AN EMPIRICAL INVESTIGATION OF VISUAL SALIENCE AND ITS ROLE IN TEXT GENERATION

E. Jeffrey Conklin *
Kate Ehrlich **
David D. McDonald ***

COINS TECHNICAL REPORT 83-14
November 1983

This work was supported in part by NSF
Grant # 1ST 810 4984

*Microelectronics and Computer Technology Corporation
9430 Research Blvd.
Austin, Texas 78759

**Honeywell Information Systems, Inc.
200 Smith Street
Waltham, Massachusetts 02154

***COINS/GRC
U. Mass.
Amherst, Massachusetts 01003

This paper is a reprint of the article by the same title which appeared in Cognition and Brain Theory, M.H. Ringle and M.A. Arbib (Eds.), Vol. VI, No. 2, Spring, 1983.

COGNITION AND BRAIN THEORY, 1983, 6(2), 197-225

An Empirical Investigation of Visual Salience and its Role in Text Generation

E. JEFFREY CONKLIN, KATE EHRLICH,
AND DAVID D. McDONALD
Department of Computer and Information Science
University of Massachusetts

ABSTRACT

An inherent part of visual perception is the derivation of information about the relative importance of each item in the image, a quality we refer to as the item's "visual salience." We have found that any reasonable natural language description of an image will reflect information about salience in its organization: highly salient items are described early and in some detail, while items of low salience are not even mentioned. In this article we report on two experiments that studied the notion of salience and its relationship to natural language text. The first experiment measured salience in color photographs of natural scenes and explored the factors that contribute to it. The second experiment explored how the organization of English descriptions of those scenes was related to the salience of the objects in them. The purpose of the experiments was to provide the basis for the construction of an Artificial Intelligence (AI) program that writes paragraph descriptions of natural outdoors scenes, using salience information in the visual representation as the basis of a direct and efficient text planning process.

INTRODUCTION

A central issue for any theory of the process of Natural Language Generation is the problem of *selection*: How to determine *what* to say, and (equally important) what *not* to say—what information about the subject matter do we include in our utterances and what information do we leave out? Put another way, the issue is how to determine what parts of the subject matter are most relevant given the discourse context and the speaker's goals. This

issue of the selection problem has been appreciated for a long time; it is the essence of Grice's conversational maxim to "be relevant" (Grice, 1975). However, progress on it has been difficult, in large part because it cannot be studied without opening the door to the little-understood realms of pragmatics, speech acts, and discourse theory.

In an attempt to make some concrete inroads into the problem of selection, we have been studying the task of generating natural language descriptions of pictures of visual scenes. We choose this task because scene descriptions, more than most other kinds of textual material, exhibit a natural and direct correspondence between the perceptual material presented in the picture image—its *meaning*—and the linguistic material in the English description. Indeed, the relative importance of the items in a picture—what we have called their *visual salience*—can be well defined, and can be matched to the relative rhetorical stress that the items receive in the text. The degree of match or mismatch can be used as an empirical guide to motivate and check the rules that one hypothesizes for item inclusion (or omission) and placement in the text.

In a series of experiments we gathered two kinds of data: (1) subject's ratings of the visual salience of the objects in pictures or suburban house scenes; and (2) brief paragraphs written by the subjects describing these same scenes. These studies were performed as a part of a larger effort to build an (AI) system that generated descriptive texts of its own (McDonald & Conklin, 1982). The studies served to provide us with an empirical measure of the salience of items in the scenes for us by the system, as well as to provide a substantial data base of textual material from which the specific rhetorical and grammatical rules by which the system was to operate could be derived.

THE ORGANIZATION OF A SCENE DESCRIPTION SYSTEM

So that the reader will have some idea of the context in which the experimental data was to be used, we briefly sketch the design of the AI scene description systems. The process of natural language generation is commonly divided into two pipelined stages (see Mann, Bates, Grosz, McDonald, McKeon, & Swartout, 1981 for a review). The first is responsible for deciding what information the text should contain and possibly its thematic structure and stylistic tone. We call this stage *deep generation*; it is the one that is responsible for the actual interface to the non-linguistic data base from which the information comes. The second stage, *realization,* is responsible for the actual production of the text given the first stage's specification. This stage is where the grammatical and morphological rules

of the language are impossed; it may also be responsible for low-level decisions about the best syntactic form to use (e.g., when to use pronouns).

In the present instance, the AI system consists of three components: (1) a simulated *perceptual representation of a picture,* structured as it would be by a computer vision system like the UMass VISIONS Systems (Hanson & Riseman, 1978; Parma, Hanson, & Riseman, 1980); (2) a *deep generation component,* the subject of Conklin's Ph.D. dissertation; and (3) a *realization component,* already developed by McDonald (McDonald, 1980).

The perceptual representation[1] is the input to the generation system. It simulates the kind of complete internal perceptual model that a successful computer vision system would construct, both for its own internal processing needs, and to represent its "understanding" of what the picture showed so that it could be used by other parts of the total cognitive system. Our version consists of a semantic net describing the objects in the scene and the relations between them in terms of a system of generic objects and relations, plus an annotation of the objects' salience as derived empirically from the experiments. This perceptual representation is expressly designed to be nonlinguistic: neither the structure of the paragraphs nor of the sentences produced by the generation system are pre-specified in the visual data base.

The deep generation component, GENARO (Conklin, 1983), contains the thematic and rhetorical knowledge of the system, embodied in a set of *rhetorical rules* that capture many of the conventions of English scene descriptions. GENARO is driven directly by the salience annotation that it finds in the perceptual data base, using the annotation to select what object to mention next, how much detail to use describing it, and when to stop the paragraph. It does no planning, in the sense of anticipating the costs of as yet unperformed actions, nor does it do any backtracking; instead the organization that might have been provided by the use of a general planning technique is "read out" directly from the annotated perceptual data base under the control of locally-operating rhetorical rules, with the result that the text can be generated in time linearly proportional to its length.

Basically, GENARO takes objects in order of decreasing salience (down to a threshold), builds a high-level specification for their textual description, and passes them sequentially to the realization component, MUMBLE [McDonald, 1980]. As each "rhetorical specification" is received by MUMBLE, an English rendering is selected for it in accordance with the contents of the specification and the linguistic context at the point in the paragraph where the realization takes place. So that general rules of discourse coherency and grammatical structure can be applied, MUMBLE

[1]The term "perceptual" is used deliberately here. We believe that there are meaningful auditory and tactile equivalents to visual salience, though nothing more will be said about them here.

constructs a full-scale linguistic surface structure tree rather than just a string of words. This tree is processed in one depth-first traversal, with the words of the text printed out as they are reached; the realization component thus also does its processing in linear time. The resulting text, whereas perhaps not Shakespeare, can be generated quickly (relative to systems that elect to do full-scale planning), and is of high enough quality that comparison with human-generated descriptions is quite favorable, demonstrating the power of salience as a heuristic in planning a descriptive paragraph.

SALIENCE

Although there has been considerable study of the processes of perceiving, recalling, and recognizing visual information (cf., [Loftus, 1982]), we have found little work directly addressing the issue of salience in pictures. Hooper (1980) studied the visual factors that contributed to the *recognition* of objects in scenes, but little of her study can be applied directly to the problem of determining what makes a particular object *salient*. Brush (1979) used a numerical rating technique similar to the one used here to study aesthetics of elements in pictures. The common sense experience of people who are "photographically literate" says that factors such as centrality in the picture, brightness, and color all can push an item into prominence (see also Arnheim, 1974); however, we were not aware of any sufficiently quantitative studies on relevant material that would have allowed us to proceed directly with the development of the AI system without performing the experiments we discuss here.

It is also apparent that the motive that a viewer has in examining a picture will affect how he or she sees it, at least in its details. Firschein and Fischler (1971) discussed the "problem of aboutness" in the context of "content analysis" during image abstraction of library pictures for storage and retrieval systems. They note that "what a document is about depends on what its reader will use it for," but they do not elaborate. Although we agree with this observation—a burglar and a real estate agent, for example, might give quite different descriptions of the same house—we also believe that it is unrealistic to attempt to model such subtleties given the limits of the present state of the art. We instead found that we were able to standardize this aspect of the perceptual process in our experiments (see Conklin, 1983 for a discussion on the effects of purpose on salience).

In looking at visual salience we found that it was useful to separate the properties of the *objects* pictured ("high-level properties") from properties of the *images* of the objects in the picture ("low-level properties"). Our use of the terms "high-level" and "low-level" is borrowed from work on com-

puter vision, in which low-level processing seeks to identify and label regions and boundaries in the image, while high-level processing seeks to match the low-level patterns with *knowledge about objects* in the world and how those objects appear in two-dimensional projections.

In addition to high-level versus low-level sources of salience, it is also important to distinguish between an item being salient just because it always is (e.g., UFO's, explosions, nudity, etc.), versus being salient because of the setting or context in which it occurs. We term this distinction *intrinsic* versus *extrinsic* salience: intrinsic salience is a fairly subjective reflection of how important the term is *in and of itself* to a person or culture (in house scenes, people are generally salient on this account); on the other hand, a bush in the middle of a highway is extrinsically salient—not because it is a bush (which has low intrinsic salience)—but because it is improbable and unexpected in its setting. This distinction applies largely to the high-level component of salience.[2]

EXPERIMENT 1

The experiments were designed with a number of specific questions in mind. In particular, we first wanted to examine how the visual salience of an object could be tied to the low-level features of the object's image. For example, below we describe how the size of an object, its centrality in the picture, perhaps even its brightness and color, all can push an item into prominence (see also Arnheim, 1974).

In addition, we wanted to study the effects of intrinsic salience: what happens to the distribution of salience in a scene when a person or other intrinsically salient item is added to the scene.[3]

Our third objective was to examine the relation between the salience of an item—as a quantitative measure—and its realization in a written description. A simple model of generation based purely on salience would predict that all and only the items up to some cut-off point should appear in the description. The model might also predict that the description is constructed so that the most salient item is mentioned first, then the next most salient

[2]To further clarify the distinction, if one took the photograph and carefully cut out the image of each object, the *intrinsic* elements of the salience of an object would be those that could be determined from examining the cut-out of that object *alone*; the *extrinsic* elements would be all others.

[3]Unfortunately, the behavior of the factor of *extrinsic* salience was not adequately studied in the initial experiments reported here, due to the fact that such studies would require pictures that contained strange and unexpected objects, properties, and relations. Such pictures generally require either careful staging, luck, or skillful darkroom manipulations. Our approach was instead to accumulate a firm foundation of data about "normal" pictures.

item, and so forth. This model is certainly simplistic: actual paragraphs obey neither a strict cut-off in salience nor a rigid ordering by decreasing salience. However, we will use this model as a starting point, if only as a yardstick against which to assess the degree to which salience alone is a useful organizing concept.

Subjects

The subjects in the experiment were 14 undergraduate and graduate students at the University of Massachusetts. All the subjects took part in the experiment voluntarily and were either paid $4.00 or were given credit for their participation in the experiment. The experimental session lasted 1 hour.

Method and Procedure

A group of subjects viewed a set of 25 pictures (as projected slides) and recorded their subjective evaluation of the relative importance of the items[4] in the picture. In our studies we used color slides of natural scenes, that is, urban, suburban, and rural scenes, and presented them in random order. All subjects saw all of the pictures, but similar versions of the pictures were separated by at least four other pictures. The scenes we used depicted little or no action.

The subjects were seated in a darkened room so that they were all roughly equidistant from the slide projection screen. The image on the screen subtended a visual angle of 15 × 12 degress, ±2 degress, depending on the viewer's position in the room. The slide projector was at its maximum brightness level, and the room illumination was made just bright enough (by opening shades at the back of the room) to make reading normal size print comfortable.

After being handed a booklet of "Item Rating Forms" (described later) the subjects were read the instructions, in which they were instructed to imagine that they worked in a library that had a large section of pictures, and that they were to rate the objects in the pictures on a zero to seven scale (where seven meant the object was the main item in the picture and zero meant the object did not even occur in the picture; they were told that their ratings would be used in cataloguing the pictures, so that they should try to be objective and consistent in their ratings. This was done by asking the subjects to " . . . imagine that you work for a large library which has a section

[4]We generally use the term *item* instead of *object,* because it occasionally happens that a non-object aspect of a picture (i.e., "snowing" or "out of focus") is salient. However, we did not include such items in our experiments.

containing only pictures of . . . real life scenes, and that your job is to rate the items in the picture by their importance. . . . The picture will be catalogued according to the values that you assign."

The entire set of slides was then presented. For each slide the subjects filled out one form (the form for that picture—see the following), and all subjects had the same form. The amount of time each slide was presented was variable, depending on the amount of time the group indicated it needed to finish filling out the form. This was done so that minimal time was spent on each slide, and yet all subjects had complete forms for all pictures. While no instruction was given against it, subjects rarely went back and changed any values. The time of presentation ranged from 1 to 4 minutes for each picture.

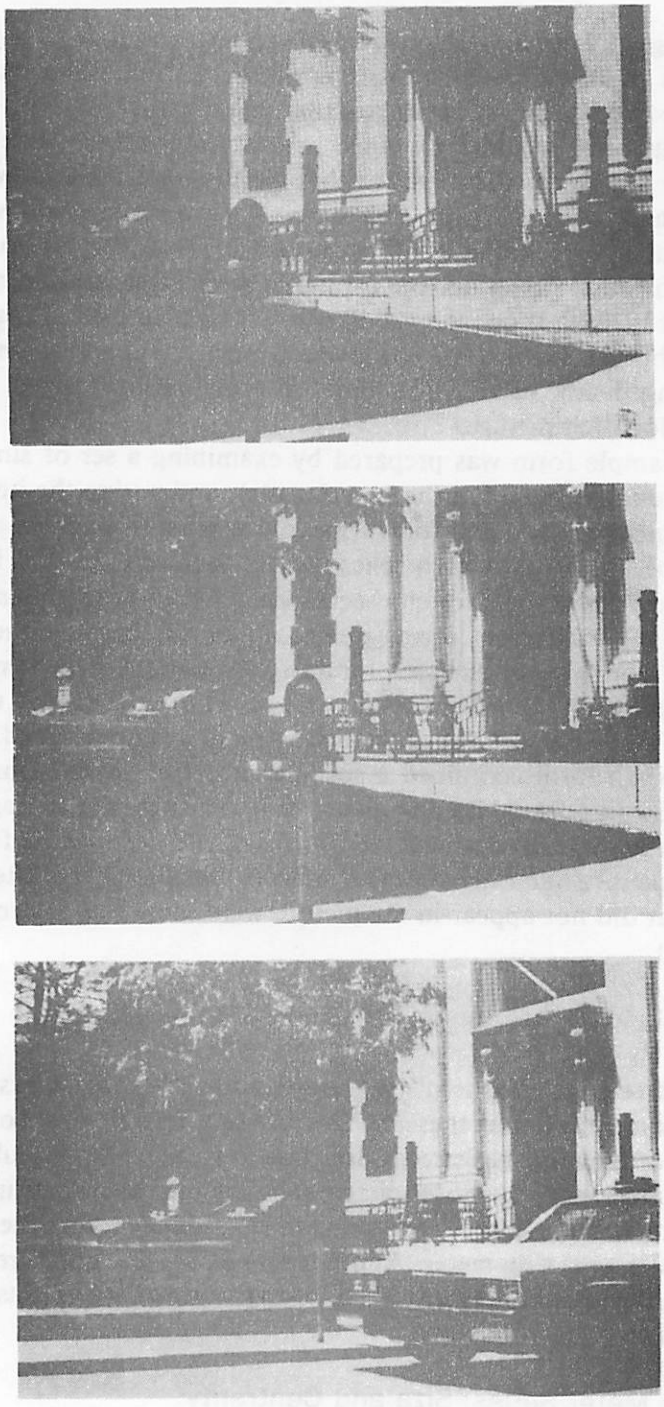
Each sample form was prepared by examining a set of similar pictures (the pictures to be used in the experiments), and noting the items that were shared among them. Each form appeared as a list of possible picture items, in alphabetical order, with a space next to each one for that item's rating. Thus, there was a form for city scenes, one for suburban house scenes, and so forth—five different forms were used in all. Because the forms were made up from a set of pictures, for any *given* picture there were items in the list ranging from the main object in the scene to objects that were not even present. There was a total of 48 items culled from our set of pictures, but because each form contained a specific subset of these items, only about half of them appeared on any given form. Further, the average picture actually contained only about three-fourths of the items on its form, thus the average picture had only about 18 objects that had to be "itemized." The items that did not appear in the picture were scored as a zero ("not in the picture").

Materials

Twenty-five photographs of city streets and residential areas were used in the experiment. Among these pictures are several series: sets of two or three pictures of the same basic scene designed to measure a particular effect. For example, one series had three pictures showing the same parking meter from slightly different distances and perspectives, thus varying the size and centrality of the parking meter. Another series contained pictures of the same house in the winter and summer, allowing study of the effects of foliage on the salience of picture items.

Parking Meter Series: Size and Centrality

This series of pictures (see Fig. 1) was of a street scene with a parking meter in the foreground and a large building and a lawn in the background. There were three versions of this scene:



a. Off-center
b. Center
c. Distant

FIG. 1 The parking meter series of pictures.

a. In one the parking meter is *large* and *off-center*; specifically, in this picture the image of the parking meter is 1.87% of the total picture area, and is located about 75% of the center-edge distance from the edge of the picture (i.e., 25% toward the edge from the center).

b. In the second the parking meter is *large* and *central*; that is, the parking meter image is 1.91% of the picture size, but is 93% of the center-edge distance in from the edge.

c. The parking meter is *central* but *small*; that is, the image is .43% of the picture, or less than one fourth of the size of the parking meters in the other two pictures, but is 90% of the center-edge distance in from the edge, almost the same location as in the second picture.[5]

We can evaluate the effect of the *size* of the parking meter on its rated importance in the picture by comparing the ratings for the parking meter in versions (b) and (c). Similarly, we can compare the assigned ratings in versions (a) and (b) to evaluate the effect of *centrality*.

Cottage Street Series: Occlusion and Intrinsic Salience

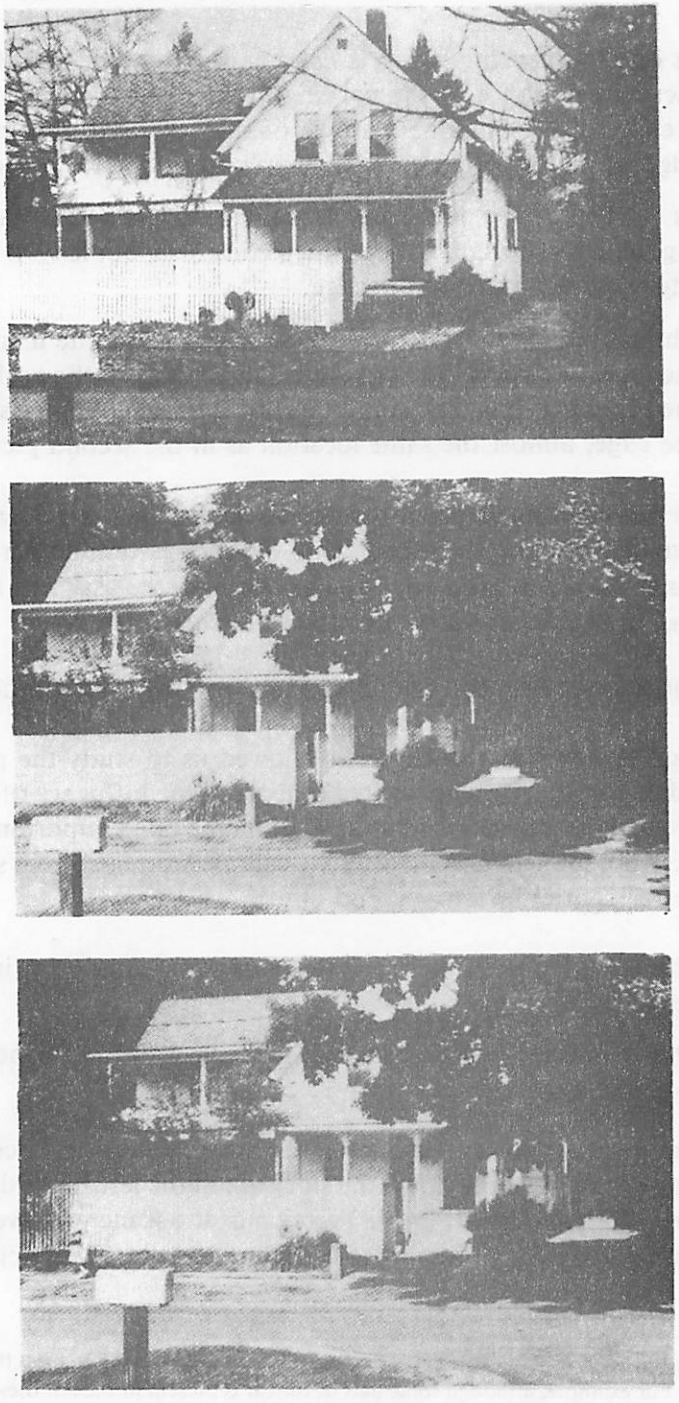
Although the Parking Meter series allowed us to study the parameters of size and centrality, we also wanted to explore the influence of other factors (such as intrinsic salience and occlusion) on the rated importance of items in a scene. Another series of pictures we call the Cottage Street series features a larger residential house (see Fig. 2):

a. The "Winter" version shows the house on a cloudy day in winter, with very little foliage;

b. The "Summer" version shows the identical scene six months later, in summer; and

c. The third version is the same as the summer version, except that there is a woman walking down the sidewalk in the left-hand side of the picture. Thus, in this "Person" version of the scene we have an item, the woman, that is not extrinsically salient, because she is neither in a cen-

[5]During the analysis it became apparent that this series of pictures was less than ideal for our purposes. For example, although some part of the car is present in each of these pictures, in the "Distant" version the car really emerges as a prominent "new" object in the scene, decreasing the similarity among the pictures. Ultimately, a much better scheme would have been to take three pictures of a scene without moving the camera, moving instead some object in the scene (like a bicycle) between foreground and background (thus varying size) and between central and non-central picture positions.



a. Winter
b. Summer
c. Woman

FIG. 2 The three pictures of the house.

tral portion of the picture nor particularly large in terms of the amount picture space she occupies. We were interested to learn if the woman's "*intrinsic*" salience was nonetheless detectable experimentally.

We also used versions (a) and (b) of the Cottage Street series to address the question: Is the part of salience that is based on image size dependent on the image size of the object *as if* you could see the whole thing, or just on the size of the visible (unoccluded) part of the object? The summer and winter versions of the Cottage Street house afforded an opportunity to answer this question because there is a large tree in the pictures, that in summer has a lot of leaves, occluding the house, and in winter is bare, exposing almost the whole house. Thus, in the summer version the image size of the tree is large and the house is small; in the winter version these are reversed, making it possible to compare just the salience of the tree and the house between the two pictures.

As with the Parking Meter series, this series of pictures was interspersed amongst the other pictures, that acted as filler items. There were always at least four other pictures between presentations of different versions of the same scene.

RESULTS AND DISCUSSION

The Parking Meter Series

The average ratings assigned to the parking meter are shown in Table 1. These data show that the parking meter was rated as more salient when it occupied a larger part of the picture (Wilcoxon Signed Rank Test: $T^* = 2.66$ (where "*" indicates the large group approximation), $N = 8$, $p < 0.0078$), and when it was more central in the picture (Wilcoxon Signed

TABLE 1
Statistics on the Parking Meter Series

Mean and Standard Error on Ratings of the Parking Meter

	Small & Central	Large & Central	Large & Off-center
Mean	5.4	6.4	5.5
S. E.	.51	.27	.48

(Ratings occurred on a 0 to 7 scale, with 7 being the highest salience, and 0 the lowest.)

Rank Test: $T^* = -1.87$, $N = 7$, $p<0.061$). Thus, not surprisingly, the rated importance of an item does seem to be influenced by features such as how central it is in the picture and how much space it occupies.[6]

The Cottage Street Series

We conducted a further test of the effect of size on salience by comparing the ratings given for the tree and the house in the Cottage Street series. In the summer version the tree had more leaves and therefore both occupied more picture space and occluded more of the house than in the winter version. The "size" of an object like the house (for purposes of its salience) could be:

* a function of the actual image area taken up by the object, or
* a function of the size that the object would appear to have if it were viewed directly (without occlusion or distortion).

In the second case the salience ratings of the tree and the house should have stayed the same between the winter and summer versions, because the tree is not actually bigger (simply "fuller"), and the house is not smaller (simply more occluded by the tree).

However, the data indicate that the former case is true—that salience is determined by the amount of the object that is visible in the image, and not how much is "really there." The mean salience rating of the tree in these two pictures went from 3.2 in the Winter scene to 4.4 in the Summer scene (Wilcoxon Signed Rank Test: $T^* = 2.215$, $N = 9$, $p<.027$), whereas the mean rating for the house dropped from 7.0 to 6.6 (Wilcoxon Signed Rank Test: $T^* = 0.0$, $N = 4$, $p<0.05$). No other objects were significantly different between the winter and summer scene.

Additionally, we asked: What is the effect of "adding" the woman to the summer house scene? Specifically, we suspected that people in general are intrinsically salient, so that even if the image of a person in a picture is not large or central it will be more salient than similarly placed images of more mundane objects. Table 2 shows the mean salience ratings for the most salient objects in the two scenes, along with the arithmetic difference between each item's pair of (averaged) ratings (one from each picture).

The first column in Table 2 presents the mean salience ratings of the objects listed in the picture without a person. The second column shows the corresponding ratings in the picture with the woman. Note that, as was

[6]Further studies will be necessary to determine the *sensitivity* of this technique: how small a change in size or centrality is detectable as significant with this method, and what other factors influence this sensitivity?

TABLE 2
Overall Effect of Adding a Person to a Scene

Item	−Person	+Person	Difference
House	6.6	6.0	−.6
Mailbox	5.6	5.3	−.3
Fence	4.9	5.1	+.2
Tree	4.4	3.7	−.7
Road	4.1	3.5	−.6
Person	−	3.4	−
Porch	4.3	3.3	−1.0
Door	3.1	2.9	−.2
Driveway	3.2	2.5	−.7
Roof	2.4	2.4	0.0
Gate	2.3	2.4	+.1
Sidewalk	1.5	2.4	+.9

predicted, the salience of the woman in this picture is distinctly higher than any other objects that are as small as remote in the image. The third column shows the differences between the first two columns. Two phenomena can be discerned in these data: the Normalization Effect, and the Locality Effect.

The Normalization Effect predicts that when a salient object is introduced into a scene the salience of the other objects drops. Thus, between the +Person and −Person versions there is a net drop of .26 in the average salience of all of the objects. Whereas this difference between the two pictures did not turn out to be significant in this case, such differences were generally observed in all of the picture pairs in which the primary difference was the presence (in one) of some salient object. We speculate that this effect is due to the fact that salience is relative, and thus that the scale "shifts" with the addition (or removal) of new salient objects. That is, an object with a rating of "6" means that the object is very important *in that specific picture*—any changes in the picture could shift that object's importance.

The Locality Effect specifies that when a salient object is introduced into a scene the salience of objects in the *image vicinity* of the new object increased as if the new salient object "pulled up" the salience of its neighboring objects.[7] In the picture, the objects near the woman did indeed show a

[7]The "*image vicinity*" is the two-dimensional area around an object's image.

relative increase over objects not in the woman's image vicinity. Moreover, this effect was found to be statistically significant (Wilcoxon Rank Sum, $W^* = -3.04$, $N = 14$, $p < .002$). The reader can get a sense of this effect simply by noting that the objects in Table 2 that increased in salience with the addition of the person are either near the person in the picture (the fence and the sidewalk) or are a part of these two objects (the gate is part of the fence).

The statistical demonstration of this effect was rather complex, and required two stages of analysis. In the first stage, the scores from *groups* of objects were treated as a single variable: the variables for objects not in the vicinity of the person (i.e., house, tree, porch, road, and driveway—refer to Fig. 2.c) were treated as a single variable, *Background*, by simply averaging the scores of those five objects. Similarly, the objects near the person (fence, sidewalk, and mailbox) were averaged into a single variable called *Vicinity*. When the two pictures were compared statistically in terms of this latter group of objects (the ones in the vicinity of the woman), the difference was not at all significant (Wilcoxon, $W^* = .71$, $N = 14$, $p < .48$). Interestingly, when the pictures were compared in terms of the Background objects, the difference was significant (Wilcoxon, $W^* = -2.23$, $N = 14$, $p < .026$)—we attribute the latter significance to the Normalization Effect: the Background objects showed a significant drop in their salience values.

The second stage of this statistical analysis is necessary because of the interaction between the Locality and Normalization Effects. That is, the Locality Effect predicts that the Vicinity variable will be *higher* in the picture with the woman in it, yet in fact there was no significant difference; on the other hand, the Normalization Effect predicts that both the Background and Vicinity variables will be *lower* in the picture with the woman in it. Thus, the impact of these two effects tend to cancel each other on the Vicinity variable.

Both the Normalization and Locality Effects were verified, however, when we tested for the significance of the *difference* between Background and Vicinity. In this case the difference between Vicinity and Background was significant at $p < .002$ (as mentioned above). In addition, the fact that this difference was the most significant arithmetic treatment of the data shows that there are two regions within the image, one around the new salient object and one covering everything else, whose saliences are moving in *opposite* directions. If the salience of the region proximal to the new object had not *increased* while the salience of the more distal region *decreased,* then the most significant treatment would not have been subtracting one region from the other—that this difference was the most significant treatment indicates that both the Normalization and Locality effects applied in this sequence.

This experiment was designed to examine whether changes in the size and centrality of an item would affect its *relative* importance; we predicted that changes in either of these dimensions would affect the rated importance of that item. Given that the results do indicate support for these predictions it is also of interest to question whether size and centrality might also predict the absolute importance of an item. The issue here is whether the largest or most central item in the scene is also rated as the most important. In a complex scene it is unlikely that any one factor will predict the rated importance of an item. However, by examining the predictiveness of one or two factors we can begin to assess the extent to which a simple analysis can account for the data. We examined the predictiveness of these two factors by determining which item was the largest in each of our 25 pictures and which item was the most central; in some pictures one item satisfied both criteria. We also determined which item in each of these pictures received the highest salience ratings from subjects. The result was that in 10 of the 25 pictures the most salient object was both the largest and most central, in three pictures the most salient objects was just the largest, in four pictures it was the most central, and in eight it was neither. However, in these eight cases the most salient item was an object that was *intrinsically* salient in the scene, such as a person (one picture), a flower in an otherwise barren scene (two pictures), or a car or tractor (four pictures).

The Object Naming Problem

One weakness of our methodology emerged in the process of analyzing the data. The technique required that subjects look at a form and note an object, and then look at the picture, locate the object, and decide on a 0–7 rating to assign to that object. The major weaknesses of this technique is that some objects have several names and some names denote several objects, or types of objects. Thus, it is difficult to measure the visual salience of some picture items simply because no good name exists for them in English. (For example, is the grassy area in front of some public buildings a "lawn," or a "yard"? Are the posts holding up the porch roof on suburban houses "columns," "posts," or "pillars"?) As a result, some items are characterized by having two peaks in the distribution of their salience values: one for the subjects who found an object corresponding to the item name, and one (at zero) for the subjects who considered the item name to denote an object not in the picture. This might be remedied by having subjects record the salience of items as they were *pointed to* in the picture by the experimenter (with or without the use of items names in the form), or by labeling the objects numerically in the picture itself.

Summary of Experiment 1

The results of this study indicate three factors that affect visual salience: the size of an item, its centrality, and the intrinsic interestingness of the item.[8] Thus, an item's salience should decrease as it shifts off-center, or as its image gets smaller in the picture. Also for a given size and centrality, items that are intrinsically interesting, or in the vicinity of such items, are perceived as being more salient. The study also provides some validation for our rating technique by demonstrating its sensitivity to these changes. Our main goal, however, is to determine the interaction between the perceptual phenomenon of salience and the way people generate descriptions of scenes. The aim of our next study was to examine the extent to which salience can be used to predict which items are included in a description and the order in which they are mentioned.

EXPERIMENT 2

In an informal pilot study (not detailed here), we asked people to write descriptions of a residential scene. We analyzed their descriptions in terms of the structure of the text and found that people often structured their descriptions around the central item, in this case the house. It appeared that the order in which objects were mentioned was guided by several *strategies*:

1. Mention and elaborate the most *salient* (unmentioned) object in the scene;
2. Mention and elaborate an object *related* to the last mentioned object;
3. Mention and elaborate an object *related* to some (viz. any) *salient* object in the scene (i.e., the house);
4. Mention objects in an order corresponding to a linear *sweep* through the picture (e.g., from left to right).

The first three of these strategies are similar in that they rely on *salience* and *relatedness* in their determination of the next object to describe. Strategy 1 uses only salience, and Strategy 2 uses only relatedness, and Strategy 3 uses both. Strategy 4 uses neither, and might be used in a scene with little organization (i.e., a Jackson Pollock painting.)

Our next study was designed to measure the extent to which the salience-related strategies (Strategies 1 and 3) were used, that is, to examine the ex-

[8]Based on the materials used in this study, neither color nor brightness had a measurable impact on salience.

tent to which salience might predict the structure and content of descriptions. Having an answer to this question would shed light on the extent to which the relatedness factor, and any other possible factors, determine the structure and content of a descriptive paragraph. (However, Strategies 2 and 4 were not studied in any detail in this study.)

Our methodology was to obtain both rating measures and descriptions of each picture for each person. We selected a set of nine pictures, culled from the 25 used in Experiment 1. In this experiment the pictures were shown twice: in the first viewing, half of the subjects did the rating task and the other half wrote descriptions of the scene; in the second viewing the tasks were reversed. Because the rating task used in Experiment 1 was used again here, this study also served to replicate the previous one.

Subjects

Thirty graduate and undergraduate students took part in the experiment. None of them had participated in the first experiment. They were either paid $5 an hour for the session, which lasted just under 2 hours, or were given "experimental credit" towards psychology courses they were taking.

Method and Procedure

Subjects were given two tasks; the rating task used in the previous experiment and a written description task, that went as follows. Subjects were read instructions and given forms for writing one paragraph descriptions of the pictures. They were asked to imagine that they worked for a picture library, as on the rating task, and to "write concise and accurate descriptions of the pictures so that they can be catalogued by their written descriptions." An example of the kind of description that got catalogued was included in the instructions.

The subjects were randomly divided into two groups: a Describe-first group, which did the description task first and the ranking task second, and a Ranking-first group, that did the tasks in the opposite order. After receiving their instructions, both groups (together) were shown the entire set of slides in random order.[9] Then, after a 15 minute break, the tasks were given to the opposite groups (i.e., the ranking-first group was given description instructions and forms and vice-versa), and the series of nine pictures was shown again (with the slides in a different random order).

---

[9]Actually, in both cases the order of the slides was non-random in the following ways: 1) the same "throw-away" scene was used as the first slide in both, as a training slide; 2) slides from the same series were not allowed to come together consecutively; and 3) slides from the same series were shown in different order the second time.

Because half the subjects did the ranking task first and the other half did the describing task first, we controlled for and were able to measure any effect that the additional exposure in the description task might have had on the ranking scores.

Each picture was displayed long enough for the people doing the describing task to finish. In all other respects this procedure was the same as that used for Experiment 1.

Materials

Nine pictures were selected from the set of 25 used in the first experiments: the three Parking Meter pictures, the three Cottage Street pictures, and three pictures of outdoor suburban scenes that were selected as fillers and warm-up items. We restricted the number of pictures for this study because we wanted to focus our attention on those pictures that had elicited significant effects in the first experiment. That is, we were interested in whether the effects found in the earlier study would be replicated and, more importantly, we wanted to examine the interactions between the ratings accorded to items and also how and when these items appeared in the descriptions.

Results and Discussion

We first discuss the results from the rating task and then turn to the results from the description task.

In addition to replicating Experiment 1, the rating data gathered in this study offered the chance to measure the effect of familiarity with the picture on the salience ratings assigned.

THE RATING TASK—REPLICATING EXPERIMENT 1

Parking Meter Series

We computed the average ratings for the parking meter series separately for the group of subjects who did the rating task first and for the subjects who did the description task first. As shown in Table 3, the data from the subjects who did the rating first replicated the results of the first experiment. That is, the first row (which shows rating-task-first results) indicates that the size difference in the parking meter is significant at well below 1% ($p < .0007$) and the centrality difference is significant at below 2% ($p < .016$).

The second row, which shows statistics on the rating scores of subjects who did the rating task *after* the description task, indicates that the process

of studying these pictures enough to write descriptions of them seemed somehow to overwhelm the size and centrality effects otherwise observable. (The third row simply shows the effect of combining the two groups of scores.)

A closer inspection of Table 3 reveals another difference between the Rating-first and Description-first groups. The rating assigned to the large central parking meter was substantially lower for the Description-first group (i.e., 5.1) as compared with the Rating-first group (i.e., 6.5). The ratings in the other two conditions remained the same. We can only guess as to the source of this discrepancy between the two groups. One additional fact may be relevant: the Rating-first group gave the Car, the Parking Meter, and the Building the highest ratings in the three pictures, respectively, while the Description-first group rated *Building* as the main object in all three pictures (hence, the low salience of the parking meter for this group). Perhaps having already studied all of the pictures, the subjects who did the rating task second moved away from basing salience on the somewhat anomalous and troubling extrinsic features of size and centrality and gave more weight to the intrinsic salience of the building. That is, by having seen the whole series of pictures before performing the rating task, the description-first group may have focused more strongly on the *theme* of buildings and houses in the pictures.

TABLE 3
Statistics on the Parking Meter Series

Means and Wilcoxon Rank Sum Significances of the Parking Meter in Experiment 2.

	Small & Central	Wilcoxon Score	Large & Central	Wilcoxon Score	Large & Off-center
First Task					
Rating (N = 15)	4.8	.0007 W = 313.5	6.5	.016 W = 175.	5.6
Descr. (N = 15)	4.2	.048 W = 280.	5.1	.75 W = 240.	5.3
Combined (N = 30)	4.5	.0002 W = 1165.5	5.8	.17 W = 822.5	5.4

Each Wilcoxon score indicates the significance of the difference between the means on either side of that score, for example, in the first row the difference between the scores on the Small & Central and Large & Central pictures was significant at $p < .0007$. The ratings were on a 0 to 7 scale, as before. (Note that the simple *unpaired* version of the Wilcoxon test was used here, since the more powerful paired version was not necessary to get significance less than 5%.)

Nevertheless, the fact that we did replicate our earlier results for the subjects who did the rating task first provides additional support for the claim that shifts in size and centrality do affect the rated importance of an item, and that our technique is reasonably sensitive to such shifts.

Cottage Street Series

In the previous experiment the data suggested that there was a *Normalization Effect* and a *Locality Effect*. When we did the same analysis on the rating data in the present experiment we found that the Normalization effect continued to be a factor in all pairs of pictures, seeming to adjust the salience values to maintain the same "total salience" for any picture. However, the Locality effect was considerably diminished, suggesting that it is a much less robust effect. Specifically, the significance of the Locality Effect, previously $p<.002$, rose to $p<.13$ in this study (using only the Rating-first data).

Description Task

The data and analysis from the description task had the greatest impact on the design of the GENARO generation system. Indeed, one of the fundamental problems facing any account of natural language generation is the issue of how the content of an utterance is determined in the first place. What should one say, and what should one leave out? The GENARO system shows that this problem of Selection is powerfully addressed by the notion of salience.

Specifically, there are two empirical questions that our study addressed. One is whether or not an object's being mentioned at all in a textual description is a function of its visual salience. Let us call this claim Hypothesis 1. Another issue is whether the order in which objects are mentioned in textual descriptions is primarily determined by the salience of those objects. Let us call this claim Hypothesis 2. This latter claims turns out to be too strong. As mentioned earlier, people use both salience and relatedness in their selection strategies.

To illustrate, we will examine rather closely the correspondence between a particular subject's salience ratings on the picture of the house in summer (Fig. 2.b), and that subject's written description of the same picture. Subject S1 produced the ratings shown in Table 4, *and* wrote the description shown in Fig. 3.

The description illustrates a number of points about this kind of data. First, the underlined terms are the *first mention* of only those items that were also included in the set of rated objects. Because this set included the generic object "tree," the second occurrence of the word "tree" in the

TABLE 4
S1's Ratings on the Summer House Scene

House	7
Fence	6
Tree	5
Driveway	4
Mailbox	3
Road	3
Sidewalk	3
Wires	3
Yard	3

This subjects' ratings ranged from 7 ("Main object in picture") to 3 ("Less than average importance") on this picture.

description actually refers to a different tree than the first occurrence (the"tree in the yard" vs. the "tree shading the driveway"), but unfortunately there was no provision for indexing and differentiating several objects of the same type in our study. Furthermore, because only objects that were in both the rating set and the written descriptions could be used for comparison, items like "power cables" in sentence (vi) had to be "translated" into "wires," and items like "beam" (in sentence (vi)) and "afternoon" (in sentence (vii)) had to be omitted altogether.

Nonetheless, Table 4 and Fig. 3 illustrate the kind of correspondence that was found between rating scores and descriptions. The first object mentioned got the highest rating, the second object got the second highest rating, and the third object (Yard) got the *lowest* rating. Our claim is that

i) This is a picture of a large white wooden *house*.
ii) In front of the house is a white *fence*.
iii) In the *yard* is a *tree*.
iv) Next to the house is a *driveway*, which is mostly shaded by a large tree.
v) In front of the house is a *street* and *sidewalk*.
vi) Across the top of the picture are *power cables*, and in the lower left is a white *mailbox* on a brown beam.
vii) It is late afternoon.

(The numbers to the left of each line were added by us. The italicized objects are those that also received values in the rating task.)

FIG. 3 S1's description of the Summer House Scene.

Yard is mentioned at this point by virtue, not of its salience, but of its *relationship* to a salient object, Tree. Note the rhetorical value of "pulling in" the related object Yard to the description of the Tree—the simpler sentence "There is a tree" is unacceptably short and plain.

Hypothesis 1

To test whether the visual salience of an object predicted whether or not it was mentioned in the text description (Hypothesis 1) we calculated the total number of items to which each of the seven ratings was assigned, over all subjects and all pictures. We then calculated the number of those items that were also mentioned in their corresponding description. These data are shown below in Table 5.

The data indicate a close correlation[10] between the rating and the probability of inclusion in the description, such that the higher the rating an object received the higher probability of that item appearing in the description. Hence, Hypothesis 1 was supported by our data.

Hypothesis 2

Having established the link between an object's salience and whether or not it is mentioned, the question became: "What is the correlation between the salience rating of an object and its *point of occurrence* in the description?"

In studying the written paragraphs from this experiment, we found that there was considerable variation in the descriptions in terms of length, order of items, and overall structure. There were also rhetorical factors operating in the generation of descriptions, in addition to the predicted salience factors. These rhetorical factors would for instance lead someone to mention an item that was not particularly salient but which had value simply on rhetorical grounds, as in the example of the Yard and Tree above. As another example, in describing the summer version of the Cottage Street house, one subject wrote:

> This picture contains a white, two-story house behind a white wooden fence and partially obscured by a tree. The house appears to be in the form of an "L" . . .

[10]The fact that there was almost no difference between the objects assigned a 3 (i.e., 0.26) and those assigned at 2 (i.e., 0.24) suggests that perhaps the phrases for these ratings, "less than average importance" and "very low importance," respectively, meant essentially the same thing to subjects. Alternatively, the distinction between the lower salience ratings simply was not as striking to subjects as that between the higher ratings.

TABLE 5
Salience as a Predictor of Mention

Salience Rating	No. of Objects w/ that Rating	No. of Objects in Description	Probability of Mention
7	72	71	.99
6	95	84	.88
5	180	118	.65
4	262	111	.42
3	155	40	.26
2	155	37	.24
1	131	15	.11

Although neither the "whiteness" of the house nor its "L shape" were particularly salient, this subject chose to mention both as part of the *elaboration* of the main picture item, the house. Such elaboration is a rhetorical convention that sometimes runs counter to the demands of salience.

However, taking Hypothesis 2 literally, if salience is the sole determinant of when an object gets mentioned in the description, then there should be a perfect correlation between the ratings received by objects and their point of mention.

A major difficulty in testing this hypothesis lies in quantifying the point in a descriptive paragraph where an object is mentioned. The simplest approach is to *rank* the objects by their order of mention. That is, the first object mentioned, House, is ranked 1, the second, Fence, is ranked 2, and so on. Although this is a coarse treatment of the subtleties of a natural language test, we used it as an initial way of testing our hypothesis.

After ranking the rating data[11] we had the ranking scores shown in Table 6.

The table shows (as discussed earlier) that when viewing this picture S1 ranked the House first both in terms of ratings and description, and likewise the Fence second in both. The statistical correlation between these two sets of ranks can be measured by the Spearman ranked correlation coefficient. In the case of the ranks in Table 6 the Spearman coefficient is 0.8, indicating quite a high correlation.

[11]The use of nonparametric statistics demands that all of the data be converted into ranked scores. For the rating data this has the effect of "inverting" the scores: the object which is rated as 7 gets ranked as 1, etc.

TABLE 6
Ranks of S1's Rating and Description Data

Object	Rating Rank	Description Rank
House	1	1
Fence	2	2
Tree	3	4
Driveway	4	5
Mailbox	7	9
Road	7	6
Sidewalk	7	7
Wires	7	8
Yard	7	3

This table summarizes the data in Table 4 and Fig. 3. The description ranks shown are by simple position in the paragraph. (Note also that when there was more than one object with the same rating score (e.g., Mailbox, Road, etc.) then all objects were assigned the same rank, that rank being the average that the objects would have received.)

However, when this test was performed on the entire set of rating/description ranking pairs (i.e., for each subject and each picture), the average Spearman coefficient was 0.52[12]. Because this is not a high correlation coefficient, either our method of ranking objects in descriptions is weak or Hypothesis 2 is too strong.

The problem with the object ranking scheme described here is that it ignores sentence boundaries—it makes the implicit claim that related objects being mentioned in the same sentence are the same "semantic distance" apart as related objects in different sentences. Similarly, it ignores the fact that there are several simple syntactic transformations that reorder the noun phrases in a sentence without changing the semantics. If the third sentence in Fig. 3 had been "There is a *tree* in the *yard*." instead of "In the *yard* is a *tree*.", then by the scheme used above the ranks of Yard and Tree would have been switched without reflecting any real difference in their salience.

Another possible scheme for ranking objects is to assign all objects in the same sentence the same rank, that is, ranking by sentence. This scheme has

[12]This correlation is based just on the rating-first subjects. The correlations for the description-first subjects was .31, and for both groups combined was .42.

the disadvantage of forcing all objects in a sentence to have equal rank, with a relatively big jump (a jump of one) to objects in the next sentence. When this scheme is applied to S1's paragraph (in Fig. 3), the rankings come out as shown in Table 7.

The Spearman correlation coefficient for this ranking set is 0.85, a small improvement (by 0.05) over the correlation based on the other object ranking scheme.

When the by-sentence ranking scheme is applied to *all* the data the average Spearman correlation coefficient does not change appreciably. This indicates that the within-sentence syntactic variation was not reducing the correlation with salience, and it also shows that the low correlation is not due to the object ranking method used. Instead, Hypothesis 2, that the salience of an object is the sole determinant of its position in the descriptive text, must be too strong. Evidently (and intuitively) there are other strategies employed in the process of selecting what to say next than Strategy 1 (i.e., "Say the next most salient thing").

However, only Strategy 1 can *begin* a description—the other strategies (listed in the beginning of this section) must be "seeded" with at least a *first* item. (It almost always happened that in each picture the object with the highest salience rating was mentioned in the first sentence of the description.) We will regard Strategy 1, therefore, as the primary strategy. To what extent do the other strategies account for the correlation-reducing "noise"?

TABLE 7
S1's Rankings—By Sentence

Object	Rating Rank	Description Rank (by sentence)
House	1	1
Fence	2	2
Tree	3	3.5
Driveway	4	5
Mailbox	7	8.5
Road	7	6.6
Sidewalk	7	6.5
Wires	7	8.5
Yard	7	3.5

Note that, as in Table 6, several objects with the same rating score (column 2), or several objects in the same sentence (column 3), all were assigned the same rank.

Hypothesis 3

Setting aside Strategies 3 and 4 for a moment, let us consider what Strategy 2 means operationally. A description produced using only Strategy 2 (after the first item was selected) would simply chain objects along, relating each object to the previous one, without any regard for the salience of each object. Adding Strategy 2 to Strategy 1, then, might be the enhancement needed to account for the data. Let us call this *Hypothesis 3*: Strategy 1 decides the first (most salient) object and, thereafter, Strategies 1 and 2 compete to decide each item, each time resolving the tension between salience and relatedness. In a description produced according to this hypothesis, one would expect to find a descripton ordering (i.e., the list of objects in the order in which they were first mentioned in the description) that was derived from the salience ordering by the occasional "movement" of objects from arbitrarily far down in the salience ordering to relatively early in the description (e.g., the way Yard moved from the bottom of the salience ordering to nearly the top of the description ordering). Hence, to the extent that the description ordering differed from the salience ordering it would be at points where a strongly related object got "pulled up" from lower down in the salience ordering. This is in contrast, for example, to high salience objects getting pulled arbitrarily far down into the description ordering.

If this asymmetry exists in the data it should be measurable. The Spearman ranked correlation coefficient is calculated by finding the average "distance" that items have "moved" in going from a first list to a second one. However, this test is *symmetrical*—two lists that correlate perfectly suffer the same loss of correlation whether an item is moved up or down in the second list. We designed a variation on the Spearman test that is asymmetrical: an item that moves *down* between the first and second list decreases the correlation much more than an item that moves *up*. Taking the salience ranking as the first list and the description ranking as the second, this corresponds to decreasing the correlation greatly if a high salience object appeared late in the description, and decreasing the correlation very little if a low salience object appeared early in the description.

Applied to S1's rating and description data in Table 6 this asymmetrical measure increases the correlation from .80 to .88, using the first description ranking method; using the "by-sentence" method, the correlation increased from .85 to .90. (The increase over the Spearman test result is because the only difference between the two rankings that matters is the appearance of the low salience Yard early in the description, and the asymmetrical test penalizes this difference less than the Spearman test does.)

Applied to all of the data, however, the asymmetrical test failed to show any appreciable increase in correlation between subjects' salience rankings

and their descriptions. The problem with the asymmetrical statistic is that it tests Hypothesis 3 in only a very narrow way: it assumes that only a few low salience items will get pulled up into early text positions. If many such items get pulled up, high salience items are necessarily pushed *down* to make room for them, and the asymmetrical test heavily penalizes the downward movement of these high salience items. This limitation illustrates a more fundamental problem: whereas it is easy to see that Yard was pulled up for rhetorical reasons in Fig. 3, in many actual texts the relationships between objects are so complex that relating the order of objects to object salience is very difficult.

Hypothesis 3, then, although more powerful than Hypothesis 2, still only accounts for about 60% of the order in which objects were mentioned by subjects in this study. Strategy 3, which uses both salience and relatedness, may account for some of the remaining 40%; or Strategy 4, or even some other strategies may come into play. In addition, it is likely that some of the variation we found was a function of personal taste, and is therefore inaccessible to merely more complex statistically-based theories.

Summary

We have shown that a moderately good correlation exists between the salience of an object and its position in a descriptive test (both absolute position and by sentence), thus lending support to our claim that a combination of salience and rhetorical factors is used in the process of selecting what object to describe next. A simple model of the generation process, as expressed in Hypothesis 3, failed to completely account for the data, indicating a complexity or rhetorical structure that is not captured by the simple notion of low salience items being "pulled up" to early text positions.

Our AI system (Conklin, 1983) is basically styled after Hypothesis 3—it uses various production rules, some that are driven by salience (capturing Strategy 1), some by relatedness (capturing Strategy 3), and some by a combination of these (capturing Strategy 2). The system produces scene descriptions that are quite good, though they are not very syntactically complex, neither do they appear to be lacking any crucial rhetorical or thematic factors.

CONCLUSIONS

As preparation for the building of an AI system that writes scene descriptions, we sought to better understand how people perform this task. In the process we discovered that the commonsense notion that "some things are more important than others" was critical to the organization of visual data

for effective linguistic presentation—without this notion of "salience," a description would at best be a stylistically-pleasant hodge-podge of items in the scene. What is more, it appears that even without the demands of linguistic processing a perceptual data base is incomplete without some labeling of relative salience in the data base (because this is a natural byproduct of the process of perception). Most importantly, our AI system demonstrated that using salience as a heuristic in text generation could greatly expedite the planning of a paragraph at the rhetorical level.

This article has reported on a series of experiments designed to study visual salience, both as a perceptual phenomenon and as an organizing principle in natural language generation. We draw three major conclusions from these studies:

1. That the notion of visual salience, as defined here, is a significant perceptual phenomenon, and that it possesses an internal structure consisting of high- and low-level and intrinsic and extrinsic (context-dependent) components;
2. That it can be studied by the simple experimental techniques described in this article;
3. That salience organizes a perceptual data base in a way that allows greatly simplified rhetorical planning in natural language generation, by playing a powerful role in determining the order in which objects are mentioned in scene descriptions.

The data from these studies (both numerical and textual) are being used directly in the design and refinement of the GENARO/MUMBLE generation system: to annotate the input perceptual representation with realistic salience values, to guide the selection process during the planning of the descriptions, and to inform both the rhetorical and grammatical rules used by the system.

ACKNOWLEDGMENTS

Michael Arbib, Lyn Frazier, and Keith Rayner provided valuable comments on earlier drafts of this paper. We also have benefited greatly from conversations with Beverly Woolf, Jeff Bonar, and Steven Levitan. Finally, many thanks to Beach Clow and Piet Vermeer for their statistics and APL assistance. This article describes work done in the Department of Computer and Information Science at the University of Massachusetts. It was supported in part by National Science Foundation grant IST 8104984.

REFERENCES

Arnheim, R. *Art and visual perception: A psychology of the creative eye.* Berkeley: University of California Press, 1974.

Brush, R. The attractiveness of woodlands. *Fores! Science,* 1979, *25,* 495-506.

Conklin, J., & McDonald, D. Salience: The key to the selection problem in natural language generation. In the Proceedings of the Association for Computational Linguistics, Toronto, Canada, 1982.

Conklin, J. *Localized planning in discourse generation using salience,* Ph.D. thesis, Department of Computer and Information Science, University of Massachusetts, Amherst, Mass., 1983.

Firschein, O., & Fischler, M. A. Describing and abstracting pictorial structures. *Pattern Recognition,* 1971, *3,* 421-443.

Grice, H. P. Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts,* Vol. 3. New York: Academic Press, 1975.

Hanson, A. R., & Riseman, E. M. VISIONS: A computer system for interpreting scenes. In A. R. Hanson, & E. M. Riseman (Eds.), *Computer vision systems.* New York: Academic Press, 1978.

Hooper, K. Picture recognition: A consideration of representational media and realism. Unpublished paper, 1980.

Loftus, G. R. Models of picture recognition. In R. Wu, & S. Chipman (Eds.), *Learning by eye.* Unpublished manuscript.

Mann, W., Bates, M., Grosz, B., McDonald, D., McKeown, K., & Swartout, W. Text generation: The state of the art and the literature. Information Sciences Institute technical report RR-81-101, Marina del Rey, California, 1981.

McDonald, D. D. *Language production as a process of decision making under constraints,* M.I.T. Doctoral Dissertation, 1980.

McDonald, D. D., & Conklin, J. Salience as a simplifying metaphor for natural language generation. In the Proceedings of the Annual Conference of the American Association of Artificial Intelligence, 1982.

Parma, C. C., Hanson, A. R., & Riseman, E. M. Experiments in schema-driven interpretation of a natural scene. In J. C. Simon, & R. M. Haralick (Eds.), *Digital image processing.* Dordrecht: D. Reidel, 1980.