating by letter versus communicating by a conversation over teletypes (while this is a somewhat unusual communicative setting, many of the people reading this paper have probably participated in it). Here again the crucial difference between the two is the possibility of feedback. In this particular task, for example, the speaker might want to ask of the listener, "Do you know the corner of Lewis and Fairview?" and base her further explanation on the response. Such an exchange would be impossible in the case of a letter

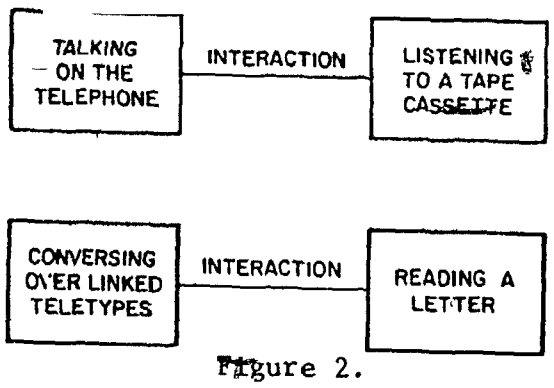


Figure 2.

Notice now that we can connect the two minimal pairs in Figure 2 by lines labelled "modality". Reading a letter and listening to a cassette form a minimal pair which differ only in modality; the same is true for teletype and telephone conversations. The modality and interaction axes together form a plane in which we can place these four language experiences.

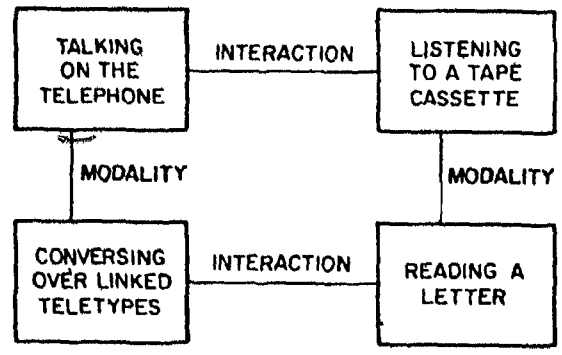


Figure 3

Other dimensions can similarly be added to this space, by first discovering a minimal pair which focuses on a particular dimension, then attempting to pinpoint each language experience already in the space on that dimension and finally filling in the holes which exist because of the added axis. As an example of one step in building the space, consider the dimension of extra-linguistic communication, that is, communication by gestures, facial expressions, etc. For the message we are considering, gestures would be particularly useful to indicate spatial features such as "right" and "left" and the relative location of objects and landmarks. None of the four media in Figure 3 admit this type of interaction, but for each of them it is possible to construct an experience which differs from it along only this new dimension. For example, garden-variety conversations differ from telephone conversations because they allow this extra dimension, and passing notes differs from conversing over a teletype link in the same way. We now see where the pair of language experiences illustrated in Figure 1 comes in and by adding two more nodes we get the following cube.

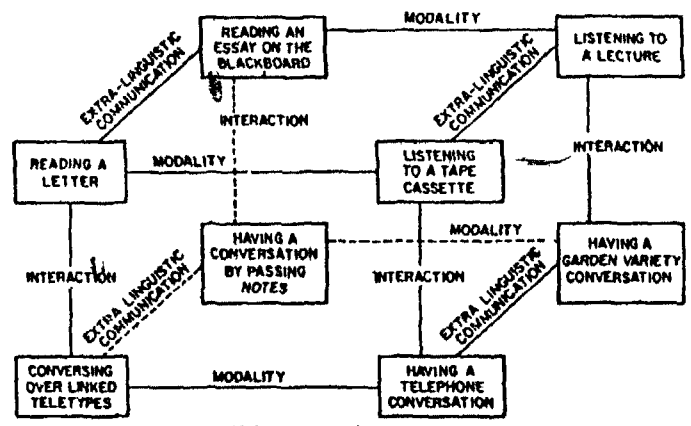


Figure 4.

The language experiences in this subspace differ in their degrees of naturalness; writing on the blackboard in such a way as to allow extra-linguistic communication but prohibit interaction, for example, seems contrived. This awkwardness is due primarily to the fact that certain dimensions generally covary and that the independence implied by a dimensional analysis doesn't really hold. Section 3 below discusses these interdependencies in more detail.

2. The Dimensions of Language Experience

In a similar fashion, we can add more dimensions to the space of language experiences. I have so far identified eight separate dimensions along which the medium of a language experience may vary. The dimensions are at least semi-independent; my informal criterion for listing a dimension separately was the existence of some minimal pair of language experiences whose media differed only along that dimension. The medium-related dimensions of language experience are: modality, interaction, extra-linguistic communication, spatial commonality, temporal commonality, concreteness of referents, separability of characters and specificity of audience. Below, each dimension is explicated by means of the question one would have to ask about a language experience to correctly place it on that dimension and additional details about its substructure and ramifications are given.

1. MODALITY - Is the message written or spoken? Even in this seemingly simple dimension are hidden at least two different characteristics which affect the communicative situation: prosody and permanence. I will briefly discuss these here, but a more extensive discussion of the components of modality may be found in Schallert, Kleiman & Rubin (1977) and Sticht, Beck, Hauke, Kleiman & James (1974) provide a review of the literature comparing auding (comprehension of oral language) and reading.

An obvious difference between an oral utterance and a text is the availability of prosodic cues. Temporal characteristics of speech such as pauses and changes in speed provide clues for the chunking of words into larger constituents. In general, pauses and breaths occur at syntactic boundaries and a more quickly spoken set of words may indicate an appositive phrase or an aside which is not germane to the top-level structure of the sentence. We rely on stress in oral language as an indicator of such discourse organizing topics as given/new, contrast and focus, as well as to aid in the disambiguation of pronominal references. Intonation is often used as an indication of the illocutionary force of an utterance or to communicate affective qualities of language such as humor or sarcasm.

While written text clearly lacks these properties, it has some compensating features. Punctuation and other textual devices provide a partial analogue of many prosodic features, including illocutionary force (. ? ' !), pauses (;), lists (, : ;), related statements (;) and contrast and emphasis (underlining and italicizing). A written message also provides the recipient with concretely indicated segments both on the lower levels of word and sentence and on the more abstract level of paragraph and section structure.

The second major distinction included in modality is the permanence of written text in contrast to the transitory nature of oral language. This permanence makes possible various "good reading" techniques such as skimming ahead to look at chapter headings, re-reading an entire paragraph whose point became clear only at the last sentence, or just re-reading a sentence which was misparsed the first time around. In oral language situations, such heuristics for dealing with misunderstanding are often replaced by an appeal to another (independent) characteristic of language experiences - interaction.

2. INTERACTION - Are the participants able to interact? In an interactive language experience, participants have the opportunity to indicate that they have not understood a previous utterance, that a pronominal or other reference is ambiguous or that they wish to change the topic. Keenan and Schieffelin (1976) in particular have represented the establishment of discourse topic as a dynamic process which includes input from all participants. This possibility for interaction means that there is less necessity for a participant to entertain and maintain a set of competing hypotheses about the meaning of some part of the message.

3. EXTRA-LINGUISTIC COMMUNICATION - Can the participants communicate via extra-linguistic means which require visual or tactile interaction? (This communication may sometimes be one-way, as in the case of a lecturer speaking to a large class.) Gestures, facial expressions and even body positions are all powerful in their communicative potential. In situations where emotions or spatial attributes are being communicated, these extra-linguistic means may be especially relevant. Children's early language experiences are especially dependent on this aspect of communication; deLaguna (1927) describes one developmental thread in children's language use as "a progressive freeing of speech from dependence on the perceived conditions under which it is uttered and heard, and from the behavior which accompanies it."

4. SPATIAL COMMONALITY - Can the participants interpret spatial deictic words such as "here" and "there" with reference to their own location? One indication that situations in which this condition is not met are difficult is the well-known situation in which two people, having arranged over the telephone to meet "here", discover that they had two different places in mind. Because the listener had to interpret the speaker's use of "here" relative to the speaker's location, it was necessary for her to know where the speaker was; incorrect information in this situation can have serious consequences. Young children may actually interpret "here" and "there" relative to their own position, rather than the speaker's. (see Tanz (1976) for details)

5. TEMPORAL COMMONALITY - Do the participants share a temporal context which allows for simple interpretation of temporal deictic terms such as "now", "today" and "last Sunday"? The correct interpretation of such words, as well as verb tense markers, requires the reader/listener to take the temporal point of view of the speaker/writer.

6. CONCRETENESS OF REFERENTS - Are the objects and events referred to visually present for the participants? If an object or event is concrete, many of its details are immediately apparent to the reader/listener besides the ones linguistically described in the message. Reading or hearing about an object which is not present often requires remembering a partial, incomplete description and then reformulating it as more information becomes available. Objects (or pictures) also provide an external "memory" for their existence and properties.

7. SEPARABILITY OF CHARACTERS Are the distinctions among different people's statements and points of view clearly indicated? In face-to-face conversations, such distinctions are obvious, as each person makes his own statements and each point of view has a physical "anchor" In reading a play, characters' lines are clearly marked, although there is no physical object to which to attach each character. In a book, the reader must parcel out comments, feelings and motivations to characters on the basis of more subtle clues: punctuation, paragraph structure and inferences based on some consistent model of each of the characters.

8. SPECIFICITY OF AUDIENCE - How complete and specific is the speaker's model of the audience for her message? Two extremes which illustrate this dimension are garden- variety conversations in which the speaker and hearer know each other well and books, which are written for wide non-specific audiences. In the former case, references to shared knowledge are possible, such as "The man looked like Uncle Joe," while in the latter, such an attempt would surely miss a large portion of the audience. To make matters worse, often a writer (or speaker) does not know who the audience is likely to be and in the case of books which are several hundred years old, the intended audience differs in significant ways from current readers.

Now that we have these eight dimensions, we can use them to generate new language experiences which begin to fill up the space. Watching a play, reading a book with pictures, viewing a movie with subtitles, reading a comic book - all these fit in the eight-dimensional space we have defined. (In Rubin (1978) I discuss quite a number of "intermediate" language experiences and show how they fit into a multi-dimensional space.) However, some areas of the space are only sparsely filled. These relatively empty sections are indications of interactions between dimensions; given a particular position on one dimension, the choices for certain other dimensions may be sharply constrained. Descriptions of some of these interactions follow.

Interactions Among the Dimensions

One fairly obvious interdependency is etween spatial commonality and xtra-linguistic communication. Since both ely primarily on the participants being in he same place, it is not surprising that most anguage experiences which exhibit the otential for extra-linguistic communication lso allow participants straight-forward use f spatial deictic terms. (In fact, in Rubin 1978), I treat the two as a single imension.) However, in a note left, for xample, on the kitchen table, the writer may se "here" and "there", but cannot use estures or facial expressions, so the two imensions do not always co-occur.

Extra-linguistic communication is also most commonly found in oral language situations. The situation with the most potential for combining extra-linguistic communication with a written message is that of two people passing notes in what amounts to a written conversation. Although it is theoretically possible for them to point and grimace, it would be difficult for them to coordinate these gestures with the words in the written text.

Interaction and temporal commonality also appear closely linked. If the participants are not communicating in "real time" that is, if the sending and receiving take place at different times - then it might even seen impossible for them to interact. However, if we allow the kind of attenuated interaction that takes place in, for example, an exchange of letters, we can maintain these dimensions as at least semi-independent.

Finally, we note that specificity of audience and interaction have an interesting relationship. Less well-defined audiences tend to occur in situations in which interaction is difficult, if not impossible. In lectures, for example, the speaker has only a vague idea of the audience's beliefs and interaction between them is limited. This covariance reflects two different facts. One is that in a large (and therefore poorly-specified) audience, interaction is restricted simply because of its size. The other is that interaction is one device by which speakers construct better models of their audiences; thus, a lack of interaction would lead to less well-specified audiences

An obvious question at this point is: Why bother to separate dimensions that are so closely related? There are really two answers: the first methodological, the second historical. In terms of getting a clean model of the complex tangle of language experiences, it is better to postulate a large number of dimensions and specify how they interact than to identify only a small number but talk about subdimensions. Having a larger number of dimensions also inspires a wider range of language experiences when the process of filling in the space is carried out. Without the separation of temporal commonality and interaction, for example, we would have missed the subtle notion of temporally attenuated interaction.

The historical explanation derives from the original motivation for this work, which was an attempt to assess the relevance of children's early language experiences to their learning to read. Even if the dimensions identified here interact significantly, each still represents a cognitive skill which a child must learn in making the transition from garden-variety conversations to reading a text. In this framework, interactions among the dimensions are interesting because they represent pairs of skills which the child may have to learn together, rather than being able to separate them and learn one at a time.

Now that we have a notion where garden-variety dialogues fit into the framework of language experiences, it is possible to see what kinds of considerations we are liable to leave out if we focus only on conversations. While all of the dimensions identified above point out areas which deserve attention, I want to focus here on specificity of audience as an example of the ways our models must be stretched to account for the diversity of language experience.

4. Limitations and Compensation in Language Experiences: Non-Specific Audiences

Certain language experiences present problems for the participants, especially in comparison with garden-variety conversations, which have many communication-facilitating features. Lack of spatial commonality, for example, poses extra difficulty in the interpretation of certain deitic words, and the absence of non-verbal communication in telephone conversations makes expressing emotion especially hard. In some cases, an aspect of the medium itself provides compensation for the limitations. Written text, for example, partially compensates for its lack of prosodic cues to structure by its permanence, which allows the reader to make several attempts at parsing and understanding the words on the page. In other cases, we ourselves take into account the limitations of the medium by expressing our message differently. In talking on the phone, for example, we express our emotions more explicitly, rather than relying on facial expressions to communicate them in more subtle ways

An important facilitating aspect of most garden-variety conversations is that they take place between people who have fairly good models of one another and who share a large set of beliefs. (See Cohen (1977) and Clark & Marshall (1978) for details on shared beliefs.) The disappearance of this feature in other language experiences can cause difficulties which require special attention from both speaker and listener. To get a feeling for the effect of an audience, consider the task of explaining the difference between analog and digital computers. Talking to a technically unsophisticated person makes the task hard enough, but it would be even more difficult if the audience were a large number of people with widely varying technical backgrounds. When one is faced with communicating with a person about whom one knows very little or, worse yet, an audience made up of many people with different beliefs it becomes necessary to use several compensating techniques to ensure that the message gets across I will describe below some of the heuristics both speaker/writers and listener/readers use to deal with complex and poorlyspecified audiences. (From here on, I will use the words "speaker" and "listener" to refer to speaker/writers and listener/readers, respectively.)

5. Speakers' Heuristics for Complex Audiences

The audience for an utterance may be poorly specified in two different ways: it may be a single person about whom the speaker has less than complete knowledge or a group of people, each of whom the speaker knows more or less well. The speaker's task is to construct an utterance which is comprehensible to those who perceive that they are part of the intended audience. The following are techniques by which speakers may accomplish this task.

Identify the Audience: In some cases, the speaker really wants to address her remarks to a single person, even though several people are physically present and "available" as audience. Straightforward techniques exist for identifying the audience in these situations: a speaker may simply look in the direction of the intended audience or address

her by name. In giving a technical talk to a large and varied audience, a speaker may analogously select a subset of those present as the audience for a particular remark by using a phrase like "for you linguists in the audience". A more subtle and interesting method for accomplishing the same goal is to include in the remark a reference which only some of the audience understands, thereby clueing the others in to the fact that they may not get anything out of the utterance. At a recent conference attended by linguists, computer scientists and psychologists, a speaker, in answering a question from a computer scientist, resorted to some technical language (what he actually said was, "It's EQ, not EQUAL"), even though he knew two-thirds of the audience would be lost. Afterwards, he remarked that he had also realized that it was precisely the computer scientists who were confused about the point he was trying to make, so the remark was doubly appropriate: it selected exactly the audience who needed to comprehend it.

Play It Safe: Cover the Audience: If he realizes that the audience consists of two or more definable subgroups, a speaker may choose to include several descriptions of the same topic, one for each set of people. In addressing an audience made up of computer scientists and psychologists, a speaker might refer to the same concept by two different terms, e.g. "cache memory" for the computer scientists and "working memory" for the psychologists. In this case, most members of the audience understand only one of the two descriptions so both are necessary. A slightly different situation exists when a speaker makes a statement such as, "He had eyes just like Paul Newman...deep, dark blue." Here the elaboration may be seen as a comment to those in the audience who don't know what Paul Newman's eyes look like. Those who do must realize that, in some sense, the elaboration was not directed at them since it was planned for listeners with different knowledge. A third example of the "play it safe" strategy is this conversation which took place in front of an audience:

B: Jerry has been studying the same thing.

M: That's right...that's Jerry Fodor...I read his paper.

Had B and M been conversing in private, the explanation that the Jerry being referred to was Jerry Fodor would have been unnecessary. However, M was aware that some people in the audience might not be able to figure out who was being referred to, so he played it safe and made the reference clear.

Embrace Ambiguity: Sometimes, the speaker is aware that an utterance may be interpreted in two different ways, but decides that the ambiguity is acceptable or even desirable. An example from a personal letter: "The weather has been beautiful...perfect for riding a motorcycle in the country." What the writer had in mind here was a particular time she and the addressee had taken a motorcycle ride; she wanted to allude to that event. If, however, he didn't remember the ride, the sentence still communicated a coherent, if less specific, message. The writer actually considered both of these possibilities and decided that either reading of the sentence was acceptable.

The following sentence taken from an advertising brochure shows a different kind of ambiguity: "The cane seats of a Mad River canoe provide excellent ventilation and drainage." The ad will be read both by people who already know that Mad River canoes have cane seats and by those for whom this is a new fact. For the latter group, the word "cane" obviously conveys new information; for the former group, perhaps, it focuses on that aspect of the seat which is relevant to the properties discussed. (I use the word "focuses" here in an informal sense, but it is similar to its use in Grosz (1977) and Sidner (1978). Again, the ambiguity exists because of the non-specificity of the audience, but both readings are acceptable and, in fact, desirable.

Rely on Interaction: The standard way speakers check whether or not they are being understood and modify their utterances appropriately is by interacting with the hearer. In the Paul Newman example above, the speaker could have watched for signs of recognition (a smile or a nod) from the listener which would have made the elaboration unnecessary. In an interactive situation, the motorcycle example could have been followed by "Do you remember that time?" Unfortunately, language experiences in which the audience is unknown or unknowable, such as books and lectures, are often those in which interaction is difficult. The presence of both a specific audience and interaction is a "positive feedback" situation in which communication is greatly facilitated. The absence of both, however, necessitates the adoption of some of the heuristics mentioned here - and even then, communication may be impaired

The heuristics identified above assume that the speaker takes the responsibility for the efficacy of the communication. Speakers in general, though, can assume that they share with the listener certain communicative principles of the type explicated by Grice (1975). They can similarly assume that listeners have certain heuristics for determining what assumptions the speaker is making about the audience and whether or not they as listeners fit those assumptions. Because they have this faith in their listeners, speakers sometimes just "broadcast" a remark, leaving it to the listener to decide who the intended audience is. Some of these listener's heuristics for interpreting broadcast utterances are described in the next section.

6. Hearers Heuristics for "Broadcast" Utterances

Integral to a listener's understanding of any utterance is a model of the speaker's model of the hearer. Where there is some mismatch between this model and the listener himself, it may be difficult for him to figure out what the speaker really intended to communicate. In language experiences where the listener suspects that the speaker is broadcasting -- that is, not being careful about specifying an audience - he must "broadreceive", using some heuristics for deciding whether or not he should consider the remark to be addressed to him.

One general technique a listener might use is to compare the intended effect of the utterance with his current state. If he has already fulfilled the effect, he can consider the remark to be addressed elsewhere. Commands have clear intended effects, so it is relatively simple for a listener to use this heuristic with such speech acts. A member of a congregation who is already standing when the minister says "Please stand up" will understand that the utterance is not meant for him, rather than yelling out, "I already am!". Signs are another medium in which this broadcast behavior is apparent. A sign asking patients to PLEASE REGISTER WITH THE RECEPTIONIST is clearly meant only for those who haven't already done so; if patients didn't use this heuristic, they would find themselves in an infinite loop of registering. A somewhat more complex example is the familiar airport announcement: Extinguish all smoking materials and have your boarding pass ready." which selects two different subsets of the audience which hears it - those who are smoking and those who do not yet have their boarding passes ready.

Deciding on the intended effect of an utterance is no mean trick, of course; the speech act literature, (e.g., Austin (1962), Searle (1969)), makes this clear. One interesting example of the interaction of these considerations with those of audience is a subway sign proclaiming SMOKING IS DANGEROUS TO YOUR HEALTH. The intended effect may be seen either as getting smokers to give up their habit or as telling smokers and non-smokers alike that smoking is not healthful. In the first case, the intended audience is smaller than in the second. Although this more restricted audience is implied by the use of "you" on the sign, it seems plausible that the informational effect and, therefore, the larger audience is intended as well. A non-smoker's interpretation of the sign is then a complex process, involving an awareness of the two possible audiences for the message.

Finally, and most obviously, speakers assume that hearers will make use of pragmatic clues to determine whether they are part of the intended audience. A volleyball player yelling "I'll set" assumes that only the members of her team will attend to the remark; she doesn't need to preface it with a direct address to her teammates. Similarly, the same player calling "You hit" hopes that the pragmatic context is strong enough that the one person who is the intended audience can identify herself.

7. Summary

This admittedly brief discussion of the specificity of audiences is meant to illustrate the kinds of considerations a narrow focus on a single language experience might overlook. It is clear from just these few examples that the process of planning a speech act must utilize heuristics like those listed above and that speakers' models of listeners must contain some explicit representation of the size and specificity of the audience. These insights would not have arisen had we restricted ourselves to two-person conversations. The multi-dimensional space developed in this paper provides eight dimensions which can provoke similar investigations and a framework in which to integrate the results.

Acknowledgements

I would like to thank Chip Bruce for comments on earlier versions of this paper, Mitch Marcus, Barbara Grosz and Al Stevens for help with examples, and Jill O'Brien for preparing the final document.

This research was supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116.

References

Austin, J.L. How to do Things with Words Harvard Universi Press Cambridge Massachusetts, 1962.

Clark, Herbert H. and Marshall, Catherine. Definite Reference and Mutual Knowledge. Paper presented at the Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting. University of Pennsylvania, May 1978

Cohen, Philip R. On Knowing What to Say: Planning Speech Acts. PhD Thesis, University of Toronto, Computer Science Department, 1978.

deLaguna, G Speech: Its function and development. College Park, Maryland: McGrath Co., 1970 (Reprint of 1927 edition).

Grice, H.P., Logic and Conversation. In D. Davidson and G. Harman (Eds.) The Logic of Grammar, Encino, California: Dickenson Publishing Company, 1975.

Grosz, Barbara J. The representation and use of focus in dialogue understanding, PhD Thesis, University of California at Berkeley, Computer Science Department, May 1977.

Keenan, E. & Scheffelin, B. Topic as a discourse notion. In Li (Ed.), Subject and topic. Academic Press, 1976.

Rubin, Ann D. A Theoretical Taxonomy of the Differences Between Oral and Written Language. To appear in R. Spiro, B. Bruce and W. Brewer (Eds.), Theoretical Issues in Reading Comprehension, Hillsdale, N.J.: Lawrence Erlbaum, 1978, in press. Also as Center for the Study of Reading, Technical Report No. 35, 1978.

Schallert, D.L., Kleiman, G.M.; & Rubin, A.D. Analyses of differences between written and oral language. Center for the Study of Reading, Technical Report No. 29, April 1977.

Searle, J.R. Speech acts: An Essay in the philosophy of language. Cambridge: Cambridge University Press, 1969.

Sidner, Candace L. A Progress Report on the Discourse and Reference Components of PAL. AI Memo 463, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1978.

Sticht, Thomas G., Beck, Lawrence J., Hauke, Robert N., Kleiman, Glenn M., and James, James H. Auding and Reading: A Developmental Model, Human Resources Research Organization, Alexandria, Virginia, 1974.

Tanz, C. Studies in the acquisition of deictic terms, PhD Thesis, University of Chicago, Department of Psychology, August, 1976.

Intentionality and Human Conversations*

Jaime G. Carbonell Jr

Department of Computer Science
Yale University - New Haven, Connecticut

Abstract

This paper analyzes principles of human conversation based on the conversational goals of the participants. Several conversational rules are proposed that seem crucial to the process of interpreting and generating conversations. These rules, as well as other aspects of the conversation process, are embodied in MICS, a computer program that generates one side of a conversation. The process model underlying MICS, and some illustrative output, are presented.

1) Formulating rules about human conversations.

This paper is an empirical approach to understanding the processes that underlie human conversations. Since the task of codifying all the knowledge required for modeling human discourse is monumental, we confine our approach to formulating rules about the conversational intent of utterances in the course of a dialog. This approach leads us to investigate the effects of shared assumptions and knowledge between the speakers the social and interpersonal relations of the speakers, and the inferences that must be made by both speakers in a conversation. We take a different approach to analyzing conversations than other research efforts, such as those adopting the speech-acts paradigm (Mann et al [1977]) or investigating task-specific dialogs (Grosz [1977]), in the hope that our new perspective will shed some light on otherwise obscure or neglected aspects of human discourse.

Consider the following conversation fragment between Bill and John, two college students sharing an apartment:

1) JOHN: Hi, what s new, Bill?
   BILL: I'm going to visit my folks tonight.

We can analyze Bill's utterance in Conversation Fragment (1) in terms of its immediate meaning, that is, a representation of Bill's utterance in Conceptual Dependency or some other meaning representation. This, however, is a very incomplete analysis of what Bill said. Why did Bill say that he was visiting his folks? Bill could just as easily have said, "I'm going to brush my teeth tonight." This utterance, however, doesn t answer John's question; brushing one's teeth is not "something new". Therefore, we could propose a rather simple conversational rule:

RULE 1: If a question is asked in the course of a conversation, the other participant should answer this question.

Rule 1, however, is a little too naive. Suppose Bill's answer was: "There are a few more microns of dust on the windowsill than the last time you asked me that question." This is indeed "something new", but we would think of Bill as a wise guy for answering the question literally rather than addressing what John "must have meant". What did John really mean? John must have been looking for something out of the ordinary and of some intrinsic importance. Let us propose a new rule to incorporate this principle:

RULE 2: In the formulation of an answer, the speaker should address the true significance of the question, not just its literal meaning.

What is the true significance of a question? In Conversation Fragment (1), Bill might have answered: "The J-particle angular momentum of +3/2 was confirmed today." John, a literature major who does not understand Physics, may not be inclined to continue the conversation. Therefore, Bill's answer is not what was called for, unless Bill intentionally wanted to end the conversation. This example suggests that Bill missed something in establishing the true significance of John's question. John did, indeed, explicitly ask to hear something new; implicitly he meant something important and out of the ordinary. The J-particle answer conforms to these requirements, but it is still an inappropriate response. Therefore, the true significance of John's answer must include John's conversational goal. Why did John ask "What's new"? The answer is, obviously, to start a conversation with Bill. Bill, being aware of this conversational goal, needs to choose an answer that attempts to initiate conversation. That is Bill should choose a topic of conversation that John can talk about and that John may be interested in. Conversational Rule (3) summarizes this discussion:

RULE 3: In introducing a new topic of conversation, the topic should be chosen so that both speakers have some knowledge and interest in its discussion.

The process of understanding the conversational import of an utterance may be conceptually divided into two primary subprocesses: 1) determine the conversational goal of the utterance, and 2) establish the real, often implicit, meaning of the utterance. Lehnert [1977] analyzes the process of establishing the real meaning of questions. Our analysis focuses on the conversational goals of the participants and the establishment of a shared knowledge base between the participants. It is this shared cultural, personal, and factual knowledge that the conversational participants leave implicit in each communication. To illustrate this fact, consider Conversational Fragment (2):

2) JOHN: Do you want to go out and try the bar at Monument Square?
   BILL: I'm going to visit my folks tonight

Real significance of Bill's utterance:
   i) No, I do not want to go to the Monument Square bar.
   ii) My reason for not wanting to go is that I made a previous commitment, and I cannot be in two places at once tonight.

iii) The previous commitment is a visit to my folks.
iv) I am telling you about the reason why I cannot go drinking with you rather than just saying "no" because I do not want you to get angry at me.
v) I may also wish to shift the topic of conversation to a discussion about my family.

Bill knows that John will interpret his answer so as to conclude its real significance; otherwise Bill would have chosen to explicitly state the real significance. How does Bill know that John will understand him correctly? Clearly Bill and John must share some common sense knowledge such as:

a) A person cannot be in two places at once.
b) Previous commitments should be honored.
c) If X's invitation or suggestion is turned down by Y without apparent reason, then X is likely to get upset at Y.
d) If a person introduces a new topic in a conversation, he may want to discuss the current topic further.

Both Bill and John are aware that they share a common cultural knowledge base. This knowledge is very crucial in determining what is said in the conversation. Bill must have considered (i) through (iv) before deciding that is was sufficient to say only (iii). How did Bill decide to say only (iii)? He must have concluded that John would infer (i), (ii) and (iv) without difficulty. Thus, Bill knew about John's general knowledge because of their common cultural background and their personal relation. Bill used this knowledge to decide what to say in the conversation.

In the course of a conversation, people make assumptions about each other's knowledge. It is sometimes easier to see what these conversational assumptions are when they turn out to be incorrect, as in the following example:

3) PETE: How are you going to vote on Proposition 13?

MARY: On what?

PETE: You know, the property tax limitation

MARY: Oh yeah. I'm not registered to vote. Which way were you trying to convince me to vote?

PETE: was hoping you would help me make up my mind.

MARY: Actually, I don't give a damn about politics.

At the beginning of the conversation Pete assumed that Mary knew what Proposition 13 was, that she was able to vote, that she would vote, and that she had already decided how to vote on Proposition 13. All of these assumptions turned out to be incorrect, and the course of the conversation turned towards clarifying the incorrect assumptions. This example is an instance of a more general rule of conversation:

RULE 4 If a participant in a conversation discovers that his assumptions about the shared knowlege between the two speakers is incorrect, then he will steer the conversation to
1) establish a common knowledge base on a specific topic, or
2) discover what their shared knowledge is in general, or
3) shift the conversational topic to some matter where a common knowledge base is more likely to exist, or
4) end the conversation.

The assumptions discussed thus far have been of a factual nature, but assumptions are also made about the conversational intent of the participants and about their interest in the conversational topic. Mary inferred Pete's conversational intent incorrectly: He was seeking advice, not trying to lobby for or against Proposition 13. Pete started the entire conversation on the wrong topic by assuming that Mary was interested in politics or taxes. A conversation about a topic that one of the participants finds uninteresting will usually digress to other topics or fizzle out as the uninterested party volunteers no new information, finds an excuse to do something else, or states outright that the conversation is boring (as was the case in our example)

Erroneous assumptions about conversational intent lead to misunderstandings because each speaker will address the perceived intent of the other speaker's utterance. It is, therefore, imperative to correctly infer the other speaker's conversational intentions in order for the conversation to proceed naturally. The type misunderstanding that often results from incorrectly perceived conversational intentions is, on occasion, exploited in creating certain types of jokes, as in example 4:

4) SON: Dad, I robbed the liquor store yesterday

DAD: How could you ever do such a thing, son.

SON: Well, I got me this gun, and I pointed at the cashier...

To illustrate the importance of the implicit conversational goals and shared knowledge between the participants in a conversation, we present a few more dialog framents between Bill and John, the two college students sharing an apartment. In each example, as in conversations (1) and (2), Bill utters the same response, but its meaning is significantly different, depending on the context of the conversation.

5) JOHN: Are you broke again? You are going to have to come up with your share of the rent this month.

BILL: I'm going to visit my folks tonight.

Meaning of Bill s utterance:
(i) Yes, I'm broke again.
(ii) Yes, I'll try to contribute my share of the rent.
(iii) My parents might give me some money if I ask them.
(iv) If I visit them and ask them in person I have a better chance of getting some money
(v) I'll visit them tonight and then I'll ask them for money.

When we read Conversation Fragment (5), we infer that Bill may be going to ask his parents for money. How do we do this? We do not share knowledge with Bill to the effect that his parents have money or that Bill is willing to ask them for money. The answer is based on a conversational rule:

RULE 5: The utterances in a conversation should be connected by continuity of topic, common conversational goals, and each participant addressing the intent of the utterances of the other participant.

Since the reader assumes that Rule (5) is true for Conversation Fragment (5), he concludes that there must be a connection between Bill needing money and the visit to his parents. The reader then infers the most likely connection: Bill will ask his parents for money. John must also make this inference based on Rule (5), unless he knows that Bill regularly visits his parents to ask for money. The significant point illustrated in example 5 is that the conversation focused the inference mechanism to find a connection between the respective utterances. Therefore, conversational principles can play an important role in focusing human reasoning processes. The principle of focusing inference processes on significant or interesting aspects of conversational utterances and events is developed into a theory of human subjective understanding in Carbonell [1978]

Let us continue with the conversational fragments between Bill and John:

6) JOHN: How come you never see your family

BILL: I'm going to visit my folks tonight

Meaning of Bill's utterance:
(i) I do visit my family.
ii) Supporting evidence: I'm going to visit them tonight.
(iii) Therefore what you just said is not true.

7) JOHN: Can I borrow your car? I got this heavy date tonight.

BILL: I'm going to visit my folks tonight.

Meaning of Bill's utterance:
Alternative I.
(i) No, you cannot borrow my car tonight.
(ii) I am going to visit my folks tonight.
(iii) I need to drive there.
(iv) The car cannot be in two places at once.

Alternative II.
(i) Yes, you can borrow my car tonight.
(ii) I am going to be at my folk's place, where I don't need to use it.

8) JOHN: Can I have the apartment to myself? I got this heavy date tonight.

BILL: I'm going to visit my folks tonight.

Meaning of Bill's utterance:
(i) Yes, you can have the apartment.
(ii) What you want is for me to be elsewhere.
(iii) I was planning on that anyway, since I am visiting my folks tonight.

Conversation fragments (6), (7) and (8) illustrate the degree to which the understanding of conversational utterances is expectation-driven. The expectations are generated from previous utterances according to rule 5; the topic, intent, and conversational goals introduced earlier in the conversation will be addressed by later utterances. In each case the same utterance on Bill's part is understood differently, depending on the context established by John's previous utterance. Utterances in a conversation do not usually have a meaning independent of the rest of the conversation; their meaning is part of the context of the entire conversation. Thus, it is easy to see why quoting only a short passage from a conversation (or a political speech) can give that passage has an entirely different meaning from what was originally intended.

The shared knowledge between two speakers depends on many different factors. Two speakers share a large amount of basic knowledge by merely being members of the human race (e.g. the basic drives that motivate humans such as hunger, self-preservation, etc.). More knowledge is shared if the two speakers are members of the same culture. (Much of the cultural and more basic human knowledge necessary to understand natural language is discussed in Schank and Ableson [1977].) If the two participants hold the same type of job, are professional colleages, or have the same special interests, then they will share some rather specific knowledge. Two people with the same special interests (such as football or radio-astronomy) will usually steer the conversation to a discussion of their common interests.

he topic of a conversation may drift to subject where the conversational participants share a great amount of knowledge.

Another factor that determines the knowledge shared by the participants in a conversation is their interpersonal relation i.e., how well they know each other. In conversational fragment (7), Bill's response can be interpreted in two different ways by the reader, but John will interpret his response unambiguously. John must know whether Bill's response means that Bill needs

the car or whether John is free to use it; otherwise, Bill would have been more specific in his answer.

Social relations and the perceived goals of conversational participants play an important role in interpreting the meaning of conversational utterances. Let us first consider the influence of the social relations between the two participants:

9) ARMY GENERAL: I want a juicy hamburger.
STAFF AIDE: Right away, sir!

10) 7-YEAR-OLD: I want a juicy hamburger.
MOTHER: Maybe next week. We are having chicken today.

11) PRISON INMATE 1: I want a juicy hamburger.
PRISON INMATE 2: Me too! Everything here tastes like cardboard

The utterance "I want a juicy hamburger" is interpreted differently in each dialog fragment. The difference in the interpretations is based on the different social relations existing between the two conversational participants. In Dialog (9) the utterance was interpreted to mean a direct order to the staff aide: "Get me a hamburger and make sure it is juicy!" In Dialog (10), the 7-year-old was expressing a request to his mother, hoping that his mother might comply. In Dialog (11) the same statement was interpreted as nothing more than wishful thinking. The first inmate made no order or request to the second inmate. Hence, the first utterance of each dialog fragment implies a different conversational goal depending upon the differences in the social relations of the conversational participants. The social context and the relationship between the two speakers generate expectations that guide the course of the conversation. A staff aide expects to be ordered about by a general. A mother expects her son to ask her for favors. Prison inmates cannot expect each other to do things that are made impossible by their incarceration. These expectations lead to a formulation of different conversational goals for the utterance, "I want a juicy hamburger " in each conversational fragment. The conversational principle exemplified in our discussion is summarized as Conversational Rules (7) and (8):

RULE 7. The social relationship between the participants in a conversation generates expectations about the intentional meaning of utterances in the conversation. These expectations are used to determine the conversational goals of each participant.

RULE 8: Each speaker s perception of the conversational goals of the other speaker determines his interpretation of the other speaker s utterances.

Differences in understanding of conversational goals lead to different responses in a dialog, as illustrated in Conversation Fragments (9), (10) and (11). We saw how a social relationship between two people can influence their interpretation of each other s conversational goals. Two strangers can also make assumptions about each other s conversational goals based on appearances, social circumstances and each other s occupation. Consider, for instance, the various responses to John s question in the example below:

Scenario: John walked up to a person in the corner and asked: "Do you know how to get to Elm Street?"

12.1) The stranger replied: "You go two blocks toward that tall building and turn right."

12.2) The cab driver in the corner replied: "Sure, Hop in. Where on Elm do you want to go?"

12.3) The person, who was holding up a map and a piece of paper with an Elm Street address on it, replied: "No, could you tell me how to get there?"

12.4) The child answered: "Yes, I know how to get there!"

The question was interpreted to mean four different things, depending on whom John spoke to. If a stranger asks, "Do you know how to get to X," the listener usually interprets this to mean "I want to go to X, but I do not know how to get there. Please give me directions." Since the occupation of a cab driver is to take people to their destination it is perfectly legitimate for him to interpret the question as: "If you know how to get to X please take me there." The person who is visibly lost and trying to find his way may interpret John's question as: "You seem to be lost. Can I help you find your way?" Response (12.3) illustrates that the responder did not infer that John s goal was to go to Elm street, in contrast with the two previous responses. A child often interprets questions of the form: "Do you know Y" literally, possibly inferring that the person asking the question is quizzing him. As in our previous examples, the differences in interpretation can be explained in terms of differences in the perceived goals of the participants in the conversation.

II) MICS: A process model of human conversation.

The phenomenon of human conversation is too complex for any single study to do justice to more than a narrow aspect of the problem. In order to fully understand human conversations we may have to understand all human cognitive reasoning processes. Our research approach can be outlined as follows: 1) Study many sample conversations; 2) try to establish some relatively general rules of conversation; 3) encode these rules into a process model; 4) see if this model accounts for certain aspects of human conversation; 5) realize that we solved hardly more than a minute part of the problem, and 6) reiterate the research process in a (hopefully positive) feed-back loop.

The conversational rules discussed in the first section address problems that need to be considered if one is to understand human

conversations. There is little doubt, as demonstrated by countless examples, that conversational goals, shared knowledge between speakers, social relationships between speakers, and the conversational import of each utterance in a dialog are aspects of human discourse that need to be analyzed if one is to understand how human conversations work. Analyzing these aspects, however, solves only a small subset of the larger problem of how conversations function. For instance, the problem of topic selection in a conversation needs to be addressed. How does a person change the topic in a conversation? How are new topics chosen? These questions are analyzed in Schank [1977]. Here we propose some additional ideas on the impact of shared knowledge and interests on topic-selection.

MICS (Mixed-Initiative Conversational System) is a fully implemented computer program that generates one side of a natural-language conversation. MICS embodies the conversational rules discussed in this paper, a topic transition mechanism based on Schank [1977], and the idea of a conversational syntax. Conversational syntax is a set of rules that help to characterize well-formed conversations. For instance, the following four rules are among the twenty conversational syntax rules in MICS:

RULE 9: Do not repeat information in the course of a conversation unless explicitly requested.

RULE 10: Do not say things that the other speaker already knows about.

RULE 11: If the other speaker says something that violates one s beliefs, then respond in kind by re asserting the violated belief.

RULE 12: If insulted, return the insult or end the conversation.

MICS illustrates that the conversational rules, topic transition rules and conversational syntax rules provide a sound if possibly incomplete approach to the modeling of human conversation. MICS is able to generate reasonable conversations about domains on which it has some knowledge, but there are still many short-comings, as illustrated in the examples that follow.

MICS "knows" about itself as a person (e.g. it believes itself to be a computer programmer) and uses a model of the other conversational participant that contains certain information about the speaker, such as his social status, his occupation, his relationship with other people and his areas of special interest and more detailed knowledge. The model of the other speaker is a necessary part of the conversational model, since it enables MICS to apply the conversational rules. When MICS talks to a stranger, it builds a model of the speaker; in fact, one of its conversational goals is to learn about the person with whom it is conversing. In the dialog that follows (an actual computer run), MICS starts out knowing nothing about Dave.

MICS CONVERSATION LOG ******** 10/11/77 *******

HELLO, WHAT IS YOUR NAME?

I am Dave.
WHAT'S NEW TODAY, DAVE?

Nothing, What's your name?
MY NAME IS MICS.

What is your job?
I AM A COMPUTER PROGRAMMER

That's nice.
WHAT DO YOU DO FOR A LIVING?

| MICS takes the initiative by asserting
| its conversational goal of learning
| about Dave.

I am still a student.
DO YOU LIKE BEING A STUDENT?

For the time being yes.
WHAT WILL YOU DO AFTERWARDS?

I want to be a professor some day.
WHAT DO YOU DO FOR FUN?

| MICS formed a person-frame for Dave. At this
| point, the conversational goal of trying to
| establish an area of mutual interest and
| shared knowledge is activated.

I smoke dope.
YOU MIGHT GET STONED IF YOU SMOKE TOO MUCH.

I know, that's the idea.
YOU MIGHT GET IN TROUBLE IF THE COPS CATCH YOU.

All right, stop moralizing, what do you do for fun?
I LIKE TO PLAY CHESS

That sounds boring.
NO, IT IS A LOT OF FUN!

My brother Eric plays chess.
HE MUST BE SMART.

I have to go. Bye Mics.
NICE TALKING TO YOU, DAVE

As a result of the conversation, MICS knows something about Dave, in particular that one of his life goals is to becomie professor. MICS's conversational goals cause it to establish its own conversational topics. When a new topic is introduced by the human speaker, MICS stays on the chosen topic or chooses a closely related topic. Another rule of conversation is invoked in the above dialog:

RULE 13: Each participant in the conversation has the conversational goal of saying things that are important to the other participant.

When the human participant says that he likes to smoke marijuana, MICS finds in its memory that smoking marijuana is illegal and can lead to the smoker getting stoned. According to MICS's

beliefs getting stoned is undesirable, and doing something illegal leads to undesirable consequences. MICS determines that it is important for Dave to know about the bad consequences of his acts. Therefore, applying Rule (13), MICS warns Dave about the evils of marijuana. It is evident that MICS should have known that Dave knows that smoking marijuana leads to getting stoned, and that, in fact, getting stoned is Dave s primary purpose for smoking marijuana. If MICS knew this fact then Conversational Syntax Rule (10) would prevent MICS from warning Dave about getting stoned. Therefore, MICS needs a better mechanism for deciding what knowledge it shares with the other speaker. It is very difficult to determine what the other person already knows and what he is able to infer, but this is a problem that cannot be circumvented if we are to model human discourse.

The conversational syntax rule of countering a belief violation (Rule 11) caused MICS to object when Dave said that playing chess (MICS s favorite pastime) was not exciting. Since he two participants could not find a common area of interest, the conversation ended without any substantitive dialog. This illustrates Conversational Rule (4), discussed in the previous section. Let us analyze a conversation between MICS and Pete, someone who MICS believes to be its friend.

HELLO, WHAT IS YOUR NAME?

Pete.
WHAT'S NEW TODAY, PETE?

Professor Stevenson is giving us an exam friday.
STEVENSON GIVES DIFFICULT EXAMS.

Yeah, do you know that Dave is probably going to flunk?
THAT'S GOING TO MAKE IT HARD FOR DAVE TO BECOME A PROFESSOR.

Too bad, Dave is not too smart anyways.
YES, DAVE LIKES TO SMOKE DOPE.

I've got to go and study.
YOU WILL DO WELL IN THE EXAM.

The above conversation illustrates the usefulness of shared knowledge between the two participants. Pete and MICS both know about taking exams, and they have a common acquaintance - Dave. MICS is able to use what it learned about Dave in its previous conversation. MICS s response to Dave possibly flunking the exam illustrates Conversational Rules (3), (5) and (7). The continuity of the conversation is preserved. The chosen topic transition is probably of interest to Pete, as inferred by MICS in the following manner: Pete introduced Dave into the conversation; therefore, he probably wants to talk about Dave. Not fulfilling a life goal (e.g. becoming a professor) is an inherently important consequence to Dave; therefore, Pete is probably interested in knowing this consequence. Rule applies because in the social relationship between Pete and MICS (they are friends) it is OK to talk about mutual acquaintances. (It is NOT necessarily OK for a judge and a defendant, for instance, to discuss mutual acquaintances.)

MICS's utterance about it being difficult for Dave to become a professor illustrates the inferenceing that proceeds in the course of a conversation. The inferences are directed at the most serious consequences of the preceeding utterance: flunking the exam. Focusing the inference process by importance criteria was found to be necessary in order to keep conversations from digressing into uninteresting topics.

When Pete states that Dave is not too smart, MICS reacts by verifying this evaluation. Since in the previous conversation Dave chose to ignore MICS's warnings about smoking marijuana, MICS concluded that Dave was not too bright. This conclusion is used as supporting evidence to back Pete's evaluation. The process of responding to opinions is captured in conversational rule 14:

RULE 14: If a speaker in a conversation states an opinion, the other speaker should consider whether or not he agrees with the opinion. The first speaker usually has the conversational goal of wanting feedback (usually agreement) on his opinion.

The last exchange in the above dialog illustrates a deficiency in MICS determination of the conversational goals. When Pete says, "I've got to go and study," he means, "I want to end the conversation," but inferring this conversational goal requires understanding what it means to study. One cannot infer from the phrase, "I've got to go ...", that the conversation should terminate. For instance, if Pete says, "I've got to go to Europe," he probably wants to introduce his upcoming trip as a new topic of conversation. The knowledge that studying and talking cannot proceed simultaneously is necessary in order to conclude that Pete wanted to end the conversation.

The primary usefulness of the MICS system is that it sheds light on problems about the rules of conversation that might otherwise remain unnoticed. The feedback from testing the implementation of our conversational model forces us to be more precise and complete in our formulation of our ideas about conversation. The following is a conversation fragment from a later version of MICS which has a more sophisticated inference capability for determining the knowledge that it shares with the other speaker. For instance, MICS infers that if a person voluntarily chooses to pursue a certain activity, then that person is aware of the intended consequences of this activity.

Ashtrays.
WHAT ABOUT ASHTRAYS?

Ashtrays are useful.
DO YOU SMOKE?

Yes.
YOU MIGHT GET CANCER IF YOU SMOKE TOO MUCH.

Who cares, anyway I smoke dope, not tobacco.
I SEE, YOU LIKE TO GET HIGH.

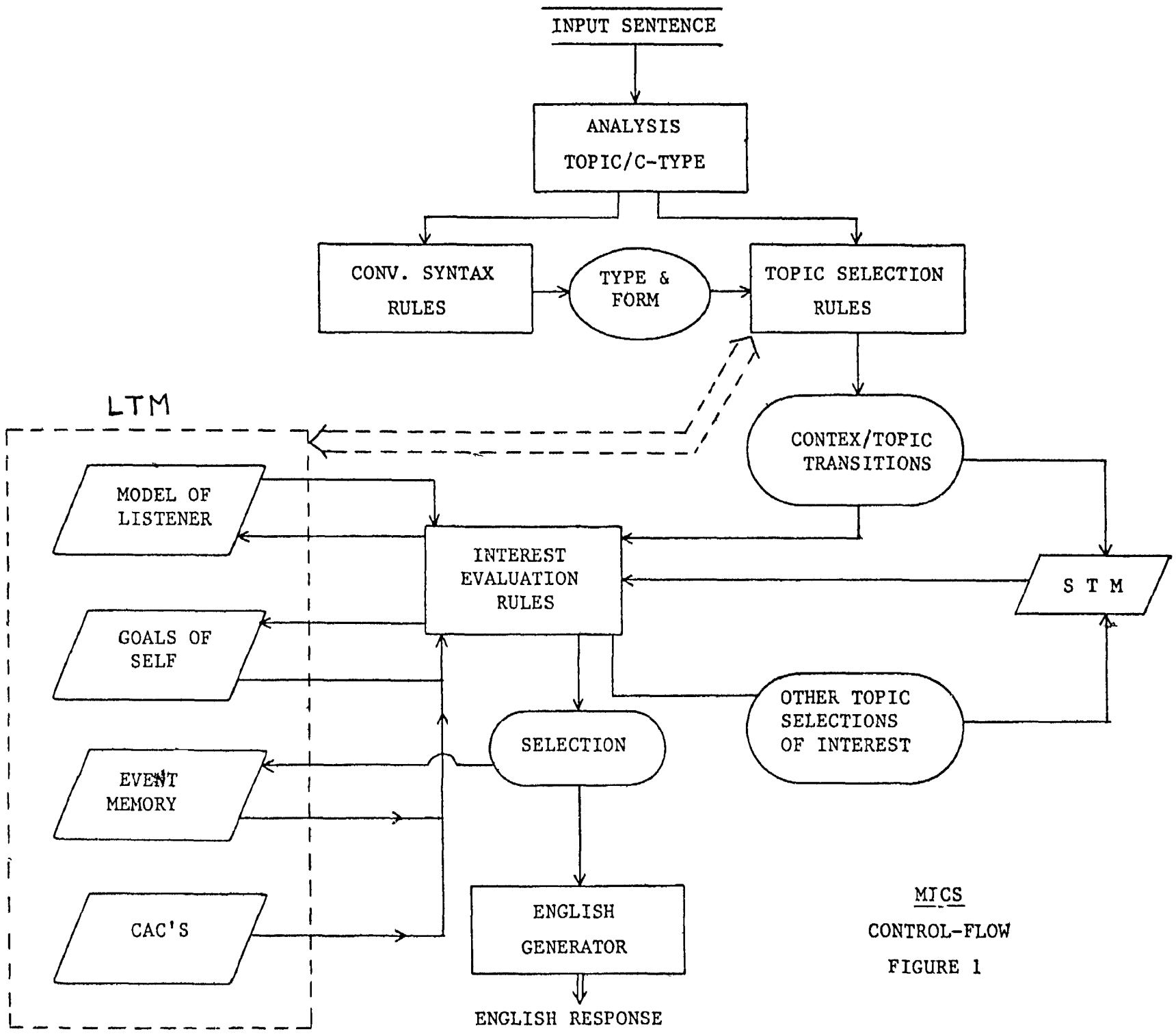
It is better than getting drunk.
NOT NECESSARILY; YOU MIGHT GET CAUGHT BY THE COPS.

Figure (1) is a diagram of the control flow of MICS. As a process model of conversation, MICS is a useful tool, but it is far from a complete or even an entirely correct model of human conversation. MICS is a 3 level production system. The first set of productions is the conversational-syntax rules. These rules formulate contraints on the nature of the utterance that MICS will generate next in order to insure that the conversation follows a proper, "natural" course. Some of these constrains take the form of conversational goals used by the other two sets of productions rules. The second set of rules is the topic selection productions. These productions generate topic transitions guided by conversational goals and the amount of shared knowledge between the speakers. Several alternative things to say may be generated in this phase of the processing. These, as well as the conversational goals are stored in a short term memory (STM) and are used by the third and theoretically most significant phase of the program: the topic evaluation rules.

The third set of productions encodes the conversational rules discussed throughout this paper. These rules use the STM information, the memory models of the self and the other speaker, and an inferencer when necessary. The purpose of these rules is to choose the most interesting topics (to both the self and the other speaker) from the alternatives generated by the second set of production rules. The inferencer is used to determine what the other speaker is likely to know and which aspect of the topic he would be most interested in discussing. Thus, the menaing of an utterance is produced by the third set of production rules.

The utterances are generated in English by a rather crude phrasal English generator. The utterances from the other speaker are analyzed for their meaning and conversational form by a primitive, key-concept oriented analyzer.

Disclaimer: MICS is a first-pass process model of a theory of conversation, not a theory of learning about other people. As such, its ability to learn about the other conversational participants is not as general as the dialogs presented in this paper may suggest. MICS learns about the other speaker by instantiating a prototypical-person frame - a data structure that encodes the more generally applicable facts about people and their social relations.



MICS
CONTROL-FLOW
FIGURE 1

Conclusion.

We believe that the best way to analyze a problem as difficult as modeling human discourse is to forge ahead by creating rules that capture important aspects of the conversation problem The usefulness of these rules should be tested in a reactive environment such as an interactive computer program. Since conversation is not a problem that can be isolated from other aspects of human cognitive behavior, we are researching it in conjunction with other aspects of Artificial Intelligence. A process-based theory of human conversation should give some insight into other Natural Language Processing issues in particular, and AI modeling of human reasoning processes in general.

References.

Carbonell, J. G. 1978. Computer Models of Social and Political reasoning, PhD. Thesis, Yale University, New Haven, Conn.

Grosz, B. J. 1977. The Representation and Use of Focus in a System for Understanding Dialogs, Proc. of the fifth IJCAI., MIT, Cambridge. Mass.

Lehnert, W. 1977. The Process of Question Answering, PhD Thesis. Tech. Report 88, Yale University, New Haven, Conn.

Mann W. Moore J., Levin J. 1977. A Comprehension Model for Human Dialogue, Proc. of the fifth IJCAI. MIT, Cambridge, Mass.

Schank, R. C. and Abelson R. P. 1977. Scripts, Goals, Plans and Understanding, Lawrence Lawernce Erlbaum. Hillside, NJ.

Schank, R. C. 1977. Rules and topics in conversation, Cognitive Science, Vol. I, No. 4.

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 & 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 scene, 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 [1976].) 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 <u>salience</u> 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 <u>organization of gestalts</u>, 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 <u>temporal sequence</u> (elementary source of cause-effect relationships and "scripts"), <u>constant copresence</u> (elementary source of notions of identity or inherent connectedness), and eventually <u>semantic relatedness</u> (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 <u>projection</u>, 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 <u>reconstruction</u> 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 <u>ball</u> 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).

<u>Properties</u> 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 Philadelphia 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 <u>not</u> 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 <u>position</u> 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 <u>possession</u> 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 <u>really</u> 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.

<u>The role of aesthetics</u>

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, ett.), in addition to evaluating whether our hypotheses are helping us get what we want.

<u>Development of a taxonomy of experience</u>

Let us assume that we start with a unitary concept of the world, and examine a plausible development of distinctions in the <u>visual</u> world.* The first sort of dist nction 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, . 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.

The Problem of Naming Shapes: Vision-Language Interface

by

R. Bajcsy*
and
A.K. Joshi*

Computer and Information Science Department
University of Pennsylvania
Philadelphia, PA 19104

1. Introduction

In this paper, we will pose more questions than present solutions. We want to raise some questions in the context of the representation of shapes of 3-D objects. One way to get a handle on this problem is to investigate whether labels of shapes and their acquisition reveals any structure of attributes or components of shapes that might be used for representation purposes. Another aspect of the puzzle of representation is the question whether the information is to be stored in analog or propositional form, and at what level this transformation from analog to propositional form takes place.

In general, shape of a 3-D compact object has two aspects: the surface aspect, and the volume aspect. The surface aspect includes properties like concavity, convexity, planarity of surfaces, edges, and corners. The volume aspect distinguishes objects with holes from those without (topological properties), and describes objects with respect to their symmetry planes and axes, relative proportions, etc.

We will discuss some questions pertinent to representation of a shape of a 3-D compact object without holes, for example: Is the surface aspect more important than the volume aspect? Are there any shape primitives? In what form are shape attributes stored?, etc. We shall extensively draw from psychological and psycholinguistic literature, as well as from the recent AI activities in this area.

2. Surface and Volume

In this section, we will investigate the relationship between the surface aspect and the volume aspect from the developmental point of view and from the needs of a recognition process. By doing so, we hope to learn about the representation of shapes. Later, we will examine the naming process for shapes and its relation to representation.

This work has been supported under NSF Grant #MCS76-19465 and NSF Grant #MCS76-19466

There is evidence that a silhouette of an object, that is its boundary with respect to the background, is the determining factor for the recognition of the object (Rock 1975, Zusne 1970). If we accept the above hypotheses then the fact that the silhouette is a projected outline of the 3-D object implies that the recognition of the 3-D object at first is reduced to the recognition of a 2-D outline. This is not entirely true, however, as Gibson (Gibson 1950) has argued. According to Gibson's theory, the primitives of form perception are gradients of various variables as opposed to the absolute values of these variables. From this follows the emphasis on perceiving the surface first and the perception of the outline only falls out as a consequence of discontinuities of the surface with respect to the background.

We are pursuaded by Gibson's argument and regard the recognition process as starting with surface properties; Miller and Johnson-Laird (Miller & Johnson-Laird 1976) have suggested some surface predicates as possible primitives, such as convex, concave, planar, edge, and corner. The 2-D outline is furthermore analyzed as a whole according to the Gestalist and some salient features (Pragantz) are detected faster and more frequently than others (Koffka 1935, Goldmeir 1972, Rosh 1973); such pragmatic features are for example, rectangularity, symmetry, regularity, parallelness, and rectilinearity.

Piaget also argues (Paiget, Inhelder 1956) from the developmental point of view that children first learn to recognize surfaces and their outlines, and only later, after an ability to compose multiple views of the same object has been developed, they can form a concept of its volume.

Volume representation becomes essential as soon as there is motion of the object or of the observer. Note that the salient features of 2-D shapes are invariant under transformations such as rotation, translation, expansion and shrinking. Features with a similar property must be found in the 3-D space for the volume representation. We feel that the most important feature is symmetry. Clark's work seem to support this (Clark 1975); he shows that in language space as in the perceptual space we have 3 primary planes of reference: ground level; vertical: left-right; vertical: front-back. While the ground level is not a symmetry plane, the two vertical ones are symmetry

planes. The fact that the ground level is not a symmetry plane is supported by the experiments of Rock (Rock 1973), who has shown that some familiar shapes are hard to recognize with 180° rotation with respect to the ground level. After a careful examination of the relevant literature to date, we find that there is a claim that we can recognize shapes via some features which are more salient than others. But does it follow from this that shape is an independent attribute like color, or is it a derived concept from other features?

In an effort to answer this question, we set out to examine labels of shapes in the hope that if there are any shape primitives (other than angles, edges, parallelness, and the like) then they may show up in labels describing more complex shapes. One immediate observation we can make is that there are very few names which only describe a shape, such as triangle or rectangle. More commenly, label of a shape is derived from the label objects which have such a typical shape, for example, letter-like shapes (V, L, X), cross-like shape, pear-like shape, heart-like shape, etc. A special category of labels are well defined geometric objects, such as circle, ellipse, sphere, torus, etc. The question is whether we store for every shape a template or whether there are any common primitives from which we can describe different shapes.

In addition to the 2-D features mentioned earlier, primarily 2-D features, we do use 3-D shape descriptions (primitives) such as: round, having 3 symmetry planes and all the symmetry axes approximately of the same length, elongated, where the size in one dimension is much longer than the two remaining, thin, where the size of one dimension is much smaller than the other, etc. Note that many of these descriptions are vague, though often there more accurate shape labels available; for example, cone stands for an elongated object with two symmetry planes, a circular crossection, and sides tapering evenly up to a point, called appex (Webster's dicitionary).

We believe that there are some descriptions of shapes which are more primitive than others; for example, round, elongated, thin, flat, circular, planar, etc., as opposed to heart-like, star-like, and so on. As pointed out earlier, these latter descriptors are derived from the names of previously recognized objects. When we use these descriptions during a recognition process we do not necessarily match exactly all features of the template shape to the recognized shape, but rather we depict some characteristic properties we associate with the given label, and only these are matched during the recognition process. In this sense, we approximate the real data to our model and primitives. The labels which encompass a more complex structure of these properties (like cone, heart, star, etc.) when they are used in describing other shapes, are used as economical shorthand expressions for the complexity that these shapes represent. (This appears to be related to the codability notion of Chafe (Chafe 1975)).

3. Analog and Propositional Representation

In this section, we will discuss certain issues concerning the form of the stored information, necessary not only for recognition purposes (matching the perceived data with a stored model) but also for recall, and introspection of images.

There are two questions:

1. At which level the analog information is converted to propositional (verbal or non-verbal) and after this conversion, is the analog information retained?
2. How much of the propositional information is procedural and how much structural?

For simplicity, we will regard analog information in our context as picture points, or retina points. Any further labeling, of a point or of a cluster of points, such as an edge, line, region, etc. leads to derived entities by one criterion or another and therefore may be regarded as propositional.*

At this point, it is appropriate to point out that any such unit as an edge, line or region can be described in at least two different ways; one is structural or organizational, and the other is parametric or dimensional. Structural information refers to the organization of perceptual elements into groups. Figure-ground and part-whole relationships are paradigm examples of structural information. Parametric information refers to the continuous values of the stimulus along various perceivable dimensions. Color, size, position, orientation, and symmetry, are some examples of parametric information.

We are not advocating that these two types of information are independent (cf. Palmer 1975). It is, for example, a well known experience that by changing drastically one dimension (one parameter) of an object (say a box), one can change the structure of the object (in this case, it becomes a wall-like object). However, we do wish to keep the distinction between structural and parametric information. The importance of this distinction is that while structural information is inherently discrete and propositional, parametric information, is both holistic (integral) and atomic (separable). The fact that parametric information is separable is quite obvious if we just recognize that different parameters represent clearly distinguishable different aspects of the visual information. For example, color, size, position, etc. On the other hand all these parameters are represented holistically in an image, and can be separated only by feature (parameter) extraction procedures (Palmer 1975).

Parametric information is separable, however, the question is whether each parameter-feature

---

* The distinction is not really as sharp as stated here. One way to make the distinction is to look at the closeness with which a transformation of a representation parallels the transformation of the object represented. The closer it is the more analog the representation is.

has continuous or discrete values. Continuous values would imply some retainment of analog information (Kossylyn 1977), while discrete values would not. Opponents of the discrete value representation argue that a) the number of primitives needed would be astronomical, and b) the number of potential relationships between primitives would be also very large (Fishler 1977). This is further supported by experiments on recall of mental images (Kosslyn, Shwartz 1977) where these images appear in continuous-analog fashion. Another similar argument in favor of analog representation is the experiment of comparing objects with respect to some of their parameters, like size, or experiments on mental rotation (Shepard, Metzler 1971).

Pylyshyn (Pylyshyn 1977) cautiously argues against the analog representation for the same object viewed under different conditions as a result of the semantic interpretation function (SIF). The SIF will extract only those invariances characteristic for the object in a given situation, and thus reduce the number of possible discrete values and their range for a given parameter. The invariances are determined by laws of physics and optics, and by the context, i.e., the object sizes will remain fixed as they move, the smaller objects will partially occlude the larger object, etc.

We would like to propose a discrete value representation for parametric information with an associated interpolation function (sampling is an inverse of interpolation) and a clustering procedure. During the recognition process, a clustering procedure is evoked in order to categorize a parameter while during an image recall an interpolation procedure is applied to generate the continuous data. Our model seems not to contradict Kosslyn's findings, that is we assume as he does, that the deep representation of an image consists of stored facts about the image in a propositional format. Facts include information about:

a) How and where a part is attached to the whole objects.
b) How to find a given part.
c) A name of the category that the object belongs to.
d) The range of the size of the object, which implies the resolution necessary to see the object or part.
e) The name of the file which contains visual features that the object is composed of (corners, edges, curvature descriptions of edge segments, their relationships etc.).

The only place where we differ from Kossyln's model is in the details of the perceptual memory. While his perceptual memory contains coordinates for every point, our perceptual memory has identified and stored clusters of these points, like corners, edges, lines, etc. From these features and the interpolation procedure, we create the continuous image. This is very much in the spirit of a constructive vision theory as proposed by Kosslyn and others. A similar argument can be used for preserving continuity in transformation of images, such as rotation (Shepard, Metzler 1971) and expansion (Kosslyn 1975, 1976). The contraction process is the inverse of expansion and therefore will envoke the sampling routine instead of the interpolation routine. The problem of too many discrete values and their relationships, as stated by Fishler, is taken care of by the fact that for each parameter there is an associated range with only a few categories such as small, medium, and large. As pointed out by Pylyshyn, it is the range of parameters which is context dependent and thus differs from situation to situation. This view also offers some explanation that often incomplete figures are perceived as whole.

We also want to postulate that analog information, as we specified it, is not retained, and if there are ambiguities due to the inadequacy of the input data, a new set of data is inputed. This is supported by several psychological experiments, for example, by asking people to recognize a building where they work from accurate drawings and sloppy pictures (Norman 1975). The overwhelming evidence is that people prefer a sloppy picture to the more accurate one, for recognizing their own building. Even the experiment of Averbach and Sperling (Averbach and Sperling 1968) concerning the visual short memory after 1/20 sec exposure to letters does not contradict our hypothesis that we maintain in this case, edges rather than picture points, although it allows the other interpretation as well.

We now turn to the second question. Sihce propositional information can be represented by an equivalent procedure (giving a true or a false value), the question of propositional information vs structural information can be replaced by the question: What are the necessary procedures that have to be performed during a recognition process and what type of data they require? Clearly, the parametric information is derived procedurally There are well defined procedures for finding color, size, orientation, etc. The part-whole relationship as well as the instance relationship clearly have to be structurally represented (Miller and Johnson-Laird 1977).

While the structural information is derived from symbolic propositional data and the transformations performed are, for example, reductions, and expansions, the parametric information is derived from the perceptual data and the transformations performed are more like measurements, detections, and geometric transformations.

In the context of 3-D shape representation we believe in a combination of procedural - parametric and propositional nodes organized in a structure. Take an example of representing a shape of a human. We have the part-whole relationship: head, neck, torso, arms, legs, etc. Head has parts: eye, nose, mouth, etc. These concepts are propositional - symbolic. From the shape point of view, however, head is round, neck is short and wide elongated blob, the arms and legs are elongated and the torso is elongated but wide. Although these labels correspong to 2-D as well as 3-D shape, there is a mechanism: projection transformation which transforms elongated 3-D into elongated 2-D shape. In any case, round,

elongated, wide, short, are procedures - tests whether an object is round, elongated, etc. We know that round (circle) in 2-D corresponds to sphere in 3-D, elongated (rectangle, or ellipse) to a polyhedra or cylinder, or ellipsoid.

When we view only one view of a scene or a photograph, we analyse the 2-D outline. However, when we have more than one view at our disposal or when we are asked to make 3-D interpretation then we reach from the 2-D information to corresponding 3-D representation. This is the time when volume primitives like sphere, cylinder, and their like come into play. These primitives do not seem to be explicit (we do not say a shape of a man is a sphere attached to several cylinders) in the representation. Rather what is in the shape representation are the feature primitives, (like the symmetry planes, the ratio of symmetry axis) attached to other pointers, which point also, if appropriate, to labels like sphere, cylinder, flat object, polyhedron, etc. These labels are in turn used for shortening a complex description.

An implementation of a 3-D shape decomposition and labelling system is under development (Bajcsy, Soroka 1977). Earlier we have experimented with a partially ordered structure as means to represent 2-D shape (Tidhar 1974, Bajcsy, Tidhar 1977) in recognition of outdoor landscapes (Sloan 1977) and in the context of natural language understanding (Joshi and Rosenschein (1975), Rosenschein (75)).

Note that not always are we able to describe a shape as a composition of some volume primitives like sphere, cylinder, or a flat object. As an example in the case is a shape of a heart. A heart has 2 symmetry planes and it is roughly round, but its typical features are the two corners centered, one, concave and the other convex connected by a convex smooth surface. Here clearly, any attempt to describe this shape, by two ellipsoids or some other 'primitive' is artificial. Thus, the representation will have only feature primitives but no volume primitives.

Of course, there are cases that fall between. As an example, consider a kidney shape where one can say it is an ellipsoid with a concavity on one side.

What are the implications from all of this?

1. We do not measure or extract spheres, cylinders and their like as primitives, but rather we measure convexity, concavity, planar, corners, symmetry planes, which are primitive features.
2. These features form different structures to which are attached different but in general, not independent labels.
3. While these structures represent explicit conceptual relationships, the nodes are either labels or procedures with discrete values denoting, in general, N-ary relations.

4. Conclusions

In this paper, we have considered the following problems:

1. How much of analog information is retained during recognition process and at which level the transformation from analog to propositional takes place?
2. How much of the information stored is procedural (implicit) and structural (explicit) form?
3. What are the primitives for two dimensional and three dimensional shapes?
4. How is the labelling of shapes effected by the way the shapes are represented? By studying the shape labels can we hope to learn something about the internal representation of shapes?

Clearly, these four questions are intimately related to the general problem: representation of three dimensional objects.

We are led to the following conclusions. Our conclusions are derived primarily on the basis of our experience in constructing 2-D and 3-D recognition systems and the study of the relevant psycological and psycholinguistic literature.

1. Analog information is not retained even in a short term memory.
2. Our experience and the analysis of the relevant literature leads us to be in favor of the constructuve vision theory. The visual information is represented as structures, with nodes which are either unary or n-ary predicates. The structures denote conceptual relationships such as part-whole, class inclusion, cause-effect, etc.
3. The shape primitives are on the level of primitive features rather than primitive shapes. By primitive features we mean, corners, convex, concave and planar surfaces and their like.
4. The labels of shapes, except in a few special cases, do not describe any shape properties and are derived from objects associated with that shape.
5. In order to preserve continuity, we need interpolation procedures. We assume that several such procedures exist, for example, clustering mechanisms, sampling procedures, perspective transformations, rotation, etc. These are available as a general mechanisms for image processing.

We certainly have not offered complete solutions to all the issues discussed above, but we hope that we have raised several valid questions and suggested some approaches.

References

1. Averbach, E., and Sperling, G . Short-Term Storage of Information in Vision in: Contemporary Theory and Research in Visual Perception, (ed.) R.N. Haber, NY, Holt, Rinehart and Winston, Inc. 1968

2. Bajcsy, R., and Soroka, B.: Steps towards the Representation of Complex Three-Dimensional Objects, Proceedings on Int. Artificial Intelligence Conference, Boston, August 1977.

3. Bajcsy, R., and Tidhar, A.: Using a Structured World Model in Flexible Recognition of Two Dimensional Patterns, Pattern Recognition Vol. 9, pp. 1-10, 1977.

4. Clark, E.V.: What's in a Word? On the Child's Acquisition of Semantics in His First Language, in: Cognitive Development and the Acquisition of Language, (ed.) T.E. Moore, Academic Press, NY 1973, pp. 65-110.

5. Clark, H.L.: Space, Time Semantics, and the Child, in: Cognitive Development and the Acquisition of Language (ed.) T.E. Moore, Academic Press, NY 1973 pp. 27-63.

6. Chafe, W.L.: Creativity in Verbalization as Evidence for Analogic Knowledge, Proc. on Theoretical Issues in Natural Language Processing, Cambridge, June 1975 pp. 144-145.

7. Fishler, M.A. On the Representation of Natural Scenes, Advanced Papers for The Workshop on Computer Vision Systems, Univ. of Massachusetts, June 1977, Amherst.

8. Gibson, J.J.: The Perception of the Visual World, Boston, MA, Houghton, 1950.

9. Goldmeir, E.: Similarity in Visually Perceived Forms, Psychological Issues 8, 1972, No. 1 pp. 1-135.

10. Koffka, K . Principles of Gestalt Psychology, New York, Harcourt, Brace 1935.

11. Kosslyn, S.M.: Information Representation in Visual Images, Cognitive Psychology 7, pp. 341-370, 1975.

12. Kosslyn, S.M.. Can Imagery Be Distinguished from Other Forms of Internal Representation? Evidence from Studies of Information Retrieval Times, Memory & Cognition Vol. 4, 1976, No. 3, pp. 291-297.

13. Kosslyn, S.M., and Shwartz, S.P.: Visual Images as Spatial Representations in Active Memory, in: Machine Visions, (eds.) E.M. Riseman & A.R. Hanson, NY Academic Press (in press) 1978.

14. Miller, A , and Johnson-Laird, P.N.: Language and Perception, Harvard Univ. Press, Cambridge, MA 1976.

15. Norman, D.A., and Bobrow, D.G.: On the Role of Active Memory Processes in Perception and Cognition, in: C.N. Cofer (ed.) The Structure of Human Memory, San Francisco, W.H. Freeman, 1975.

16. Palmer, S.E.: 'The Nature of Perceptual Representation: An examination of the Analog Propositional Contraversy, Proc. on Theoretical Issues in Natural Language Processing, Cambridge, June 1975 pp. 151-159.

17. Piaget, J., and Inhelder, B : The Child's Conception of Space, New York: Humanities Press, 1956.

18. Pylshyn, Z.W.: Representation of Knowledge: Non-Linguistic Forms, Proc. on Theoretical Issues in Natural Language Processing, Cambridge, June 1975 pp. 160-163.

19. Rock, I.: Orientation and Form, Academic Press, Inc. Ny 1973.

20. Rock, I.: An Introduction to Perception, MacMillan Publ. Co., NY 1975.

21. Rosh, E.H.: On the Internal Structure of Perceptual and Semantic Categories, in: Cognitive Development and the Acquisition of Language. (ed.) T.E. Moore, Academic Press, NY 1973, pp. 111-144.

22. Shepard, R.N., and Metzler, J.: Mental Rotation of Three-Dimensional Objects, Science, 171, 1971, pp. 701-703.

23. Tidhar, A.: Using a Structured World Model in Flexible Recognition of Two Dimensional Pattern, Moore School Tech. Report No. 75-02, Univ. of Pennsylvania, Philadelphia, 1974.

24. Zusne, L.: Visual Perception of Form, Academic Press, 1970, NY and London.

25. Sloan, K.: World Model Driven Recognition of Natural Scenes, Ph.D. Dissertation, Computer Science Department, University of Pennsylvania, Philadelphia, June 1977.

26. Joshi, A.K., and Rosenschein, S.J., "A Formalism for Relating Lexical and Pragmatic Information: Its Relevance to Recognition and Generation", Proc. of TINLAP Workshop, Cambridge 1975.

27. Rosenschein, S.J., "Structuring a Pattern Space, with Applications to Lexical Information and Event Interpretation", Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA 1975.

An Argument about the Composition of Conceptual Structure

Ray Jackendoff

Brandeis University

In order for people to be able to talk about what they perceive, there must be a level of mental representation at which information conveyed by language is compatible with information from sensory systems such as vision, nonverbal audition, touch, and so forth. I will call this level conceptual structure. Though the existence of conceptual structure has been more or less taken for granted (especially by the AI community), the need to consider it seriously has been brought to the attention of linguists rather recently, by such works as Fodor (1975) and Miller and Johnson-Laird (1976) This paper will present a combination of linguistic and visual evidence which bears on the nature of conceptual structure.

1. General properties of a theory of conceptual structure

A linguist's questions about conceptual structure can be separated into two major issues. The first, which the linguist shares with many branches of psychology, concerns the form of conceptual structure itself; the second, particular to linguistics, concerns the mapping between conceptual structure and syntactic structure. An answer to the first question, within the theoretical paradigm I will assume, will consist of a set of well-formedness rules for conceptual structure. The second will be answered by a set of correspondence rules which relate some subset of coneptual structures (the verbally expressible concepts) into some subset of syntactic structures (the meaningful sentences).

It seems reasonable to assume for a first approximation that the well-formedness rules for conceptual structure are universal and innate, i.e. that everyone is born with the capacity to develop the same concepts. However, the actual concepts that one does develop will depend to some extent on experience--including possibly linguistic experience, so there is room for a certain amount of "Whorfian" variation if necessary.

On the other hand, this position is not consistent with what I gather is the strongest version of Piagetian developmental theory, which could be construed in the present framework as a claim that certain conceptual well-formedness rules must be learned. Rather, the development of the child's conceptual ability must be attributed to increasing richness and interconnection of concepts, or to growth either in the well-formedness rules or in computational capacity, over which the child and the environment have little or no control. (The kind of growth I have in mind here is akin to the growth of bones and muscles: the environment must furnish nourishment, but it hardly can be said to control the interesting aspects of structure. See Chomsky (1975) for discussion.)

In addition to the assumption of universality and innateness of the conceptual well-formedness rules, I will make three other assumptions about the theory of conceptual structure and its relation to language. First, a theory of conceptual structure must be observationally adequate: as the level linking language and other perceptual systems, it must at least be able to express all the conceptual distinctions made by natural language. In practice, this calls for an attempt to account for a lexically and grammatically significant fragment of the language without artificial assumptions about the semantics (such as restriction to a microworld).

Second, a theory of conceptual structure must provide some principled way for the meanings of the parts of a sentence to be combined into the meaning of the whole sentence. This requirement of compositionality may be taken more or less strongly, depending on whether or not one requires each syntactic constituent (or even each word) of a sentence to correspond to a well-formed concept.

Related to this second assumption is a third, that the correspondence rules relating syntax and conceptual structure be relatively simple. As motivation for this constraint, we observe that the language learner must relate syntactic form to understood meanings--in fact he must probably learn many aspects of syntactic form in part by figuring out from context what meaning is intended by other speakers. Since syntactic form varies from language to language, the correspondence rules must be at least partly learned. In order to be able to explain how the child manages to acquire language, we should strive for a theory in which at least the language-particular part of the correspondence rules is fairly straightforward.

A second argument for the simplicity of correspondence rules is more heuristic. Language is, after all, an information transmission system, conceptual structure being the information which language conveys. It would be perverse not to take as a working assumption that language is a relatively efficient and accurate encoding of the information it conveys--despite generations of philosophers who have assured us that language is impossibly unsystematic and vague. To give up this assumption is to refuse to look for principles in natural language semantics. Accepting it entails that all deviations from efficient encoding be rigorously justified; what appears to be a quirky relationship between syntax and conceptual structure may turn out to be merely a bad theory of conceptual structure. (See Goldsmith and Woisetschlaeger (1976), Jackendoff (1978) for arguments to this effect.)

2. The argument: Figure formation and pragmatic anaphora

The preliminaries complete, we turn to the main argument, which begins with a discussion of one aspect of visual perception before turning to linguistic matters. We then will draw consequences for conceptual structure, where visual and linguistic information interact.

One of the most important and well-studied phenomena of visual perception is the emergence of a <u>figure</u> against a <u>background</u>. Intuitively, the figure is what attention is directed to; coherence or "thingness" inheres in the figure. It is often reported that the figure seems to stand out from the ground or to be imbued with greater vividness than the ground. The study of the figure-ground opposition has been one of the major preoccupations of the school of gestalt psychology (see e.g. Köhler (1947), Koffka (1935)).

For a simple and hopefully illustrative example, consider the contents of this page, in particular the geometric configuration in Figure 1.

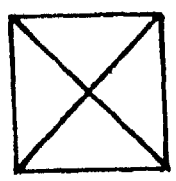


Figure 1

The whole of this configuration can form a visual figure seen against the background of the page. Parts of it can also emerge spontaneously as figures; probably the most prominent are a square and an X, each of which can seen against the rest of the page (including the rest of the configuration) as background. Among less natural figures, which emerge only with more deliberate effort from Figure 1, are such configurations as these (in order of decreasing salience, from left to right):

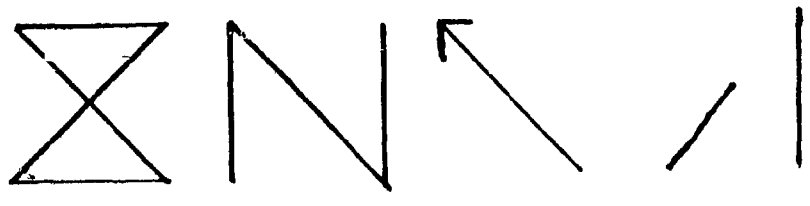


Figure 2

A number of important observations can be drawn from this simple example.

1. The number of possible figures that can be be perceived in a given configuration is very large, perhaps unlimited; however,
2. Only a small number of these are particularly salient.
3. Relative salience of perceived figures is a function of both features of the physical signal and properties of the visual system.
4. Features of the visual <u>context</u> can affect relative salience of figures. For example, the configurations in Figure 2 become much more likely to emerge from Figure 1 upon presentation of Figure 2; certain other possible figures (such as other arrow-shaped configurations) undoubtedly become somewhat more salient than before, while certain other possible figures not presented here remain as unlikely as before.
5. Features of the visual signal interacting with the viewer's <u>intention</u> can make certain figures more salient than they otherwise might be. For example, more figures will emerge from Figure 1 if the viewer is instructed to find all possible figures. Similarly, this aspect of figure formation is crucial in children's puzzles which ask the reader to find three rabbits and two bears hidden in the forest, or in the Hirschfeld cartoons in the New York Times in which the reader in on the joke is to find a stipulated number of instances of the configuration NINA. These figures would not emerge at all were it not for the reader's intention to find them.
6. Features of the visual signal interacting with the viewer's <u>knowledge</u> may make certain figures more salient. Someone who has worked with automobile engines will perceive more distinct figures upon looking under the hood than a mechanical novice; a botanists may see a number of distinct plants where a layman sees only a tangled confusion.

It is important to distinguish the conscious decomposition of a figure into parts from the process of figure formation itself. The former is available to awareness, and is in fact governed by principles of figure formation: as was seen in the example above, the perceived parts are themselves figures. On the other hand, the mental processes which bring about the emergence of a figure are themselves not open to awareness, and the aspects of the configuration which are relevant to the character of the perceived figure may have little to do with the intuitive (i.e. conscious) decompositon of the figure. To make this clearer, consider the not atypical example of facial recognition. Though one can recognize and distinguish thousands of faces, one cannot in most cases consciously decompse faces and say specifically what makes them recognizable and different from each other. In present terms, each recognized face forms a remembered figure, but many of the distinctive features of faces do not themselves form figures and hence are not available to conscious awareness (see Carey (1976), and also Helmholtz's (1885) brief but insightful remarks (p. 369)).

The observations we have made about the figure-

ground phenomenon have a bearing on the nature of conceptual structure. They suggest that there is a privileged set of conceptual structures which encode figures as unitary entities, and that this set is somehow related to conscious awareness. As far as consciousness is concerned, there is no representation of how figures are composed; though to other, unconscious, processes the composition of figures is accessible. Let us call this set of structures figural expressions.

We now relate figural expression in conceptual structure to language. Suppose someone points and simultaneously utters (1).

(1) I bought that yesterday.

What must the hearer do to fully understand the speaker? He must of course understand the words and the syntactic structure, and be able to use the correspondence rules involved in interpreting the sentence; but he also must interpret the word that. In this particular utterance, that is a case of what Hankamer and Sag (1976) call "pragmatic anaphora." In order to understand the intended referent of a pragmatically controlled pronoun like that in (1), the hearer must pick something out of his visual field, perhaps aided by the speaker's pointing gesture.

To make clearer the process of interpreting pragmatic anaphora, consider an example where no figure emerges which can correspond to a pragmatically controlled pronoun. Suppose speaker A utter (1) and points to a blurry photograph: "I bought that yesterday--isn't it gorgeous?" Speaker B, unable to make out anything in the picture, doesn't fully understand the utterance and responds "What are you talking about?" Suppose A then says, "That boat!" B peers at the picture and sure enough the figure of a boat emerges. He has a mino' aha-experience: "Oh, that! How could I miss it?" He now has received the message and discourse can continue.

Every reader has probably had an experience like this; its relevance in the present setting is as follows: in order for a pragmatically controlled pronoun to be understood, its intended referent must emerge as a figure in the mind of the hearer, that is, it must have a representation as a figural expression in conceptual structure. Thus we have established an important connection between the figure-ground phenomenon and pragmatic anaphora.

So far we have dealt only with figures that correspond to things (or their shapes). By and large this has been the kind of figure that has been investigated. But as Hankamer and Sag (1976) point out, there are many sorts of pragmatic anaphora.

(2) Here and there:
Your coat is here (pointing) and your hat is there (pointing).

(3) Do it:
(Hankamer attempts to stuff a 9" ball through a 6" hoop)
Sag: It's not clear you'll be able to do it.

(4) It happen:
That (pointing) better not happen again.

(5) Nominal identity-of-sense anaphora:
a. (Sag produces an apple)
Hankamer: Did you bring one for me
b. Those (pointing, e.g. to a (single) Cadillac) are expensive.

(6) Manner adverb:
You shuffle cards { thus / so / this way } (demonstrating)

(7) Measure phrase:
The fish that got away was { this / that / yay } (demonstrating) long.

There was more than that much (pointing) in the jar when I left

The same conditions hold on the comprehensibility of these sorts of pragmatic anaphora as on that in (1). For example, if the hearer is unable to see or figure out what goings-on the speaker is pointing at in (4), he will not fully understand the utterance (in the sense of having received all the information he is intended to receive).

Given that the existence of an appropriate figural expression in conceptual structure, supplied by the visual system, is necessary for the comprehension of pragmatic that-anaphora in (1), we must conclude similarly that a figural expression is necessary for all the sorts of pragmatic anaphora in (2)-(7) as well. But from the selectional restrictions involved in these constructions, we see that the figures involved cannot be things or shapes. Rather, each corresponds to a different sort of figure, distinct from things. Roughly, here and there correspond to places; do it to actions; it happen to events; nominal identity-of-sense anaphora to categories or kinds; manner adverbials to manners; and measure phrases to amounts. Each of these types of figures represents a different organization of the visual field than do figures corresponding to physical objects.

The existence of this variety of types of pragmatic anaphora suggests three points. First, the mind has the capacity to form figures of a number of distinct types on the basis of visual perception. Second, conceptual structure can represent such entities as places, actions, events, etc. as figural expressions, and this is why we can talk about them. Third, by the criterion of simplicity of correspondence rules, these entities are conceptually simple, since they correspond to something syntactically simple. More explicitly, that, a maximally simple NP, represents a minimally specified thing in (1), and the visual field is the source of the remaining information about the intended message. Similarly, the other expressions of pragmatic anaphora are maximally simple PPs, VPs, etc., and therefore should likewise correspond to minimally specified entities of the proper type; again, the remainder of the intended message is conveyed through the visual system.

3. The alternative to psychological and philosophical reductionism

One might object that all these different types of entities should be reduced by the theory of conceptual structure to concurrences of physical objects over time (a four-dimensional space-time map, for example), and that such entities as places and events should play only a derivative role in linguistic semantics. But such a view generally assumes that the psychological notion thing can be fairly simply correlated with physical objects via patterns of retinal stimulation; and this assumption is patently false. Most of the literature of perception is concerned with how we manage the remarkable feat of construing the world as full of more or less stable things, given constantly shifting patterns of retinal stimulation, and with how the things we see are or are not correlated with actual physical facts. It's not easy.

What seems to me a more productive approach is to abandon the goal of reduction and to claim that the types of entities referred to by the anaphoric expressions in (1)-(7) are all present as primitives of conceptual structure. Formally, this means that the well-formedness rules for conceptual structure must allow for figural expressions which correspond to each type. Furthermore, the well-formedness rules must provide an algebra of relationships among the types: a thing can be in a place, an event may have a certain number of things and places as constituents, some events consist of an action performed by a thing (the agent), and so forth. Under this approach, linguistic semantics is not concerned with reducing out events, places, and so forth, but with clarifying their psychological nature and with showing how they are expressed syntactically and lexically.

If any reduction is to take place, it will be in the theory of perception, which now must explain the relation of retinal (and auditory, etc.) stimuli to event- and place-perception as well as thing-perception. If this view is correct, one would expect these other aspects of perception to have many of the same gestalt properties as thing-perception. dependence on proximity, closure, "good form," and so forth. In fact, the few pieces of work I know of on perception of entities other than things (Michotte (1954) on causation, Jenkins, Wald, and Pittenger (1976) on events, remarks of Köhler (1947, pp. 89-90) on temporal grouping) do reveal just what we are led to expect. This suggests that there is no fundamental new difficulty for perception in admitting entities other than things into conceptual structure--just more of the old problem of how we perceive anything at all.

The argument of section 2 also explicitly addresses the important philosophical issue of the ontological commitment of natural language semantics: what entities should semantic theory allow language to talk about? The predominant philosophical tradition, modeling itself after mathematics, tries to minimize primitives and axioms. (It is perhaps not insignificant that Frege and Russell, the founders of modern logic, were most deeply concerned with foundations of mathematics.) Thus logicians mostly confine themselves to semantic systems which contain only things and sets as primitive ontological types. A few, such as Davidson (1967, 1969), have tried to argue that events and actions are necessary as well. But Davidson himself has some qualms about such entities, because criteria of individuation are not easy to define.

The view of language taken here, however, takes the ontological commitment of natural language semantics not to be a question of elegance, but an empirical question: what possible ontological types are psychologically real? The argment presented here, based on the interaction of language and visual experience, provides simple preliminary evidence for a relatively rich ontology. As further justification, one would hope to show that this ontology is necessary on both language-internal and perception-internal grounds independently. (Jackendoff (1978) gives linguistic arguments to justify the notion place.) One would also hope to come to terms with the philosophical traditions concerning reference, which will be drastically affected by the expansion. I am not prepared to reconstruct the entire edifice at this point; some arguments will appear in Jackendoff (in preparation).

4. Conclusion

The main argument of this paper combined perceptual and linguistic evidence to show that figural expressions in conceptual structure must include entities of a great number of ontological types. I take this to be a prototype for a novel sort of linguistic argumentation--one that treats descriptive semantics as fundamentally a psychological rather than logical discipline, and which seeks to account for the nature of thought and of human experience through grammatical structure. It is not clear that this is linguistics in the usual sense any more. Rather it is an attempt to use linguistic theory as a tool of cognitive psychology. This seems to me to be a promising way to go.

References

Carey, Susan (1976) "A Case Study: Face Recognition," in Explorations in the Biology of Language, Report of the MIT Work Group in the Biology of Language, mimeo, pp. 229-264.

Chomsky, Noam (1975) Reflections on Language, New York, Pantheon.

Davidson, Donald (1967) "The Logical Form of Action Sentences," in N. Rescher, ed., The Logic of Decision and Action, Pittsburgh, Univ. of Pittsburgh Press.

----- (1969) "The Individuation of Events," in N. Rescher et al., eds., Essays in Honor of Carl G. Hempel, Dordrecht, Reidel.

Fodor, Jerry (1975) The Language of Thought, New York, Crowell.

Goldsmith, John, and Erich Woisetschlaeger (1976) "The Logic of the Progressive Aspect," Bloomington, Indiana University Linguistics Club.

Hankamer, Jorge, and Ivan Sag (1976) "Deep and Surface Anaphora," Linguistic Inquiry 7.3, 391-428.

Helmholtz, Hermann (1885) On the Sensations of Tone, New York, Dover reprint, 1954.

Jackendoff, Ray (1978) "Grammar as Evidence for Conceptual Structure," in M. Halle, J. Bresnan, and G. Miller, eds., Linguistic Theory and Psychological Reality, Cambridge, MIT Press.

----- (in preparation) Semantics and Cognition

Jenkins, James J., Jerry wald, and John B. Pittenger (1976) "Apprehending Pictorial Events An Instance of Psychological Cohesion," in Minnesota Studies in the Philosophy of Science, Vol. 9.

Koffka, Kurt (1935) Principles of Gestalt Psychology, New York, Harcourt, Brace.

Köhler, Wolfgang (1947) Gestalt Psychology, 2d ed., New York, Liveright, Mentor Books reprint.

Michotte, A. (1954) La Perception de la Causalite, 2d ed., Louvain, Publications Universitaires Louvain.

Miller, George, and Philip Johnson-Laird (1976) Language and Perception, Cambridge, Harvard University Press.

Wittgenstein, Ludwig (1953) Philosophical Investigations, Oxford, Blackwell.

This research was supported in part by a Fellowship for Independent Study and Research from the National Endowment for the Humanities.

ON THE ONTOLOGICAL STATUS OF VISUAL MENTAL IMAGES

Stephen Michael Kosslyn
Harvard University

There has long been considerable controversy over the ontological status of mental images. Most recently, members of the A.I. community have argued for the sufficiency of "propositional representation" and have resisted the notion that other sorts of representations are functional in the human mind. The purpose of this paper is to review what I take to be the best evidence that images are distinct functional representations in human memory. Before reviewing these data, however, I offer a preliminary definition of what I mean by a "visual mental image." This definition arises out of the "cathode ray tube" metaphor originally introduced in Kosslyn (1974, 1975, 1976) and later implemented in a computer simulation by Kosslyn & Shwartz (1977a, in press). On this view, images are spatial representations in active memory generated from more abstract representations in long-term memory; these spatial representations are able to be interpreted ("inspected") by procedures that classify them into various semantic categories.

1.0 A preliminary definition of a visual mental image

I wish to define a "visual mental image" in terms of five basic kinds of properties. Images are often distinguished from more discrete, propositional or linguistic representations because they supposedly have "analogue" properties. Thus the first two properties noted below describe analogue representations as a class. Goodman (1968), Palmer (in press), Shepard (1975), Sloman (1975), and others have provided informative and detailed discussions of relevance here, and I will draw freely on these sources in the present discussion.

1) Images can capture continuous variations in shape. This continuity property implies that image representations are both semantically and syntactically "dense" or "undifferentiated" in the extreme (Goodman, 1968, p. 136 ff.). For example, a reading on a tire pressure gauge is an analogue representation to some extent, because every reading along the continuous scale has meaning (and so it is semantically dense); if the gauge had an infinity of markings of pounds-per-square inch, the scale would be syntactically dense and readings on it would be purely analogue. In contrast, discrete representations are not semantically or syntactically dense, but are differentiated (i.e., separable and distinct). For example, each reading of a digital clock, in contrast to the traditional dial variety, is entirely unambiguous in terms of its identity (i.e., is syntactically distinct) and its meaning (i.e., is semantically distinct). Images are both semantically and syntacically dense.

2) Part and parcel of the continuity property is the property that analogue representations are not arbitrarily related to their referents. Because analogue representations can be arranged on a continuum (e.g., of size), a symbol indicating a value falling between two others (e.g., an intermediate size) must refer to a value of the referent falling between the two indicated by the others (e.g., an object of intermediate size). Hence, unlike discrete representations, any given analogue representation cannot be assigned an arbitrary meaning (this point was first brought to my attention by Wilkins, 1977).

Because of this requirement, portions of images of surfaces or objects (involving two or three dimensions) bear a one-to-one structural isomorphism to the corresponding portions of the referent. That is, portions of the representation correspond to portions of the referent, and the spatial relations between portions of the referent are preserved in the image. This property has been described by Shepard (1975) as an "abstract first-order isomorphism." In

The work reported here was supported by NSF Grants BNS 76-16987 and BNS 77-21782. I wish to thank Willa Rouder for her assistance in preparing the manuscript.

this case, there is not a genuine first-order isomorphism, where a triangle is actually represented by something triangular in the brain, but there is a more abstract isomorphism where a triangle is represented by a set of representations corresponding to the vertices and sides standing in the proper relations. Thus images depict, not describe. While any symbol can be used to represent an object or part thereof in a description, the particular representation of such in an image is constrained by other representations--given that the interportion spatial relations must be retained in the image representation.

The following three additional properties follow from our CRT metaphor:

3) Images occur in a spatial medium that is equivalent to a Euclidean coordinate space. This does not mean that there is literally a screen in the head.[1] Rather, locations are accessed such that the spatial properties of physical space are preserved. A perfect example of this is a simple two-dimensional array stored in a computer's memory: There is no physical matrix in the memory banks, but because of the way in which cells are retrieved one can sensibly speak of the inter-cell relations in terms of adjacency, distance, and other geometric properties.

4) The same sorts of representations that underlie surface images also underlie the corresponding percepts. Hence, in addition to registering spatial properties like those of pictures, images depict surface properties of objects like texture and color. Thus, although the image itself is not mottled, or green, or large or small, it can represent such properties in the same way they are represented in our percepts. That is, the image representations must be able to attain states that produce the Qualia, the experience of seeing texture, color, size and so on.

5) Finally, by dint of the structural identity of image representations and those underlying the corresponding percept, images may be appropriately processed by mechanisms usually recruited only during like-modality perception. For example, one may evaluate an image in terms of its "size" (i.e., being depicted--the representation itself is neither large nor small) in the same way one would evaluate the representation evoked while actually seeing the object.

Images, then, share virtually all the properties of percepts, as opposed to properties of pictures or objects themselves. I refrain from making a complete identity because of a crucial difference: Perceptual representations are "driven" from the periphery, whereas images are somehow formed from memory. Hence, in both cases there may be particular kinds of "capacity limitations" that influence properties of the representation. For example (and this is an empirical question), images may be coarser and less detailed than the corresponding percept because of memory capacity limits.

These properties of images can be further understood in contrast to properties of "propositional" representations. Consider the two representations of a ball on a box illustrated in Figure 1. A propositional representation must have: 1) a function or relation; 2) at least one argument; 3) rules of formation; and 4) a truth value.

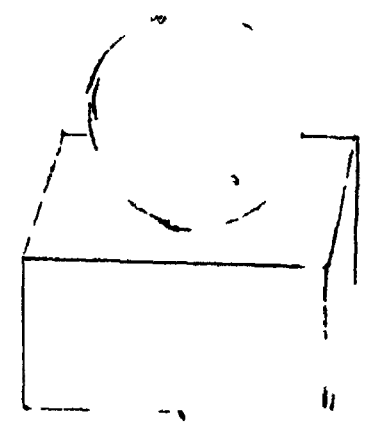


Figure 1. Two representations of a ball on a box.

In contrast:

1) Images do not contain identifiably distinct relations; relations only emerge from the conglomerate of the components being represented together. Thus, one needs two components before a relation like "on" can be represented.

2) Images do not contain arguments. The components of an image are not discrete entities that can be related together in precise ways. The box, for example, can be decomposed into faces, edges and so on--and these are certainly not arguments in and of themselves.

3) Images do not seem to have a syntax (except perhaps in the roughest sense). That is, a relation like "on" requires two arguments in order to create a well-formed proposition; "on box" is an unacceptable fragment. In contrast, any syntax dictating "well-formedness" of pictures or images will probably depend on some sort of interaction with a "semantic component," will depend on what an image is supposed to be an image of As we all know, "impossible pictures" are created regularly (e.g., by artists such as Escher), and rules that govern the nature of objects in the world may not

1. Although there could be, if images occur as topographic projections on the surface of the cortex; this kind of space is a subset of the one I am defining here, however.

necessarily constrain the things that one can depict in a picture.

4) Finally, unlike a proposition, an image does not have a truth value. In fact, as Wittgenstein (1953) pointed out, there is nothing intrinsic in a picture of a man walking up a hill that prevents one from interpreting it as a picture of a man sliding downhill backwards. The meaning of an image, and hence its truth value, are assigned by processes that work over the representation and are not inherent in the representation itself.

2.0 Five classes of empirical findings supporting the functional reality of visual mental images

2.1 Experiments on scanning visual images

A key property of images is that they embody spatial extent. If images are functional, then, we should expect this property to affect some forms of processing that involve using images. Kosslyn, Ball, & Reiser (1978) report a number of experiments that demonstrate that more time is required to scan further distances across mental images. In one study, people imaged a map containing seven locations and scanned between all possible pairs of locations. Time to scan increased linearly with increasing distance between the 21 possible pairs of locations, each of which was separated by a unique distance. There were no effects of distance in a control condition where subjects focused on a location in the image but then simply decided whether another object was present, without being asked to scan to that location.

In another experiment, people imaged schematic faces wherein the eyes were either light or dark and located either 3, 4, or 5 inches above the mouth; in all other respects the faces were identical. After a given face had been removed, a subject was asked to focus on the mouth and then to image the face as large as possible without it seeming to overflow, or image it half of this size, or image it so large subjectively that only the mouth was left visible in the image. Following this, the word "light" or "dark" was presented. As soon as either word had occurred, the subject was to "glance up" to the eyes of the imaged face and see whether or not they were appropriately described by the word. Time to judge whether the eyes were light or dark increased linearly with distance from the mouth. Further, overall scanning times were reduced when people were asked to "shrink" an imaged face mentally prior to scanning it, and times were increased when subjects "expanded" a face before scanning. These results are difficult to explain if images are simple "abstract propositional" list structures, but follow naturally if images are spatial representations that preserve metric distance information.

2.2 Measuring the visual angle of the mind's eye

The notion that images embody spatial extent suggests that they may have spatial boundaries; after all, they do not extend on indefinitely. If images occur in a spatial representational medium, then their maximal spatial extent may be constrained by the extent of the medium itself. Kosslyn (in press) used the following paradigm in an attempt to test this idea: People were asked to image an object as if it were being seen from very far away. Then, they were asked to imagine walking towards the object and were asked if it appeared to loom larger; all subjects reported that it did (of the subjects who could do the task at all, which was usually only about 80% of the people tested). Further, these subjects claimed that the image loomed so large at one point that it seemed to "overflow." At this point, the subject was to "stop" in his/her mental walk and to estimate how far away the object would be if s/he were actually seeing it at that subjective size. We did this basic experiment in a variety of ways, having subjects image various sorts of pictures or image animals when given just their names and sizes; in addition, subjects estimated distance by verbally assessing feet and inches or responded by moving a tripod apparatus the appropriate distance from a blank wall.

If images occur in a spatially constrained medium, then the larger the imaged object, the further away it should seem at the point of overflow. In addition, a constant angle should be subtended by the imaged objects (which ranged in actual size) at the point of overflow. Using simple trigonometry, we were able to compute the "visual angle of the mind's eye" from the estimated distances and longest axis of each imaged object. In all of our experiments, the basic results were the same: First, people claimed that smaller objects seemed to overflow at nearer apparent distances than did larger objects (the correlation between object size and distance was always very high), and distance usually increased linearly with size of the imaged object. Second, the calculated "visual angle" at the point of overflow remained constant for different-sized objects when subjects imaged pictures or objects that had just been presented. The actual size of the angle varied, however, depending on instructions: More stringent definitions of "overflow" resulted in smaller angles. These last findings imply that images do not overflow at a distinct point, but seem to fade of gradually towards the periphery. (The best estimate of the maximal angle subtended by an image while still remaining entirely visible seemed to be around 20 degrees.)

In another experiment, we asked

people to scan images of lines subtending different amounts of visual arc and we calculated how many msec were required to scan each degree. These people also scanned an image of a line they had constructed to be as long as possible without either end overflowing. The visual arc subtended by this "longest possible non-overflowing line" was inferred from the time required to scan across it. This estimate was very close to one obtained using the technique described above and to one obtained by simply asking people to indicate the subjective size of a longest non-overflowing line by holding their hands apart so as to span the length of the longest line.

2.3 Effects of subjective size on ease of "seeing" parts of mental images

If asked which is higher off the ground, a horse's knees or the tip of its tail, many people claim to image the beast and to "inspect" the image, evaluating the queried relation. It makes sense to suspect, then, that images might be appropriately processed by the same sorts of classificatory procedures used in categorizing perceptual representations. If so, then we might expect constraints that affect ease of classifying parts perceptually also to affect ease of imagery classification. Parts of smaller objects are "harder to see" in perception, for example, and also may be harder to "see" in imagery. This result was in fact obtained (see Kosslyn, 1975); parts of subjectively smaller images of objects did require more time to classify mentally than did parts of subjectively larger imaged objects. In addition, simply varying the size of the part per se also affected time to examine an image. In this case, smaller parts--like a cat's claws--required more time to see on an image than did larger parts--like its head. This last result was obtained (Kosslyn, 1976) even though the smaller parts were more strongly associated with the animal in question, and were more quickly verified as belonging to the animal when imagery was not used (more highly associated properties are typically affirmed as appropriate more quickly than less associated ones in studies of "semantic memory"--see Smith, Shoben & Rips, 1974 These findings, then, not only are consistent with the notion that images are functional spatial representations that may be interpreted by other processes, but also serve to distinguish between processing imaginal and non-imaginal representations.

2.4 Effects of subjective size on later memory

If parts of subjectively smaller images are less distinct, then one might expect that the imaged object itself would be more difficult to identify. Thus, if one actually encodes a subjectively small image into memory, one's ability to recall the object later should be poorer than if the image had been larger--if in fact the image itself is recalled and inspected when one tries to recall the encoded words or objects. Kosslyn & Alper (1977) asked subjects to construct images of the objects named by pairs of words. Sometimes one of the images was to be very small subjectively and sometimes both images were to be "normal" sizes. When a surprise memory test for the words was later administered, memory was in fact worse if one member of a pair initially had been imaged at a subjectively small siz This result was replicated in seve studies, each of which controlled for different possible confoundings (e.g., less "depth of processing" may have occurred when people constructed subjectively smaller images).

2.5 Transforming visual images

Cooper & Shepard (1973a, 1973b) and others have demonstrated that increasingly more time is required when one "rotates" a mental image through progressively greater arcs. Similarly, we have found that more time is required to expand or contract images to greater degrees (Kosslyn & Shwartz, 1977b), as did Sekular & Nash (1971). A propositional model of the sort offered by Gips (1974) does not lead us to expect these results. A spatial model, wherein a pictorial image is transformed, seems to imply in a straightforward manner that images will pass through intermediate positions as they are transformed, given that the same image is being retained and processed.

3.0 Conclusions

On my view, the most parsimonious, straightforward accounts of all these data will include the notion that images are functional representations in human memory. I have no doubt that alternative non-imagery accounts can be formulated for each set of results, but the collection of each of these individual accounts will likely be more ad hoc, post hoc and cumbersome than the imagery accounts.

References

Cooper, L. A. & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973a.

Cooper, L. A. & Shepard, R. N. The time required to prepare for a rotated stimulus. Memory and Cognition, 1973b, 1, 246-250.

Gips. J. A syntax-derived program that performs a three-dimensional perceptual task. Pattern Recognition, 1974, Vol. 6, 189-199.

Goodman, N. Languages of Art: An Approach to a Theory of Symbols. Indianapolis, Indiana: Bobbs-Merrill, 1968.

Kosslyn, S. M. Constructing Visual Images. Ph.D. dissertation, Stanford University, 1974.

Kosslyn, S. M. Information representation in visual images. Cognitive Psychology, 1975, 7, 341-370.

Kosslyn, S. M. Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. Memory and Cognition, 1976a, 4, 291-297.

Kosslyn, S. M. Measuring the visual angle of the mind's eye. Cognitive Psychology, in press.

Kosslyn, S. M. & Alper, S. N. On the pictorial properties of visual images: Effects of image size on memory for words. Canadian Journal of Psychology, 1977, 31, 32-40.

Kosslyn, S.M., Ball, T. M., & Reiser, B. J. Visual images preserve metric spatial information Evidence from studies of image scanning. Journal of Experimental Psychology Human Perception and Performance, 1978, 4, 47-60.

Kosslyn, S. M. & Shwartz, S. P. Two ways of transforming mental visual images. Psychonomic Society Meetings, Washington D. C., 1977b.

Kosslyn, S. M. & Shwartz, S. P. Visual images as spatial representations in active memory. In E. M. Riseman & A. R. Hanson (Eds.), Computer Vision Systems. New York: Academic Press, in press a.

Palmer, S. E. Fundamental aspects of cognitive representation. In E. H. Rosch & B. B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, N. J.: Lawrence Erlbaum Associates, in press.

Sekuler, R. & Nash, D. Speed of size scaling in human vision. Psychonomic Science, 1972, 27, 93-94.

Shepard, R. N. Form, formation, and transformation of internal representation. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1975.

Sloman, A. Afterthoughts on analogical representation. Paper presented at the Conference on Theoretical Issues in Natural Language Processing. Cambridge, Mass. June, 1975.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory A feature model for semantic decision. Psychological Review, 1974, 81, 214-241.

Wittgenstein, L. Philosophical Investigations. New York Macmillan & Co., 1953.

What has language to do with perception? Some speculations on the Lingua Mentis.

Zenon W. Pylyshyn

partments of Psychology and Computer Science
University of Western Ontario
London, Canada

1. Introduction.

The topic under consideration in this conference session (viz. Language and Perception) is not the one to which the greatest amount of attention has been devoted in philosophy of mind and philosophy of language. There a major concern has been the relation between language and thought. As everyone knows there has been a long standing dispute regarding whether or not it makes sense to view thought as being carried out in the medium of natural language or whether some other form of representation is involved. There has not been a comparable dispute over the relationship between perception and language. For one thing no one to my knowledge has proposed that perception occurs through the medium of natural language (though some early behaviorist writings come close, especially in respect to memory for perceptual events). What I propose to consider in this brief note are some respects in which the language-thought relationship is similar to the language-perception relationship.

2. Language and Thought.

At least since Aristotle there has been speculation and argument concerning the form (or language) of thought. Many contemporary philosophers (e.g., Quine, Sellars, Harman) as well as some past students of language (e.g., Whorf, Humbolt) believe that we think in our "outer" natural language: that knowing a language is being able to think in it. Harman (1975) takes a sophisticated approach to this position. He argues that in thinking in one's spoken language one need not parse or disambiguage it--since that would get us into the vicious circle of having to parse the thought into something which itself would be a thought and hence in need of further analysis. In Harman's view our thoughts are carried by "sentences under analysis" or by ambiguity-free already analysed sentence structures (e.g., P-markers). One problem with this view is that it denies the possibility of thought in animals and pre-verbal children. Other difficulties were recognized by psychologists. In the beginning of this century the Wurzburg school was able to argue that much of our thinking was unconscious. A more modern view (e.g., Paivio, 1975) takes the conscious experience of thoughts as occurring in language or in imagery as its starting point and demonstrates by operational means that at least two distinct modes of thought need to be postulated. This "dual code" approach is quite widely held in psychology although it is not precisely clear what intrinsic properties are being claimed for the imagistics mode of thought. But more on this later.

My own view, which I have been espousing for some half dozen years, is that an adequate account of the process underlying thought will show it as occurring in a symbolic mode which has few of the properties we would normally ascribe to either natural language or to images. For example, the vehicle of thought does not require words (but only concepts) nor does it have such intrinsic properties as size or shape. Rather it consists, as do all computations, of the transformation of formal symbolic expressions whose terms are given an intentional interpretation by the theoretican. In other words, thought is a symbol manipulation process. Because the data structures representing thoughts have an implicit syntax and because its terms and composite expressions are interpreted, one can think of them as expressions of an internal language-or lingua mentis--call it "mentalese".

While the particular arguments and examples I have presented in support of this position have varied over the years the thrust of the arguments has always been a two-pronged one. On the one hand I maintain that criteria of explanatory adequacy require one to give an account of certain specifically cognitive phenomena in a manner which neither presupposes certain crucial properties which themselves require a cognitive explanation, nor avoids a complete process explanation (involving a reduction to primitive mental operations) by attributing certain phenomena to intrinsic features of the brain. On the other hand, the argument has always appealed to empirical evidence. It is the dual requirement of meeting explanatory criteria and empirical evidence that has, for me, been the basis of my rejection of specific imagistic models such as those of Paivio, Kosslyn, and Shepard.

This is obviously the wrong forum in which to continue this debate especially since many of the details are peripheral to our present concerns. However, I do want to elaborate very briefly on what I referred to above as criteria of explanatory adequacy since I believe that this is the real crux of the debate, not only over imagery accounts of thought but also over some of the

issues about language and perception I want to raise later. Further details can be found in Pylyshyn (1978a).

The issue about explanatory adequacy is the following. Positivist doctrine not withstanding, an explanation of a phenomenon has to do more than predict or duplicate aspects of the phenomenon. It must also explicitly characterize the properties of the system by virtue of which the observed (or predicted) behavior occurs. Since some of these properties are adventitious or ad hoc while others are principled, such a characterization is essential. Furthermore, the account must separate *properties which are fixed and universal from those* which vary from task to task. To use an analogy from logic, it must separate the contribution of the notation, the logical axioms and inference rules from the particular premises used in deriving entailments. In the case of a process theory it is not sufficient to simply provide a procedure which generates behavior similar to that observed in humans. We must, in addition, explicitly isolate those properties and mechanisms which will remain fixed over all cognitive processes (the underlying system architecture), those which can vary gradually with learning or accomodation but whose component parts and intermediate states are not available to the whole system (the compiled skills), and those which represent particular methods adopted for particular tasks or which represent particular knowledge which the system possesses (and thus which can change freely). Furthermore, this parametrization or attribution of behavior to separate sources must be individually empirically justified--e.g., we must show that it is reasonable to postulate such properties of the architecture as we do by appealing to empirical evidence. If we can in this way isolate the fixed properties and show how these can be combined to produce the observed behavior, then we would have an account of the behavior which refers it to both fixed universal properties and to particular task specific ones. Such an account would not only capture cross-task generalizations but it is the best we can do from a cognitive or functional point of view. Further explication would involve describing, for example, how the fixed properties are realized in neural tissue or how and why the variable aspects got to be the way they are given the nature of the environment-organism interactions. An account partitioned this way would provide a means of deducing current behavior from fixed universal properties of mind and hence would provide a basis for explanation.

My main objection to such notions as analogues and to such hypothesized mental operations as scanning and rotation (to cite just two) is that the empirical evidence does not support the position that these are primitive properties of the mental architecture. I have argued that in all the proposals I am aware of which postulate analogues or analogue-like operations on images, there is independent evidence that the phenomenon in question must be attributed, at least part, to tacit knowledge which the system or person) possesses or to more articulated and piece-meal processes than those claimed. In other words these analogue operations cannot be taken as explanatory primitive operations in the mental architecture. Consequently to explain the experimental findings that these terms were introduced to account for we are forced to show how they could be carried out in an architecture in which scanning and rotation are not primitive operations. In such an architecture the processes might be quite different (e.g., while there might be a subroutine that accomplishes scanning or rotation, these particular terms would only be descriptive and not explanatory since the functions implied by them would in turn have to be explained in terms of more detailed computations using other more primitive and independently justified operations). The exact form of the argument against the hypothesis that scanning or rotation are primitive operations in the fixed mental architecture can be found in Pylyshyn (1978a, 1978b). Essentially they depend on showing that certain empirical facts (e.g., that rate of rotation depends on properties of the figure, the probe, and the task in general) require for their explanation that we specify more detailed processes which carry out the function described as rotation or scanning, thus demonstrating that the function was not a primitive.

The general conclusion I draw from these arguments is not that talk of analogues or other non-symbolic systems is incoherent or logically ruled out, but only that none of the phenomena which people typically appeal to have been shown to require them--and even if they were admitted they would, at least in these instances, not be explanatory in the required sense, though they might well be predictive (but then so would a multiple regression equation). Within the information processing paradigm (i.e., excluding phenomenological or purely neurophysiological explanations for reasons which we cannot go into here) the only remaining candidate paradigm for explaining the nature of thought is computation, in the sense of transformations on symbolic expressions. Of course within this alternative we may still posit different symbols, and even different composite data structures for different areas of cognition. What I am saying, however, is that this most basic level of symbolic representation is the modality independent medium of thought, the "mentalese" in which goals, beliefs, hypotheses, knowledge, and other cognitive states are expressed.

What makes this point of view on the relation between language and thought relevant to the perception-language discussion is that mentalese is not only taken to be the form in which thoughts are carried. it is also proposed as the appropriate representation of percepts.

3. Language and Perception.

Before discussing the similarities between the language-perception relation and the language-thought relation it may be useful to consider why one might be motivated to ask about the relation between language and perception in the first place. An obvious connection between the two is the fact that we can talk about what we perceive. But that tells us little about how the two are related. We get hints that the relation may be more intimate from the widespread use of perceptual terms (especially spatial relation and movement or transfer terms) to refer to abstract relations in general. The experiments on imagery by people like Shepard, Kosslyn, Moyer, Paivio, and others show,

if nothing else, that perception and thought are closely related. Thus the issues raised in discussing language and thought become relevant here too.

But perhaps one of the main reasons why language and perception are inextricably related is that the perceptual system is the primary means through which language acquires a semantics. A system which contained a body of data and a language processor might conceivably be able to carry on a coherent dialogue. But without a perceptual component it would, in an important sense, not know what it was talking about. We could, in principle, change the ASCII coded strings in its lexicon and it might conduct an equally intelligent conversation on an entirely different topic without anything (other than the external tokens) having changed. This is possible because the only constraints in the system are intra-linguistic ones and hence only linguistic and data-base consistency can be detected. In such a system there is no correspondence between internal symbols and things and hence the system makes no reference to the world.[1] This argument is made with painful force by Fodor (in press). It would be more obvious to people in A.I. that this is indeed the case if they heeded McDermitt's (1976) suggestion and refrained from using English words and phrases inside their programs and only employed nonsensical atomic symbols (GENSYMS). In that case it would be clear that only the programmer (and not the program) knew what it was talking about.

There is in fact a general and largely ignored problem of the distribution of explanatory burden between program and programmer that needs to be explicitly acknowledged in discussions in which programs are presented as theories. I have come to realize over the years that any crackpot theory can be implemented on a computer in some sense or other simply by assigning the appropriate names to various things in the program (e.g., call this buffer "consciousness", that data structure an "image" and this procedure "the mind's eye"). Elsewhere (Pylyshyn, in preparation) I have suggested a number of ways in which some of the arbitrariness can be taken out of this enterprise. They include independent validation of the "fixed mechanisms" that are to serve as the primitive components out of which cognitive processes are constructed (what I called the mental architecture) and the independent provision of at least a partial intrinsic semantics for symbols in the system by relating them to perceptual and motor subsystems. A further step might also be to provide the system with a learning component (in the very general sense of a history-dependent relationship with an environment) which would also serve to constrain the interpretation of symbols by connecting it to the physical world through a historical causal chain (c.f., Kripke's, 1972, causal theory of reference).

Now if we accept that in order for a system to have a semantics, as opposed to merely a complex intra-verbal deductive system operating on uninterpreted symbols (or "logical forms"), it must at least have a perceptual component, a number of fundamental questions arise. Though the whole issue of semantics is fraught with difficulty I will take advantage of the invitation to speculate by rushing in where many have been lost. The questions I shall in a sketchy way comment on concern the nature of the perception-language correspondence, the way in which this correspondence might be represented, and how such a correspondence could arise in the first place.

3.1 The nature of the language-perception correspondence.

Since the set of perceptual patterns and the set of definite descriptions are both unbounded, the correspondence between the two cannot be through existing associative links. The mapping can only be given by a recursive procedure which associates subpatterns of the language with subparts of the percept--in other words the correspondence is between some analysis of both descriptions and percepts. We are of course no more aware of the conceptual analysis of percepts than we are aware of the analysis of linguistic inputs. Given the necessity of an analysis of both, the most parsimonious story of how this occurs is one which assumes that both are analysed into a structure in the same interlingua--viz., mentalese. Contrary to some of my critics on this point (Kosslyn & Pomerantz, 1977, Anderson, 1978) such a view is neither inconsistent nor unnecessarily complex. Independent arguments suggest that at least this much analysis or translation is necessary and there is, to my knowledge, no convincing argument that more than one form of interlingua is needed. Though this latter possibility is not ruled out, the relatively weak constraints placed on the formal properties of the representing medium at present (viz., that it consist of sumbol structures) make this possibility seem unlikely. Furthermore, the freedom we have in thinking about information received through all modalities and the readiness with which we forget (outside of experimental settings) how we came to know something argues that at least memories and thoughts might appropriately be viewed as being amodal.

Another question that arises in connection with the nature of the language-perception correspondence is whether the formal properties of the two are independent or whether one might be able to explain linguistic properties in terms of perceptual or general cognitive ones and vice versa. Such a possibility is most attractive since it would increase the explanatory power of the resulting theories. On the other hand, there is no a priori necessity that such an explanatory link exist. As Chomsky (1975) has frequently pointed out we do not expect to be able to explain why humans have certain physical characteristics (e.g , why they have 10 as opposed to 8 toes, etc.) so why should we expect to explain why the noun-verb dichotomy appears to be a linguistic universal. Still one might be permitted to hope for some economy of explanatory principles by unifying over cognitive domains.

There is already reason to believe that at least some of the lexicon can be explained in terms of universal properties of perception Perhaps the clearest and most familiar example is the case of color terms. Berlin and Kay (1969) have demonstrated that color terms in various cultures form a strict hierarchy so that languages with more color terms invariabley include the terms used by languages with few color terms. In this example, however, it has been

possible to go further and demonstrate universal color perception properties paralleling the linguistic findings and even to relate these to visual physiology. Denny (in press) has cautiously suggested the possibility of a similar hierarchy across cultures of lexical systems for spatial deixis. For example, compared to English's two terms "here" and "there", Kikuyu has 8 spatial deictic terms and Eskimo has 88, all forming an inclusive hierarchy.

It is not inconceivable that the structure of the lexicon will exhibit many such points of contact with perception--at least for concrete descriptive terms. Is there any reason to believe that this parallel might also hold for other parts of language--specifically for grammar? There have been suggestions that syntactic classes such as noun or verb or even adjective correspond to conceptual categories--to ways of conceptualizing the named entitities. There have even been occasional suggestions that grammatical rules are a reflection of how people conceptualize what they perceive.

We must be quite clear about what such claims can mean There is a sense in which these claims are very likely (but perhaps not too interestingly) true. For example, when I choose to say "that's a red ball" as opposed to "the color of that ball is red" it seems reasonable that I select a part of speech and grammatical form which highlights certain aspects of what I intend to assert. Grammar provides many options on how essentially the same propositional content can be asserted. These alternatives may differ in respect to which items are treated as figure and ground (or topic and comment). Which option we take on a particular occasion no doubt depends, at least in part on how we conceptualize the situation This, however, is very different from the claim that grammatical categories represent conceptual categories. Even less does it suggest that syntactic rules can be expressed in terms of conceptual properties In spite of considerable effort devoted to the problem no one has, to my knowledge, provided even a glimmer of hope that any particular grammatical rule of language bears anything but a conventional relation to things in the perceptual field. It is as though syntactic structure provides a sort of system of codes which can be exploited to carry conceptual distinctions even though the system of codes itself is independent of what it can be used to express. In fact the linguistic code is rather severely constrained by properties of the communication channel into which it encodes ideas, for example by the serial nature (i.e., low bandwidth) of our speech and hearing apparatus in contrast with the richness of our conceptualizations an our perception in general.

Since, however, language is in all likelihood a function of the same cognitive apparatus as is available for other cognitive domains, we might expect an influence to be apparent at some level--even if not at the level of rule structures. For example, if figure-ground organization was a primary mode of structuring perception and thought one might expect syntactic features of some kind to be used consistently to reflect this organization--even though the code could in principle also be used to represent quite a different type of conceptualization or the same conceptualization in a different way. Thus, it is entirely conceivable that some predicate-argument type of characteristic might be found in grammar, whether represented as a surface taxonomy or some less obvious way. Whether or not this is the case is an empirical question in respect to which I don't believe there is wide agreement at present.

When it comes to more abstract properties of language, such as some of the putative linguistic universals, I believe the possibility of showing parallels between language and other areas of cognition may be more hopeful. My rather tentative view on this is based on the belief that whereas the form of grammar may well be an unexplainable consequence of some properties of brain structure together with properties of channels of communication, sentence comprehension must be implemented on a system with the same architecture as that used in other areas of cognition. Consequently, there may be some very general processing constraints that might show up as linguistic universals. In any case, if they appear in linguistic data at all the effects of system architecture will be seen in abstract universals rather that particular language specific syntactic rules.

For example, one very general universal property which Chomsky (1975) has cited as evidence for the innateness of Universal Grammar is that of "structure dependent rule" Rather than infer the apparently simplest rule (or the rule whose features are most evident on the surface of the set of samples) the child infers more complex structure-dependent ones. For instance, rather than infer that declaratives and questions are related by virtue of a certain pattern of permutation of substrings of the sentences, the child learns that the permutation applies over an analysis of the sentence into abstract phrases. Thus, while the simple rule accounts for the relation between "The man is tall" and "Is the man tall?", this would produce the incorrect transformation of "The man who is tall is in the room" as "Is the man who tall is in the room". Yet children never make the latter error, thus suggesting that their hypothesis formulation capacity is constrained in ways characterized by Universal Grammar.

But structure-dependence is not only a phenomenon of language, it is also ubiquitous in perception. Even a casual examination of what is involved in visual tasks, such as the solution of geometrical analogy problems, makes it clear that the rules employed must be sensitive to various level of abstract structure as opposed to more superficial features of the figure. In fact it is characteristic of all of perception that the structuring of the perceptual field must be hierarchical. If we were to describe what a child learns in learning to perceive its world we would come to the same conclusions about vision as Chomsky does with language--viz., that the way in which the regularities of the visual field are captured is constained by innate mechanisms in a way which would be described as "structure dependent".

There have also been attempts to explain more specific linguistic universals--such as the Specified Subject or Subjacency constraints--in terms of general properties of the processor (e.g., Marcus, 1977). Such studies are only beginning but I have no doubt that some linguistic properties will eventually be attributable to architectual or strategy properties unique to the human cognitive

system. How much will be explainable this way remains an open question.

3.2 Representing semantics.

The much misused term "semantics" refers to the interpretation of a symbol system (in this case language) into some other domain. In a computer without a perceptual component the only symbols which strictly speaking have a semantics are ones which are either directly executable by the hardware or are translated into other symbols which are executable.[2] All other symbol structures which are referred to as semantic are really supports for the deductive apparatus. They simplify the process of deducing new expressions from old ones in such way as to maintain the truth of the expressions under a consistent interpretation. This interpretation, however, is provided by the user, not the system.

Often what is referred to as the semantic representation has some of the properties of a model. For example, it provides a set of objects which can be used to evaluate expressions, the way models are used in mathematics. In a sense then, these models form a domain of interpretation. They are not, of course, the ultimate intended domain of interpretation. Expressions are typically intended to refer, for example, to beliefs about objects in the real world, not to other symbols. But this formal model can itself be taken to represent such cognitive objects and so provides a formal semantics for the symbolic expressions which hopefully is valid in the intended domain. The design of such formal models is a major concern in A.I. and the computational version of such systems are typically hybrid mixtures of models and inference schemes. I will have very little to say about them here.

In a system which does contain a perceptual component there has to be some facility for translating between the perceptual analysis and the linguistic analysis. In order to deal with the "semantic content" of sentences and percepts we must provide the potential for cross-modality and extra-linguistic correspondence. I have suggested that the most parsimonious view of how this occurs is that the end products of both perceptual and linguistic analyses are conceptual structures, or expressions in a single symbol system which we call mentalese. Other alternatives are occasionally proposed. We shall very briefly examine one below.

There have sometimes been objections to the view that percepts are conceptually analysed into articulated symbol systems. Some people feel that this loses the holistic and continuous aspect which seems intuitively to characterize percepts. It is hard to know what to make of such intuitions. They seem to suggest to people something more than that we see distributed features (e.g., roundness) or continuous properties and therefore that the percept must represent such properties. Rather these intuitions seem further to suggest that the percept must <u>have</u> such properties--i.e., it must not only represent the property of continuity but it must actually be continuous. This is a dangerous direction to pursue, however, since it could lead one to also claim that percepts actually are large, blue, warm, heavy, etc., running us right into Leibniz's problem.

The only proposals I have seen for dealing with the holism concern are ones which propose unanalysed objects such as templates or holograms as perceptual representations. These are not only atomic wholes but are clearly relatable to the proximal stimulus, at least in the case of vision.

I have discussed such proposals elsewhere (Pylyshyn, 1974; 1978b). Their inadequacy stems from several sources. One is that by considering the percept to be holistic in this sense one loses the ability to attend selectively to parts or aspects of it or to notice the respects in which two such representations differ. Of course, one can gain this facility back by positing a process of comparison or analysis which yields the more detailed features--but this is just to postpone the translation into mentalese. Alternatively one might posit that the comparison itself is done by a non-symbolic holistic process like that used in matching holograms. But here we run into trouble with the sheer empirical facts concerning the cognitive structure of percepts. The type and degree of perceived similarity among stimuli cannot be matched by a uniform interpretation-independent process like the hologram one. To what extent and in what respect two things are perceived to be different depends entirely on what we perceive those things to be. In other words similarity must be defined over an already interpreted --and hence conceptual, nonuniformly detailed, preanalysed, and articulated--representation.

Even a compromise in which the representation is an articulated structure with something like "imagoids" or pieces of templates at its nodes will not help. For if those template pieces need in some cases to be further analysed then we are back with the problems sketched above. If, on the other hand, they do not need to be analysed then there is no distinction between this proposal and one in which the templates are replaced by atomic symbols--i.e., terms in the mentalese vocabulary. Recall that mentalese terms appear in the output from the perceptual system and thus can arise from such perceptual properties as "large", "round", "red" or ones for which there is no single word in English, such as "sand-like texture" or ones best displayed graphically. What mentalese terms there are--i.e., what well-formed perceptual categories exist--is an empirical question.

Whatever merits the proposals for imagistic or analogue representations may have they clearly do not help the language-perception interface problem since sooner or later the representation must be analysed in such a way as to be commensurable with natural language terms. Whether this is done at the time of perception, or postponed by storing an unanalysed proximal stimulus so that it must be done at the time of sentence generation, does not affect the basic problem. Other independent considerations, discussed in Pylyshyn (1973, 1974, 1978b), argue against the view that unanalysed stimulation is stored in memory.

3.3 The genesis of the language-perception correspondence.

In an earlier paper I noted three major preconditions for learning a language (Pylyshyn, 1977).

1. Sensory experience must be structured. The "blooming, buzzing confusion" of William James must be susceptible to segmentation, analysis, and re-

construction. Some aspects must be foregrounded relative to others so that the environment becomes articulated or differentially noticed in some fashion.

2. Communication codes (both verbal and nonverbal) must likewise be structured The stream of vocal or gestural behavior must be perceived as segmented and a distinction between signifying and nonsignifying variation must be made (in generation and/or perception).

3. The occurrence of a speech act must be recognized. This is perhaps the most important but most neglected aspect of preconditions for language acquisition Not only must a child attend to the appropriate aspects of his environment, but he must do it within the context of what Merleau-Ponty would call (loosely) an "intention to mean".

In this section I wish to deal primarily with the first of these preconditions and with what has to happen in order for a simple naming or describing correspondence to occur. I will not dwell on the other two preconditions except to note that, as the third precondition suggests, a simple associative pairing will not make one perceived pattern (e.g., a word) refer to another. The pairing must be conceptualized and subsequently treated as a particular kind of asymmetrical irreflexive relation called naming or reference. This in turn means that one pattern (e.g., a word) is not simply an indicator that, say, the other pattern is about to appear but rather becomes a symbolic surrogate for its referent. It can then be used in arbitrary cognitive combinations with other such surrogates. It can be used not only instrumentally to anticipate or to ask for objects, but also to think about, hope for, question, assert something about, plan for, and vicariously play with the designated object.

What I would like to consider in a general way is how a linguistic sign or word can come to refer to something in the perceptual field Take the simple example of naming by ostention. A child is shown a dog and the word "dog" is uttered. Suppose the preconditions are fulfilled. The first problem to be faced is the well known difficulty of how the child is to know that what is being pointed at is the object rather than any of its properties. Alternatively, how is the child to know whether the word refers to that very object lying on the carpet with a collar around its neck and a bone in its mouth or any member of the Cocker Spaniel family or any canine or mammal or living creature, and so on.

First of all it is clear that what the speaker is referring to must be a conceptually integral unit for him--something he can conceptually detach from his cognitive or phenomenal field Secondly, if the hearer is to have any chance of acquiring the same referent for that word he will also have to to have conceptualized the field in such a way as to individuate the same entity as the speaker. Given the unlimited number of in-principle possible ways of analyzing the entire ostention situation, nothing short of a miracle could ensure that the same analysis was given by both participants. Nothing, that is, except a highly constraining universal innate mechanism that severely limits the set of alternatives which are humanly conceivable.[3] What this in turn comes to is the claim that the terms of mentalese are innate This outrageous claim, which is argued for in considerable detail by Fodor (1975), is also pressed on us by other considerations which we take up below. Thirdly, the listener must use both his perception of the physical situation and his understanding of the social context to infer the intentions of the speaker. This gives definition-by-ostention a problem-solving character.

John Macnamara (1972) has revived interest in the view, often associated with St. Augustine, that "...infants learn their language by first determining, independent of language, the meaning which the speaker intends to convey to them, and by then working out the relationship between the meaning and the language (p. 1)." In other words the child has various sources of evidence concerning such things as what objects, classes and properties are in his environment and what the adult intends to convey, say, by pointing and speaking a word. His task is then to make the inference to the best hypothesis concerning the correspondence between these events But the question arises, how is the hypothesis formulated? Clearly this view assumes that the relevant aspects of thought and perception (my first precondition) are present prior to language learning. This in turn presupposes that the terms of mentalese are also available prior to language learning since the hypothesis must be expressed in mentalese. But how then is mentalese acquired?

The answer is that if it is "acquired" at all no one has the slightest idea how this could possibly occur. The only notion around (as Fodor, 1975, has argued) regarding how a new concept (or term of mentalese) could be learned is one which says that what people learn is the relation of the new concept to some relational structure of already known concepts. But this precludes the learning of any concepts that are not definitional composites of old ones, and therefore strictly eliminable. Unfortunately, this appears to include most natural concepts Like the theoretical terms in science, most natural concepts cannot be given a context-free definition but rather depend on the entire system of concepts for their meaning (which is why dictionary definitions are invariably circular). While one can speak of the accomodation of linguistic usage (e.g , the referents of words can vary as we discover new empirical facts--such as that both steam and ice are really just forms of water), the accomodation of the mental concepts, in terms of which the linguistic terms can be understood, remains a mystery. The mystery is not lessened, moreover, by talk of motor schemata or "equilibration" as Piaget does. In each case of putative conceptual change the process either depends on assimilating new concepts into arrangements (or schemata) made up of old concepts, thus severely limiting the type of conceptual change possible, or it is left unexplained There is no explanation, nor even the beginnings of an approach, for dealing with the accomodation of schemata or conceptual structures into ones not expressable 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.

This is in fact an extremely deep problem about which very little sense has been made. People are sometimes mislead by certain computational metaphors into believing that the problem can be dispensed with by something like compilation. But however attractive that

notion is, as a way of talking about how new procedures can come into being which are themselves expressed in terms of new operations, it does not generalize to concepts in general Such a notion works in the case of procedures because the set of computable functions is closed and reduceable to elementary (Turing machine) operations in a way that the set of conceptualizations of the world is not.

It seems to me that there are two general avenues open for dealing with this dilemma, both of which simply raise more questions than they answer. In both cases what we are doing is opting for a different locus for the mystery, rather than resolving it.

The first approach is to simply 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 means accepting that virtually all unitary concepts of which we are capable are genetically determined. Compound concepts (such as circular red object) can also be constructed as well as definitional composites, but these constitute a minority of our mentalese vocabulary. Of course, there need not be (and in fact certainly will not be) a one-one correspondence between concepts and words in the spoken language. It is quite likely that most words do correspond to concepts, though there have been suggestions that some words are represented by compositions of more primitive concepts (e.g., kill = do something to cause to die; never = not ever). So far few, if any, of these suggestions have withstood empirical tests (c.f., Fodor, Fodor, and Garrett, 1975). Clearly, however, not all mentalese terms correspond to words. Not only do societies differ in their basic vocabulary but the view of mentalese we have been discussing requires terms for stable perceptual features which are not encoded in our language, at least not as single words

While the notion of all our concepts being innate is repugnant to the contemporary Zeitgeist, part of this attitude may be due to the connotations of this way of speaking. If we thought of the innate mentalese vocabulary as corresponding to the fixed structural properties of the computational system, together with the input-output transducers, this might not seem as distasteful. Even the simplest modern computer has a considerable amount of fixed hardware (i.e., innate) structure--including a facility for discriminating an unlimited number of formal atomic symbols. If each of these symbols had predetermined potential referents (say by virtue of the way they were wired to mechanisms which were eventually connected to transducers), they could be considered innate concepts. Of course this is not the whole story since it is hard to see how many of the required concepts (e.g., Kant's transcendental categories such as space, time and cause) could be thought of as wired to transducers. The problem here is that it is still not very clear what the force of the claim is when we say that concepts, qua interpreted symbols, are innate Conceivably it could mean little more than that the constraints on the system of symbols is so great that the class of possible interpretations (like the class of realizeable grammars) is extremely limited. In fact one way that the class of possible interpretations could be characterized might be to formulate them in terms of the requirement that the only concepts the organism can hold are ones expressable in terms of a certain "innate vocabulary". In that case, "innate vocabulary" has the same status as "universal grammar"--viz., they both somehow characterize the endowed cognitive capacity of the organism.

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?

Another approach to this dilemma is to locate the puzzle in yet another quarter. We think of the "innate concepts" as being the representational capacity of the fixed hardware architecture--so that mentalese becomes identified with machine language. The innate concepts are thus not truly concepts but, as suggested above, symptoms of the interpretive constraints imposed by the computational architecture on the system of available symbols. Now the symbols do have to be exploited in representing the world, and for any particular machine architecture their interpretability is constrained in certain ways For example, if a certain subset of available atomic symbols is treated in a certain way by the motor transducer (e.g., cause the hand to open or the arm to reach out) then they cannot consistently be interpreted as, say, referring to phonemes

Now the problem we had was to explain how new concepts can develop which are not definable in terms of old ones. This is the essense of radical conceptual change or accomodation. The paradox arose because the only formal mechanism which seemed to be available was symbolic composition (or definition). A whole new realm of possibilities opens up however if we allow non-symbolic changes to occur--i.e., if we allow the actual hardwared connections or architecture to change Concepts can then drift or mutate insofar as the constraints on symbols can change in novel ways.

The trouble with this proposal, of course, is that it is nothing more than a burying of the problem into hardware. So long as the relation between hardware and symbolic levels is not systematically understood--so that, for instance, we had some formal rules for how the underlying architecture could change in response to programmed instructions --then this proposal is not a real alternative. It does, however, contain one recurring suggestion which seems to surface in many different contexts and for many different reasons (most, in my view, are invalid)--viz., that there are some cognitive functions whose realization will require that we transcend the symbolic mode and deal with physical (or, at any rate, a quite different set of symbolic) processes. Maybe that's what Kant had in mind when he spoke of "transcendental reasoning".

Footnotes

1. The fact that a system without instrinsic semantics could conceivably still pass the Turing test and meet Newell's criterion for

understanding (viz., "S understands knowledge K if S uses K whenever appropriate") suggests that such criteria may show that there is a difference among (a) achieving "understandin" (b) knowing what things, properties, etc. in the world are being referred to, and (c) explaining what such understanding consists in, or what it means to comprehend on utterance. As noted earlier, criteria of performance are distinct from criteria of explanation.

2. Even numerals are not interpreted by the machine. The transformations of numerals into numerals carried out by what are called arithmetic commands are just formal operations on symbols. The user typically interprets the symbols as designating numbers and the operations as designating the usual arithmetic operations but he could just as well interpret the symbols as, say, propositions and the operations as deductions (though the interpretation function might be quite complex)--or any other interpretation which happens to maintain its coherence.

3. It is understandably not easy to provide an example of a humanly inconceivable unitary concept. Goodman's "Grue" and "Bleen", introduced to highlight certain problems of induction, may be such examples. Grue is the unitary concept which in English corresponds to the color description "Has a green color up to time t and a blue color after". Thus in the new system green would be the name given to that strange color which is Grue up to time t and Bleen afterwards So far as anyone knows, concepts like Grue and Bleen never occur in human cultures. However we must not be too presumptive about what concepts actually can exist. Exotic societies frequently provide examples of what are for us inconceivable ways of carving up experience. For example Foucault (1972, xv) quotes Borges' citation of an ancient Chinese encyclopedia which has the following strange taxonomy. "Animals are divided into (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies." If very strange concepts do exist we might find it very hard to decipher them, given our constrained schemata

References

Anderson, J. R. The status of arguments concerning representations for mental imagery. Psych. Review, in press.

Berlin, B., & Kay, P. Basic color terms. Berkeley: Univ. of California Press, 1969.

Chomsky, N Reflections on language. New York. Pantheon, 1975.

Denny, J. P. Locating the universals in lexical systems for spatial deixis. Papers from the 14th regional meeting of the Chicago Linguistics Society, 1978, in press.

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

Fodor, J. A., Tom Swift and his procedural grandmother. Cognition, in press.

Fodor, J. D., Fodor, J. A., and Garrett, M. F. The psychological unreality of semantic representations. Linguistic Inquiry, 1975, 6. 515-531.

Foucault, M. The order of things. London: Tavistock publication, 1970.

Harman, G. Thought. Princeton, N.J.: Princeton Univ. Press, 1973.

Kosslyn, S. M., & Pomerantz, J. R. Imagery propositions and the form of internal representations. Cognitive Psychology, 1977, 9. 52-76.

Kripke, S. A. Naming and necessity. In D. Davidson and G. Harman (eds.), Semantics of natural language Dordrecht D. Reidel, 1972.

Macnamara, J. Cognitive basis of language learning in infants. Psych. Review, 1972, 79, 1-13.

Marcus, M. P. Theory of syntactic recognition for natural language. M.I.T. Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, 1977.

McDermitt, D. Artificial intelligence meets natural stupidity. Newsletter of the ACM Special Interest Group on Artificial Intelligence (SIGART), 1976, 57, 4-9.

Paivio, A. V Neomentalism. Canadian Journal of Psychology, 1975, 29, 263-291.

Pylyshyn, Z. W What the mind's eye tells the mind's brain. a critique of mental imagery. Psych. Bulletin, 1973, 80, 1-24.

Pylyshyn, Z. W. The symbolic nature of mental representations. Paper presented at a conference on Objectives and Methodologies in Artificial Intelligence, Canberra, Australia, May, 1974 (mimeo)

Pylyshyn, Z. W. What does it take to bootstrap a language. In J. Macnamara (ed.) Language learning and thought. New York: Academic Press, 1977.

Pylyshyn, Z. W. The explantory adequacy of cognitive process models. Paper presented at a workshop on mental representation, M.I.T., January, 1978a (mimeo).

Pylyshyn, Z. W. Imagery and artificial intelligence In C. Wade Savage (ed.). Perception and Cognition: Issues in the Foundation of Psychology, (Vol. IX of Minnesota Studies in the philosophy of science). Minneapolis, Minn. University of Minnesota Press, 1978.

Pylyshyn, Z. W. Towards foundations for Cognitive Science. Book manuscript, in preparation.

SEMANTIC PRIMITIVES IN LANGUAGE AND VISION

Yorick Wilks
Department of Language and Linguistics
University of Essex
Colchester, England

The purpose of this brief note is to argue that, whatever the justification of semantic primitives for language understanding may be [see Wilks 1977] there is no reason to believe that it relates to vision in any strong sense

By "semantic primitives" I mean the general sort of item proposed within Artificial Intelligence (AI) by Wilks (1972, 1977), Schank (1973) and within linguistics by Fodor and Katz (1963), Jackendoff (1975) among many others, in both case The generality of these items is essential to my argument, and I shall not count as semantic primitives items used for special tasks, whether or not those tasks are related to vision, as are the visual description primitives of Johnson-Laird (1977)

Spatial versus visual

What follows is highly naive and speculative it will rest largely upon the opposition of linguistic knowledge to spatial and visual knowledge respectively I take it for granted that the latter are not necessarily connected, and so to establish that we need spatial knowledge to understand language (to name a task at random) does not establish that we need visual knowledge The lack of necessary connexion is shown by such hackneyed examples as the person blind from birth, who has no visual, but a great deal of spatial, knowledge

One initial reason for distinguishing the two is the great deal of argumentation in linguistics in recent years that falls under the general heading localism This thrust of argumentation has sought to establish the central role of spatial concepts in linguistics, and among its best known proponents are Anderson (1971), Fillmore (1977) and Jackendoff (1975) One stand in this view is to argue that temporal expressions are in general reducable, in some sense, to spatial ones that in ten minutes (a time expression) is dependent on the spatial sense of such forms as in five miles This is a very difficult and general debate there is contrary evidence from cultures where space is indicated by time (The airport is is about ten minutes away), and there is a strong philosophical tradition, centred round Kant, that our sense of time is logically prior to our sense of space That is to say, we could conceive of structuring our experience without the concept of space, but not without that of time because, if we could not know that one event preceded another, then we could probably not know anything at all, not even mathematics if that consists at bottom in sequences of operations Michotte's famous experiments on the willingness of subjects to attach the word cause to moving pictures of pairs of "striking billiard balls" is sometimes cited as providing a visual basis for causality (Clarke & Clark 1977), although the notion of causality may well in fact make no sense without the concept of time We could assert (wrongly, as it happens) that lightening causes thunder without the aid of a spatial concept, but not without a temporal one

The logical or linguistic priority of space to time is by no means a settled matter, and neither therefore is the thesis of localism have argued that the role of the visual in language is not necessarily supported by the need for spatial knowledge, and so the status of the latter need not be discussed Nonetheless, I have questioned the self-evidential truth of

localism, just in case anyone should think that, if it were true, it would support the centrality of visual knowledge in language understanding.

Let us now, as the brief substance of this paper, look at three arguments that might be put forward to support the dependence, or inter-dependence, of linguistic and visual knowledge.

Evolutionary arguments

This comes in phylogenetic and ontogenetic forms. The former is the ingenious argument (Gregory 1970) that, since the human race has been able to see for many times more millenia than has been able to speak or write, then it might seem reasonable to believe, on evolutionary grounds that the brain "took over" the existing visual structures for language understanding and production. This argument may well be true, but at present there is no independent evidence that would count for or against it.

The "ontogenetic form" of the argument - in the individual, that is - is that one first learns words essentially through the visual channel, and so again our linguistic knowledge is essentially dependent upon visual criteria and experience. The best quick answer is to turn to the sort of word often used as a semantic primitive in AI language understanding systems: STUFF (=substance), ATRANS (=changing the ownership of an entity), CAUSE (=preceding and necessitating an event) It is highly dubious that such very general concepts are, or can be, taught by visual/ostensive methods. Can one point at substance as such? One may want, or mean, to, but can one in fact reliably do so?

One structure for many purposes

This is a widespread view in AI that has been argued for explicitly by Minsky (1975) and Rieger (1976), among others. Roughly speaking, it is that implemented systems should use a single knowledge structure for a range of purposes: language understanding, problem solving, etc. It is an additional assumption that human beings do function in this way.

The thesis can be expressed at many levels, and at a sufficiently general level it is almost certainly true. But it might then mean no more than that a single programming language could express general sub-routines for parsing, noise reduction etc. for a number of input channels. At a more specific level was the thesis, not now widely supported, that language and vision in some sense shared the same "grammar", in the sense of Chomsky's transformational grammar (Clowes 1972). Striking evidence from the parallelism between visual and linguistic ambiguity was found, and the fact that Chomsky's grammars no longer seem such plausible candidates for such a role does not mean that the thesis itself is false at that level.

Let us concentrate for a moment on two more specific levels. First, consider the well-known contrast between such sentences as:

The paper moved

The dog moved

Linguists who differ about much else would want to ascribe a notion of agency to the subject of the second sentence but not the first. Many in AI working on natural language would agree, and add that the notion of agency is essential if other important inferences are to be made. But, surely no one would argue that agency is, in any useful sense ascribed a visual criteria, that could be reduced to the visual differences of paper and dogs. It is in fact a complex theoretical notion dependent upon our beliefs and theories about the world: we do not now attribute agency to trees, though some fellow humans do. But this difference is a theoretical (including linguistic) one, not one of difference of visual perception.

Secondly, we may return to general semantic primitives of the sort already mentioned (and similar inventories may be found in (Bierwisch 1970) and (Leech 1974)).

There are many possible ways in which one might seek to justify such primitives (see Wilks 1977), and Bierwisch (1970) has gone on record as saying that they do denote, and are to that extent dependent upon visually observable entities. I suggested above that that may not be so: one may point at treacle, water or elephant meat, but it is not so clear one can point at SUBSTANCE, yet this notion has a role to play in language understanding for how, without it, can one economically express such axioms as "A quantity of a substance plus a quantity of it

yield a quantity$_3$ of it" This axiom is not true of physical objects, as distinct from substances

A well-known confusion must be avoided here it may well be true, as the model theoretic semanticists like Montague claim, that any contentful notion, primitive or not, refers to a function of sets In that sense move might be said to refer to a set of entities that move

However, this point about logical reference has no consequences for the point about whether or not such primitives denote entities in the real world

Visual and spatial imagery

Finally, it is sometimes argued that the structures underlying language must depend upon those underlying vision if only because natural language is so full of visual imagery In whatever sense "visual imagery' is taken, this fact is, I believe, irrelevant to any precise assertion under discussion, by which I mean any of

I) Language understanding processes in humans depend, either as to primitive elements or structure, on visual experience and the mechanisms that interpret it

II) The specification of language in humans has no significant overlap, in terms of primitive elements or structure, with that of other faculties, like vision

III) Visual processes in humans depend, either as to elements or structure, on linguistic experience and the mechanisms that interpret and produce (sic) it

For all three theses only anecdotal evidence is available, though I would be strengthened by empirical evidence that the blind from birth were less able to understand the use of visual imagery in language Those with a predeliction for motor theories should be tempted to consider the Whorfian thesis III (Whorf, remember, believed we might perceive, say lightning, as an entity, rather than an activity or process because we denoted it by a member of the theoretical category NOUN, rather than VERB) since, as the structural difference of I and III makes clear, language is an activity in a way vision is not

Thesis II will be agreeable to those who are impressed by the way in which confusion can arise when one tries to bring together information on the same topic, but obtained via different channels As when one refers to two cities whose mutual relation of position one knows from a map, between which one can drive 'without thinking', and also about both of which one has a great deal of textual/factual information Readers of (Fillmore 1977) will recall his attempt to describe the relation of a text-based frame and an experientially-based scene to the same situation I think AI workers at this particular interface could profit from considering the extent to which such possible inconsistencies can be matters of theory rather than superficial fact an observer who is asked whether two sides of a long railway line meet at the furthest point he can see will give an answer not independent of of his abstract (possibly linguistically based) theory of parallel lines

In conclusion, this note has tried to do no more than ward off certain confusions, and to stress how many points of view are still open, since the evidence for and against them is no more than anecdotal, even when the anecdotes come from Psychology labs The choice between theses I/II/III is a metaphysical one, in the more red-blooded sense of that overtired word it cannot be made on empirical grounds now, but it can have important practical consequences about where one chooses to look for answers

References

Anderson, J. (1971) The Grammar of Case (London: Cambridge U.P.)

Bierwisch, M. (1970) "Semantics" in Lyons (ed.) New Horizons in Linguistics (London: Penguin)

Clark, E. & Clark,H. (1977) Psychology and Language (New York: Harcourt Brace)

Clowes, M.B. (1972) "Scene Analysis and Picture Grammars" in Nake & Rosenfeld (eds.) Graphic Languages, (Amsterdam: N. Holland).

Fillmore, C.J. (1977) "Scenes and Frames Semantics" in Zampolli (ed) Linguistic Structures Processing (Amsterdam: N. Holland)

Gregory, R. (1970) "The Grammar of Vision". The Listener. (London: BBC)

Jackendoff, R. (1975) "A system of semantic primitives" in Schank & Nash-Webber (eds.) Theoretical Issues in Natural Language Processing (Cambridge, Mass.: BBN)

Johnson-Laird, P. (1977) "Psycholinguistics without linguistics" in Sutherland (ed.) Tutorial Essays in Psychology (Hillsdale N.J.. Erlbaum)

Katz, J & Fodor, J. (1963) "The structure of a semantic theory". Language.

Leech, G. (1974) Semantics (London: Penguin)

Minsky, M. (1975) "Frame Systems" in Schank & Nash-Webber (eds.) Theoretical Issues in Natural Language Processing. (Cambridge, Mass.: BBN)

Rieger, C. (1976) Computers and Thought Lecture at IJCAI4, and published in Artificial Intelligence.

Schank, R. (1973) "Identification of Conceptualizations underlying Natural Language". in Schank & Colby (eds) Computer Models of Thought and Language. (San Francisco: Freeman)

Wilks, Y. (1972) Grammar, Meaning and the Machine Analysis of Language. (London & Boston: Routledge)

Wilks, Y. (1977) "Good and bad arguments for semantic primitives", Memo No.42, (Edinburgh: Dept. of Artificial Intelligence).

Wilks, Y. (1975) "Primitives and words" in Schank & Nash-Webber (eds) Theoretical Issues in Natural Language Processing. (Cambridge, Mass.: BBN)