20

Pictorial Relationships - a Syntactic Approach

M. B. Clowes
Division of Computing Research,
C.S.I.R.O., Canberra

1. INTRODUCTION

Grammars or syntax specifications address themselves to the characterisation in symbolic terms of the structure of complex expressions. Two types of expression of empirical interest have been studied: sentences in English and other 'natural' languages, and programs written in some high-level procedural language like ALGOL. Expressions in these languages consist of sets of elements (words and characters) co-ordinated with one another according to the *sensorily* manifest relationship 'alongside', more commonly termed 'followed by'. (In the case of context-sensitive phrase-structure grammars it may also be 'on both sides of', more commonly termed 'bounded by' or 'in the context of'.)

A grammar seeks to relate, by translation or mapping, this manifestation of the expression into another in which the same elements together with others (e.g. 'Noun Phrase', 'Simple arithmetic expression', etc.) are co-ordinated by *abstract* relationships which in the case of English and ALGOL is the single relationship 'parts of'. The notation used to exhibit this relationship is some tree-structure representation in which elements are associated with (i.e., label) the nodes of the tree.

A syntactically motivated parser is a device which *accesses* elements of the sensorily manifest expression, by application of an addressing procedure, e.g. '*Next char*' which embodies their sensorily manifest relationship. The parser *develops* a representation in which elements are co-ordinated by the *abstract* relationship ('parts of'), through application of an addressing procedure, e.g. '*Tl*' or '*Cdr*' (Woodward 1966) which *embodies* this abstract relationship.

In a free paraphrase of Chomsky we might say that a parser translates some set of *surface* relationships on elements, into some set of *underlying* relationships on those (and other) elements. Thus, according to Chomsky

(1965, p. 171), the functional relationship 'Subject of' is to be understood as the relation of 'dominated by' obtaining between a Noun Phrase node and the S(entence) node which immediately dominates it. Needless to say this cannot be translated simply into the surface relationship 'followed by' – hence Transformational Grammar.

The foregoing differs from 'traditional' accounts of generative grammar in according a central rôle to the *relationships* manifest in the surface and the underlying representation of an expression rather than focusing attention upon the categories of element (e.g. *NP*, *s.a.e.*, *N*, *var*) and the properties or features which they might be thought to possess. The need to reformulate the account of generative grammar in this way appears essential to the characterisation of expressions of a non-linguistic kind and in particular to pictorial expressions (Clowes 1968).

There have been several attempts to write generative grammars for pictorial expressions notably those of Kirsch (1964), Narasimhan (1966) and Shaw (1967).

In the case of Kirsch's 'triangle grammar' – a set of rules which express addressing (and labelling) operations upon a two-dimensional array of positions – we observe the implicit identification of the surface relationships 'to the left of', 'below', etc., predicated on positions as elements. The 'grammar' specifies no other kind of relationship. Specifically it fails to exhibit as *parts* of the triangle whose structure it purports to describe, the 'edges' of the triangle.

Both Narasimhan and Shaw utilise surface elements consisting of two or more distinct and distinguished positions (called Head, Tail by Shaw and vertices numbered 1, 2, 3 by Narasimhan). Shaw defines these elements by providing co-ordinate values for Head, Tail; thus we may compute the relationship between them, e.g. length or relative position, but this relationship is not *structurally* exhibited. Both authors predicate the surface relationship 'coincidence' on pairs of positions belonging to two different elements. The grammars both assign the underlying relationship 'parts of'. Furthermore on this view of syntax it becomes clear that Minsky's (1961) picture language is one in which a wide *variety* of named relations, e.g. *ABOVE, LEFT, INSIDE* is assumed. Constraints such as 'parallel' in Sketchpad (Sutherland 1963) would appear to have a similar rôle. An immediate distinction between string expressions (English sentences, say) and pictorial expressions now emerges.

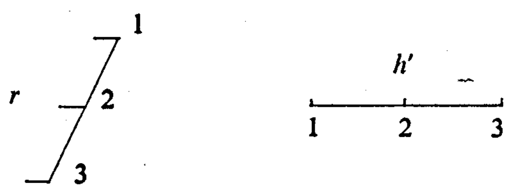
The variety of relationships which we can readily identify and name is much greater in pictorial expressions than in string expressions. We should note that one way to look at Chomsky's 'Aspects of the Theory of Syntax' is an attempt to account for, among other things, the *grammatical relationships* Subject-Verb Verb-Object (pp. 64, 73). This attempt fails in the author's view precisely because of the failure to reformulate the purpose of generative grammar in relational terms. It seems likely that other problems in generative grammar might benefit from this reformulation, when this distinction between string and pictorial expressions would lose much of its force. It would remain true, however, that linguistic relationships are harder to *identify and name*. This is the essential fact that makes it necessary – if we are to adopt any of the methods of generative grammar – to reformulate syntax as *having to do with relationships rather than with what is related.*

Given that the foregoing analysis is true, it follows that we will have to make provision in the metalanguage (in which we will couch picture grammars) for the overt characterisation of (possibly) large numbers of distinct relationships. Given such a metalanguage, the empirical task becomes that of providing formal definitions of just those relationships as do mediate our grasp of the structure of pictorial expressions.

2. THE METALANGUAGE

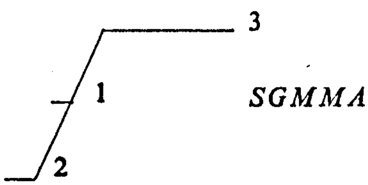
As we have noted, both Narasimhan and Shaw are concerned in their notations to exhibit a specific relationship – that of 'joined' or 'connected'. Thus in Narasimhan's notation the primitives r and h'

(1)



may be considered to be parts of *SGMMA*.

(2)



according to the composition rule (3).

(3) $\quad$ *SGGMA* (1, 2, 3)$\rightarrow r \,.\, h'$ (1 1:2, 3; 3)

We read this as stating that r and h' are 'joined' at positions or 'vertices' designated as 1 of r and 1 of h'. Furthermore that three positions or 'vertices' on *SGMMA* are to be identified with positions 2, 3 of r and 3 of h'. The descriptions of r and h' – as each having distinguished positions 1, 2, 3, – implied by (3) is made pictorially explicit in (1). We must imagine of course that in any formal procedural account of this notation, this informal pictorial characterisation would be replaced by a specification of r and h' in which the positions 1, 2, 3 would be given co-ordinate values. Thus we might replace (1) by

(4a) $\quad r(1(x,y),\ 2(x+p,y+p),\ 3(x+2p,y+2p))$

(4b) $\quad h'(1(x,y),\ 2(x+d,y),\ 3(x+2d,y)),$

where (x,y) are of course variables assuming different values in each of the three sets of parentheses – that is, h' might have the literal form $h'(1(x,y), 2(x+1,y), 3(x+2,y))$.

The intention of (3) is that it is these co-ordinate values of designated vertices of r and h' which should be 'transferred' to designated vertices of *SGMMA*, rather than the vertices themselves. Similarly that it is an equality of the co-ordinate values associated with 1 of h' of r which underlies this 'join' relationship between these two primitive parts of *SGMMA*. In other words, several pictorial relationships, 'join', 'coincidence', 'same position as' are assumed in our reading of (3). A syntax should provide a formal description of these assumed conventions insofar as they reflect pictorial intuitions. We can rewrite (3) in a form which makes these assumptions explicit and *names* the varieties of relationship involved.

$$(5)\quad \mathit{join}\,\langle r_1(1_2(x_3,y_4),\ 2_5(x_6,y_7),\ 3_8(x_9,y_{10})),\ h'_{11}(1_{12}(x_{13},y_{14}),\ 2_{15}(x_{16},y_{17}),\ 3_{18}(x_{19},y_{20}))\rangle\ [\mathit{Coinc}\,\langle 2,12\rangle]$$
$$\Rightarrow \mathit{SGMMA}(1(6,7),\ 2(9,10),\ 3(19,20))$$

This states that *SGMMA* is formed from two parts r and h' – enclosed within angle brackets. The relationship which these two parts enjoy in order that they be 'capable' of forming *SGMMA* is *join*, which entails a further relationship namely of coincidence (*Coinc*) between designated elements of the *descriptions* of r and h'. The inferior or (suffixed) integers used in (5) merely provide an explicit referencing mechanism to replace the implicit ordering convention of (3). This permits us to state the 'same position as' requirement in respect of the distinguished positions of *SGMMA* and those of r and h'.

The left-hand side of (5) is descriptive of the structure of *SGMMA* in exhibiting its parts and specifying the relationship between them. On the right-hand side a further description of *SGMMA* is provided which does not explicitly state the relationships between the elements (1, 2, 3) which comprise it. The same is true of the descriptions of r, and h' on the left-hand side of (3).

A further difference between (3) and (5) is the use of, and the direction of the $\Rightarrow$. What we have in mind here is that in discerning that *SGMMA* is 'made up of two parts' we are recovering relationships, specifically *join*, *Coinc*, on these parts, thereby assigning an underlying structure to *SGMMA*. We identify this process with *parsing*. Accordingly, regarding (5) as a rule of grammar, the arrow points in the generative direction, i.e., towards the surface form of the pictorial expression.

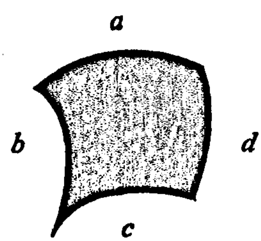
This account of (5) identifies it with a rule of transformational grammar (a T-rule) and we may note a fairly consistent correspondence between the syntactic structure of (5) and the syntactic structure of the *generalised transformation* (Chomsky 1965). Specifically, *join* is the name of the transformation; the pair of descriptions enclosed in angle brackets, are *SD*1, *SD*2; the relationship within the square brackets is the *condition* restriction on the T-rule; the right hand side of (5) is the derived structure.

(5), then, exhibits the metalanguage which we will deploy in discussing pictorial relationships and their rôle in determining our intuitive apprehension of form and shape. The problem now becomes that of determining the relationships and the structures which they co-ordinate. Published accounts of 'picture syntax' have not provided any systematic accounts of the variety of pictorial relationships with which they deal, much less a discovery procedure for those relationships. This omission may of course be intentional in the sense that no attempt is being made to capture our intuitive knowledge of picture structure in these picture syntaxes. In this account, however, we adopt as goal the formal description of our pictorial intuitions and accordingly we shall adopt a more or less systematic methodology for ascertaining what these intuitions are.

3. THE METHODOLOGY

The methodology to be employed in deciding the specification of a picture grammar is based upon that of Chomsky. Since (as Chomsky remarks) our intuitions are not always immediately apparent it may prove necessary to resort to consideration of particular expressions and pairs of expressions which have the property of rendering our intuitions clear cut. This is the purpose of the study of ambiguous, anomalous and paraphrastic expressions. A good example of the use of ambiguity would be the pictorial expression in (6) which may be seen in two ways: as a 'bellying sail' or a 'sting-ray'.

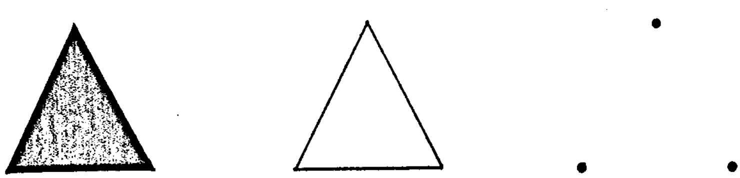
(6)



In the former interpretation we group sides *a* and *c*, *b* and *d*. In the latter 'reading' we group *a* and *d*, *b* and *c*. Any adequate picture grammar must provide the symbolic apparatus by which to exhibit this 'grouping' of edges.

As examples of paraphrase we might take the three pictorial expressions illustrated in (7) which evidently have the same shape.

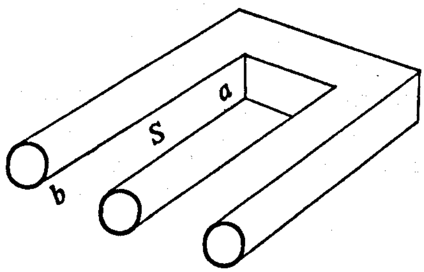
(7)



Our intuitions about this similarity of shape apparently involves relations between edges or lines, between the positions we call 'corners', and involves among other things the idea that lines can 'function' as edges. We require that the picture syntax give us an adequate formal account of these intuitions.

Finally we might introduce varieties of anomalous picture such as that in (8).

(8)



Attempts to assign an interpretation to (8) break down due to the inconsistency involved in assigning to the region *S* the status of *figure* in the vicinity *a*, but *ground* in the vicinity *b*. Making these assignments is evidently tied up with the recovery of certain relationships between the edges – denoted by lines – in the vicinities of *a* and *b*. The picture grammar must give some account of what sort of relationships between 'edges' mediate or force the assignment of such distinctions as 'figure'/'ground'.

4. THE PICTURE GRAMMAR

4.1. The structural representation of position

Restricting ourselves to mechanical means for displaying or addressing pictorial data (that is, excluding the retina) it is clear that the primitive elements of pictorial expressions must be distinct positions in a two-dimensional array. This conventional view is of course based upon the insight of Cartesian co-ordinate geometry which established the (x,y) notation; that is, the representation of position in a plane by two magnitudes which reflect the operations required to address the given position starting from an origin on a defined axis.

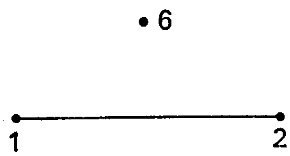
The notation (x,y) implies an axis and an origin but nowhere states it. For our purpose it will be necessary to do so since it (x,y) denotes 'position relative to the origin', and we are committed to making all relationships overt. Thus our representation of position would take the form

$$(9)\qquad \mathit{Relpos}\,\langle \mathit{axis}\langle \underset{1}{p},\underset{2}{p}\rangle[\underset{3}{\mathit{bint}}(\underset{4}{t},\underset{5}{\mathit{bint}})],\underset{6}{p}\rangle[\underset{7}{\mathit{bint}},\underset{8}{\mathit{bint}}]$$

This exhibits the relationship *relative position* (*Relpos*) of a point (or position) suffixed 6 with respect to an axis of co-ordinates defined on the positions suffixed 1,2 as depicted in (10). The value of this relationship is given by a pair of *basic integers* (*bints*) suffixed 7,8 corresponding to the variables

(x,y) in the traditional notation. The characterisation of magnitude by a *bint* rather than some more conventional notation has a specific purpose.

(10)



4.1.1. Bint. In keeping with the desire to symbolise everything by syntactic structure *bint* is defined by the context-sensitive phrase-structure grammar (11):

(11)

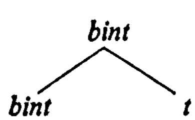
$$\text{(i)}\quad bint \rightarrow \left\{\begin{array}{ll} \left\{\begin{array}{c} bint \\ \# \end{array}\right\} & ,\, t \\ t, & \left\{\begin{array}{c} bint \\ \# \end{array}\right\} \end{array}\right\} \Big/ \_\_\_\_$$

$$\text{(ii)}\quad bint \rightarrow \left\{\begin{array}{c} bint \\ \# \end{array}\right\}, t \Big/ \_\_\_\_\, t$$

$$\text{(iii)}\quad bint \rightarrow t, \left\{\begin{array}{c} bint \\ \# \end{array}\right\} \Big/ \; t\,\_\_\_\_$$

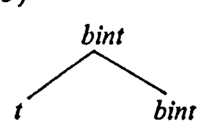
The usual notational conventions are implied here, that is $/\_\_\_\_\,t$ specifies that the category *bint* may only be rewritten according to (ii) if it is in the context of – specifically: immediately precedes – a *t*. Braces indicate alternatives. Thus applying (i) we get two main alternatives:

(12)



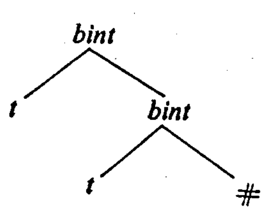
or

(13)



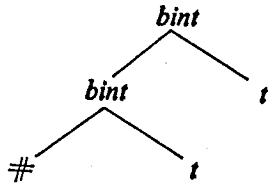
Rules (ii, iii) 'separately' develop (12,13) so that a typical structure resulting from (13) might be (14):

(14)



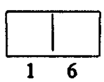
We may think of (14) as specifying the magnitude 1 and of (15) as the magnitude -1

(15)

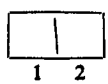


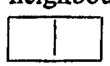
4.1.2. Axis. With this interpretation of *bint* we see that the *bint* suffixed 3 in (9) is of unknown but positive (because of its right-branching structure) magnitude. The mapping of *axis* into pictorial relationships is given by the recursive productions (16,17)

(16) $$axis\langle \underset{1}{p},\underset{2}{p}\rangle[\underset{3}{bint}(\underset{4}{t},\underset{5}{bint})]$$

$$\Rightarrow axis\langle \underset{6}{p},2\rangle[5],$$ 

(17) $$axis\langle \underset{1}{p},\underset{2}{p}\rangle[bint(t,bint(t,\#))]$$

$$\Rightarrow$$ 

Rule (17) asserts that the relationship *axis* of unit magnitude (*bint*(*t*,*bint*(*t*,#))) between positions suffixed 1,2 is identical with the relationship of nearest horizontal neighbours. This pictorial relationship is exhibited in pictorial form thus . In some mechanical device such as a television tube, it would take the form of an incremental voltage applied to the *x*-deflection plates of the tube.

(16) asserts that an axis defined upon points which are not nearest neighbours – the general case – is manifested as a sequence of co-ordinated overlapping pairs of nearest neighbour positions. The number of such positions is determined by the magnitude of 3 in (9). The ',' in the right-hand side of (16) is to be interpreted as '&'.

4.1.3. Relpos. While (9) characterises the notion relative position (and, unlike the Cartesian notation (x,y), explicitly identifies the axis of measurement), it does not exhibit the surface (pictorial) manifestation of this underlying relationship; that is, we have so far failed to provide a derived structure for (9). This mapping can now be specified by a recursive production rule having a similar form to (16,17).

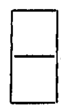
(18) $\quad Relpos \langle axis\langle p,p\rangle [bint],p\rangle [bint(t,bint),bint]$
1 2 3 4 5 6 7 8

$\Rightarrow Relpos \langle axis\langle p,p\rangle [3], 4\rangle [7,8],$ 
9 $\quad$ 1 9

and

(19) $\quad Relpos \langle axis\langle p,p\rangle [bint],p\rangle [bint, bint(t,bint)]$
1 2 3 4 5 6 7 8

$\Rightarrow Relpos \langle axis\langle p,p\rangle [3], 4\rangle [5, 8],$  9 1
9

(18) and (19) 'unpack' *Relpos* into a series of nearest neighbour relationships of two types: one corresponding to the x dimension being , the other corresponding to the y dimension being . (18) and (19) define (recursively) relative position for non-zero magnitudes of x and y. *Relpos* $\langle axis, p\rangle [bint(t,\#), bint(t,\#)]$ characterises a position coincident with the origin of co-ordinates. Accordingly we may complete the formalisation of *Relpos* by

(20) $\quad Relpos \langle axis\langle p,p\rangle [bint], p\rangle [bint(t,\#),bint(t,\#)]$
1 2 3 4

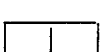
$\Rightarrow axis \langle 4, 2\rangle [3].$

Evaluating a *Relpos* is akin to changing the pen position on an incremental plotter. We may illustrate this for a very simple case. Consider the position (2,2) i.e., $x = 2$, $y = 2$, which in our terms is

(21) $\quad Relpos \langle axis, p\rangle [bint(t,bint(t,bint(t,\#))), bint(t,bint(t,bint(t,\#)))].$

Represent the discrete incremental positions (of the pen) as a square array. Then *axis* of (21) defines some pair of positions labelled 1,2 as shown in (22a) with the associated values of *Relpos*. (For brevity integers are shown as 2, 2 rather than in the explicit structural form of (21)). Applying (18) to this yields a new axis pair denoted 1′,2′ in (22b), and so on until we reach (22e) where the Cartesian value of *Relpos* is now (0,0). (20) now applies yielding (*f*) which provides a labelled position for 4 correctly positioned with respect to the initial origin 1.

The significance of this lies only in the fact that just two pictorial relationships – sensorily manifest relationships, that is – are utilised, namely

 and .

The other relationships *Relpos* and *axis* are defined on these and are, according to our earlier discussion, 'abstract' or underlying relationships. (22) demonstrates, moreover, that any formulation of pictorial relationships which is reducible to *Relpos* can be effectively computed, that is, defines addressable positions on some plane surface, given only that two directions and a unit separation are defined on that surface.

(22)

$$\text{(a)}\quad \begin{array}{ll} 1 & \qquad 2 \end{array} \qquad \textit{Relpos}\,\langle \textit{axis}\langle 1,2\rangle[\textit{bint}],p\rangle[2,2]$$

$$\text{(b)}\quad \begin{array}{ll} 1\ 1' & \qquad 2\ 2 \end{array} \qquad \textit{Relpos}\,\langle \textit{axis}\langle 1',2'\rangle[3],4\rangle[1,2]$$

$$\text{(c)}\quad \begin{array}{ll} \ \ 1'' & \qquad\ \ 2' \\ 1\ 1' & \qquad 2\ 2' \end{array} \qquad \textit{Relpos}\,\langle \textit{axis}\langle 1'',2''\rangle[3]4\rangle[1,1]$$

$$\text{(d)}\quad \begin{array}{ll} \ \ 1''\ \ 1''' & \qquad\ \ 2''\ 2''' \\ 1\ 1 & \qquad 2\ 2' \end{array} \qquad \textit{Relpos}\,\langle \textit{axis}\langle 1''',2'''\rangle[3],4\rangle[0,1]$$

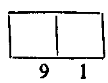
$$\text{(e)}\quad \begin{array}{ll} \qquad 1'''' & \qquad\qquad 2'''' \\ \ \ 1''\ \ 1''' & \qquad\ \ 2''\ 2''' \\ 1\ 1' & \qquad 2\ 2' \end{array} \qquad \textit{Relpos}\,\langle \textit{axis}\langle 1'''',2''''\rangle[3],4\rangle[0,0]$$

$$\text{(f)}\quad \begin{array}{ll} \qquad 4 & \qquad\qquad 2'''' \\ \ \ 1''\ \ 1''' & \qquad\ \ 2''\ 2''' \\ 1\ 1' & \qquad 2\ 2' \end{array} \qquad \textit{axis}\langle 4,2''''\rangle[\textit{bint}]$$

(18), (19) and (20) define position for the first quadrant only. A further version of (18) and of (19) is required dealing with negative magnitudes.

(18a) $\textit{Relpos}\,\langle \underset{1}{\textit{axis}}\langle \underset{2}{p},\underset{3}{p}[\underset{4}{\textit{bint}}],p\rangle[\underset{5}{\textit{bint}}(\underset{6}{\textit{bint}},\underset{7}{t}),\underset{8}{\textit{bint}}]$

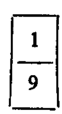
$$\Rightarrow \textit{Relpos}\,\langle \textit{axis}\langle \underset{9}{p},p\rangle[3],4\rangle[6,8],$$



and

(19a) $\textit{Relpos}\,\langle \underset{1}{\textit{axis}}\langle \underset{2}{p},p\rangle[\underset{3}{\textit{bint}}],\ \underset{4}{p}\rangle[\underset{5}{\textit{bint}},\ \underset{6}{\textit{bint}}(\underset{7}{\textit{bint}},\underset{8}{t})]$

$$\Rightarrow \textit{Relpos}\,\langle \textit{axis}\langle \underset{9}{p},p\rangle[3],4\rangle[5,7],$$



A corresponding version of (20) differs only in applying to the negative version of zero.

Notice that the same two pictorial relationships are used; we have merely changed the 'ordering' of the position pairs they relate. Thus, the definitions of *Relpos* express our intuition that if 'up' or 'left' is thought of as 'positive' then 'down' or 'right' is 'negative', where both 'positive' and negative' are given explicit definitions in terms of structural manipulations of bints.

The representation of integer magnitude in a structural form[1] is thus

[1] Originally introduced by my colleague D. J. Langridge to provide a syntactic account of the arithmetic operations plus, multiply, etc.

associated with distance and measure. It is of course possible to formulate *Relpos* using not *bint* but more conventional integer symbolism, e.g. n, $n-1$, etc. We have adopted *bint* here in order to avoid the implication that our intuitions of position are based upon arithmetic concepts. Rather, we would like to suggest that the reverse is the case.

4.2 The accuracy of position judgements

Consider the following simple experiment. We provide a blank, square sheet of paper upon which we have marked a point. The subject (S) is invited to estimate the position of the point relative to the edges of the paper as axes, and to verbalise this estimate as an integer pair. The magnitude of these integers relates to an 'assignment', on the part of S, of an *interval scale* to the vertical and horizontal edges of the paper. It has been found (Klemmer and Frick 1953) that the accuracy of these integer estimates is about 20 per cent; that is, S can discriminate roughly 25 positions in this square. This limitation seems a fundamental one (Miller 1956, Clowes 1967); what does it imply for the structural description of position formulated here? The experimental observations are consistent with the view that the *bints* (suffixed 7, 8) in (9) both having a limiting value of 5. The magnitude of these integers is of course dependent upon the magnitude of the *nearest neighbour interval* (e.g.

).

We may, therefore, say that in judging relative position – recovering, that is, a *Relpos* such as (9) from its pictorial manifestation – S 'chooses' a nearest neighbour interval to define an axis. This interval is sufficiently large that each of the bints in 9 will not exceed 5 in magnitude.

4.3 Pairs of positions

Pairs of points may form an entity which is related to an axis. We may think of a pair as defining a line, i.e., as the positions of the ends of a straight line. Our grasp of the line as an entity involves a relation between the end points of the line which betray an inherent axis. Thus we might say the line is 'vertical', 'sloping' or 'long'. All of these epithets which apply to the relationship between the end points imply an axis and an interval on which that axis is defined. We shall designate this relationship *coord* expressed as

$$(23)\qquad coord\langle p_1,p_2\rangle[Relpos\langle axis\langle 1,p\rangle[bint],\ 2\rangle[bint,bint]]$$

$$\Rightarrow Relpos\langle axis_4\langle p,p\rangle[bint],1\rangle[bint,bint],$$

$$Relpos\langle 4,2\rangle[bint,bint]$$

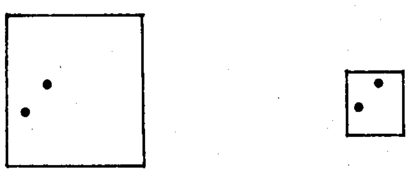
(23) shows how choice of an *axis* (4) commits us to a particular image of the relation between 1 and 2. If we rotated 4 we would grasp 1, 2 differently. (23) *fails*, however, to bring out the fact that the *axis* on the left-hand side is 'parallel' to that on the right hand side (*see* Postscript).

4.3.1. Near. In describing a line as 'long' or 'short' we are implicitly relating its endpoints: the judgement appears to apply to positions, minimally to pairs of positions. If position is characterised by *Relpos*, and we are to seek a formal specification of position judgements in structural terms then it is natural to seek to characterise *near* as, say, an identity condition upon the independent structural characterisation of two positions. Thus we might argue that two *Relpos*'s (of the form (9)) having the same *axis* and the same *bint* (7,8) value would describe two positions which are near to each other. We have seen that the use of different nearest neighbour intervals in the definition of *axis* implies a labile metric for judgement of position. It follows, then, that in judging that two positions are near to one another we are assigning an axis – common to both *Relpos*'s – which makes them so. We may express this as a relation.

(24) $$near\langle \underset{1}{p}, \underset{2}{p}\rangle [Eq\langle \underset{3}{bint}, \underset{4}{bint}\rangle, Eq\langle \underset{5}{bint}, \underset{6}{bint}\rangle] \Rightarrow Relpos\langle \underset{7}{axis}, 1\rangle [3,5], \quad Relpos\langle 7,2\rangle [4,6]$$

The crucial question, in making this judgement therefore rests upon the choice of axis.

(25)



(a) (b)

Thus, in (25a), relative to the rectangular frame (an *axis*) the pair of points appear close, but in (25b) they do not. Of course the axis may not be an external one.

(26)

(a) (b)

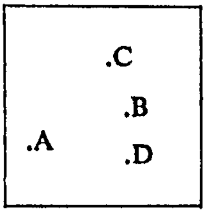
For example, where we are judging the proximity not of positions but of complex pictorial forms as in (26), the *axis* may be provided by the forms themselves. Thus large characters (26a) appear nearer than small ones (26b).

In introducing the metalanguage (§2) we made use of relationship *Coinc* (5) between a pair of positions. What is the distinction between 'coincident' and near? We shall take it that coincident is an identity relationship upon a pair of positions, i.e., that they are the same positions.

4.4 Side

A weaker form of positional judgement than that involved in *Relpos* is the judgement as to which side of a pair of points a third point lies. For example in (27) the points C and D are on opposite of the line AB, while A and D are on the same side of the line CB.

(27)



In making this judgement it is intuitively apparent that a reorganisation of structure is involved. Thus, prior to the remark about 'side', we see the points as occupying unrelated positions in the rectangle. *It* acts as the frame of reference or origin of coordinates. In making the 'side' judgements we use first AB and then CB as the frame of reference. Thus (27) is an ambiguous picture and as such reveals the syntactic structure of the 'side' relationship.

The first organisation would be characterised by four distinct expressions of the form of (9) each taking A,B,C,D as the 'point' of the *Relpos*, i.e., as the item suffixed 6 in (9), all identifying the same positions as *axis*. Informally we may associate the latter with the base of the rectangular frame in (27). The 'side' judgement however takes a pair of these positions, say A,B as axis.

(28) $$side\langle \underset{1}{p},\underset{2}{p},\underset{3}{p}\rangle[\underset{4}{bint}] \Rightarrow Relpos\langle axis\langle 1,2\rangle[bint],3\rangle[bint,\ 4]$$

Thus in (28) the positions suffixed 2, 3, 5 might be A,B,C. The derived structure of (28) is a *Relpos* relation between C and A,B as axis. (28) thus defines *side* as that relationship of relative position in which the '*x*-co-ordinate' is undetermined.

The judgement 'D and C are on opposite sides of AB' is the result of a comparison of two relational structures of the form typified by the left-hand side of (28) both having the same *axis*. It follows, of course, that the *bint* (suffixed 6 in (28)) in one structure will be of opposite sign to that in the other – it is this which we will regard as underlying the judgement.

The judgement 'A and D are on the same side of CB' is another comparison of a pair of relational structures each involving CB as *axis*, with A,D as the positions.

We may treat 'opposite sides' and 'same side' as *relations* between two points and a pair of points. Thus *sside* (same side) might take the form

(29) $$sside\langle \underset{1}{p},\underset{2}{p},\underset{3}{p},\underset{4}{p}\rangle\ [Ssign\langle \underset{5}{bint},\underset{6}{bint}\rangle] \Rightarrow side\langle 1,2,3\rangle[5], side\langle 1,2,4\rangle[6]$$

Ssign (same sign) is the relationship 'same structure type' on the pair of *bints* to which it applies (remembering that the distinction between positive and negative is exhibited in the structure of *bint*).

What (29) does of course is to specify formally what was earlier stated informally as the 'comparison of two relational structures of the form typified by the left-hand side of (28)'. That is *sside* is assigned to this set of four points just in case the separate *side* relationships on the right-hand side of (29) satisfy the *Ssign* relationship specified on the left-hand side.

4.5. Discussion

The foregoing has established that, with the notational system introduced in §2, we may characterise a variety of relationships between positions. The relationships to be characterised are made evident by considering various pictorial examples in accordance with the general methodological approach outlined in §3, that is, we are characterising our intuitions about picture structure, *not* erecting some arbitrary picture calculus.

Our perception of these varieties of pictorial organisation can be identified with the assignment of these functional descriptions, e.g. *Relpos*, *sside*, etc., to the primitive sensorily-manifest data. The process of assignment is essentially a *parsing* process. It will be evident that given say the illustration (27) there are very many (probably an infinite number of) relationships which could be assigned to this collection of points and the frame. We suggest that this is entirely consistent with observation: there *are* many ways of looking at (27). Significantly, however, we cannot hold these multiple views simultaneously – we switch between them. Formally, that is, we can only assign a single relational structure at a time, although this structure may relate a number of items, e.g. *sside*, in quite a complex manner.

There is one major respect in which the whole of the presentation to this point has been inadequate. In exhibiting various pictorial expressions manifesting varieties of position relationship, we have *assumed* that a position is denotable by a 'point' and that a 'straight line' has two salient positions associated with its two 'end points'. These positions are abstractions which underlie the *forms*, 'line', 'point', 'frame', etc. The whole apparatus is empty of empirical interest (i.e., has no application to picture interpretation) unless we can also characterise how these abstractions are possible, that is, characterise *form* with the same apparatus used to characterise *position*.

5. THE CHARACTERISATION OF FORM

The recovery of Form may be stated to be the discovery of the *significant* positional relationships exhibited in the picture. Thus, in our preceding account we have denoted position by a point and pairs of positions (as in *coord* for example) by a straight line. This denotation presumes that from these two types of form – point and line – the reader may easily recover the positional structure which underlies them. In picture interpretation, the raw data is a very large number of possible relationships between positions (sampled by a scanner) having distinguishable colours, e.g. between all pairs of raster positions. A picture containing a line clearly has more than one significant positional relationship exhibited in it although there may be only one (the relationship between end points of the line) which we wish to utilise. The positional relationships between say black and white points which subsume the *edge* of the line (equivalently the edge of a point), are obviously necessary to the exhibition of the line itself and the recovery of these more primitive relationships must precede (in parsing the picture) the recovery of the relationship between the end points of the line.

The essential difference between the interpretation of computer graphics and their hard-copy equivalents lies precisely in the fact that these 'end point relationships' evident in both, are the raw data when input at the graphics console, but *extensively parsed* data when recovered from the hard-copy *via* a scanner.

The object of parsing is in some sense (one which we will progressively make sharper) the recovery of objects with which we can associate some 'position information', i.e., a line having as 'position information' its end points. We may regard an object to be defined as 'position information' upon which a variety of position relationships are specified. It will be convenient to utilise much the same notation as already deployed except that the name of the relationship, e.g. *side*, will be replaced by the object name, e.g. *LINE*.

5.1. Straight edges

We shall adopt the notation p (*colour* i) to designate a position having colour i. Clearly, objects or forms are ultimately dependent for their exhibition upon *distinctions* of colour between 'sets' of positions enjoying certain varieties of spatial relationship. In fact, we could regard $p(colour\ i)$ as an abbreviation for an object definition involving two or more sets of positions having colours i,j respectively, enjoying the relationship $i \neq j$. Underlying the form 'straight edge' (abbreviated to *SEDGE*) we discern a pair of positions – the 'ends' of the edge – and a colour relationship between the opposite sides of the edge.

(30)

$$SEDGE\langle \underset{1}{p}, \underset{2}{p}, \underset{3}{colour}, \underset{4}{colour} \rangle [side\langle 1,2,p(3) \rangle [0], side\langle 1,2,p(4) \rangle [-1], Diff\langle 3,4 \rangle]$$

(30) provides a partial formalisation of this form in terms of the *side* relationship. The values $(0,-1)$ of the two *side*s employed restricts the scope of the colour contrast to be local to the edge in the y direction. However, it leaves unspecified the range of the contrast in the direction of the edge, i.e., in the x direction. This is a direct consequence of the formulation of *side*; note, however, that since isolated straight edges are pictorially anomalous, that is they can only occur in the context of some extended boundary, no parsing problems should arise from this imprecision. The relationship *Eq* ⟨*colour*,

colour⟩ may be said to have the value *true* when the two colours are discernibly different, otherwise *false*. The relationships predicated as underlying *SEDGE* are essentially those recovered by various types of 'edge follower' (Greanias 1963, Ledley 1964) in so far as these programs only examine some restricted neighbourhood of positions in the picture in order to assign an 'edge'.

5.2. Convex boundaries

The simplest 'context' in which a straight edge can occur is that whose underlying 'positional information' is what we apprehend as a convex polygon. Let us call this form a convex boundary (*CBND*).

(31)

$$\mathit{CBND}\langle \underset{1_i}{\mathit{SEDGE}}\langle \underset{2_i}{p},\underset{3_i}{p},\underset{4_i}{\mathit{colour}},\underset{5_i}{\mathit{colour}}\rangle[\quad],$$

$$\underset{1_{i+1}}{\mathit{SEDGE}}\langle \underset{2_{i+1}}{p}, \underset{3_{i+1}}{p}, \underset{4_{i+1}}{\mathit{colour}}, \underset{5_{i+1}}{\mathit{colour}}\rangle[\quad],\ldots\rangle$$

$$[\mathit{Diff}\langle 4_i, 4_{i+1}\rangle, \mathit{Between}\,\langle 2_i, 3_i, 2_j, 3_j, p(4_i)\rangle,$$
$$\mathit{Coinc}\,\langle 3_i, 3_{i+1}\rangle, \mathit{side}\,\langle 2_i, 3_i, 3_{i+1}\rangle[\mathit{bint}(\mathit{bint},t)]$$
$$\mathit{sside}\,\langle 2_i, 3_i, 2_j, 3_j\rangle[\quad] - \textit{for all } i \neq j]$$

The formulation (31) is in terms of five types of relationship (*Diff*, *Between* *Coinc*, *side* and *sside*) predicated upon an ordered set of *SEDGE*s. The ordering is determined by the values of *Coinc* and *side* which characterise continuity (or connectivity) and a clockwise (because *side* is negative) progression around the boundary. Convexity is specified by *sside*. The 'fact' that the spatially distributed colour relationships which support a form are not local to the edge (the assumption underlying *SEDGE*) is exhibited by the relationship *Between*.

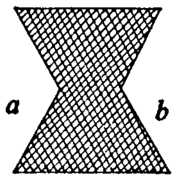
(32)

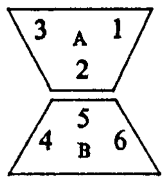
$$\mathit{Between}\langle \underset{1}{p},\underset{2}{p},\underset{3}{p},\underset{4}{p},\underset{5}{p}\rangle[\mathit{Ssign}\langle \underset{6}{\mathit{bint}(t,\mathit{bint})},\underset{7}{\mathit{bint}(t,\mathit{bint})}\rangle, \mathit{sside}\langle 1,2,4\rangle]$$
$$\Rightarrow \mathit{side}\langle 1,2,5\rangle[6], \mathit{side}\langle 3,4,5\rangle[7]$$

Of course such a form – one in which the 'interior' colour is uniform – is idealised. Naturally illuminated objects present interiors whose illuminance varies in a highly complex fashion. Thus the two-dimensional distribution of retinal illumination produced by a uniformly illuminated sphere, is dark at the edges (Lambert's law), brightening uniformly towards the centre.

5.3. Compound forms

All forms do not of course have an underlying *CBND*. It is an essential part of this approach to picture syntax that we regard Forms containing concavities of boundary, as concatenations of Forms which *are* convex. Intuitively, the varieties of concatenation would be described as 'join', 'overlap', and 'touch'. Such descriptors evidently apply to complete convex Forms in such a manner as to merge the two *CBND*s deleting one or more *SEDGE*s and introducing concavities characterised by positive *side* relationships. Thus, if we take (34) to be the perceptual organisation underlying (33),

(33) 

(34) 

then the *BND* derived from joining the *CBND*s of A and B would have 1 followed by 6, 4 followed by 3 and positive *side*s relating 1 and 6, 4 and 3. The *SEDGE*s 2 and 5 would have been 'deleted'.

The concatenation requires several relationships including an 'agreement' between the 'colour pairs' (4_i, 5_i of (30)) for each *CBND*. In the event that A and B are of different colour *SEDGE* 2 will be 'merged' with *SEDGE* 5 not deleted, in the derived structure there will be a single *SEDGE* replacing 2 and 5, this *SEDGE* having an appropriately modified colour contrast. This form of join is appropriate to pictures such as maps: it will not be explored here.

Whether 2 and 5 are deleted or merged there will be two positions (*a*,*b* in 33) in the resultant *BND* where essentially new side relationships will be introduced. At least one of these positions a concavity (i.e. a negative *side*) must be introduced if we are to be able to recover the underlying pair A,B. [There is a weak sense in which any n-gon ($n>3$) may be decomposed into $n-2$ triangles and so on, even if the n-gon is convex. We do not consider this case here.]

The key relationship between A and B to which the term 'join' applies is of course the 'coincidence' of the *SEDGE*s underlying 2 and 5; that is, (33) is decomposable into the two parts A, B of (34) at which 'point' the relationship between A, B 'emerges'.

5.4. Discussion

In the foregoing sketch, we see how the *relations* defined in §4, mediate our apprehension of Form. From the standpoint of analysing pictures, we may say that the relationships underlying *CBND* (31), are predicates whose value must be *true* over the set of *SEDGE*s which constitute the arguments of these predicates. We may think of these arguments as the parts of *CBND*.

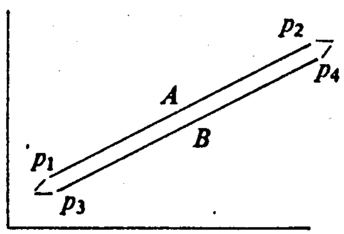
The formulation (31) is close to that developed by Evans (1968) and Guzman (1968). The crucial issues, however, are identified as being concerned with questions having to do with relationships not objects or forms, that is, 'How many relationships are there?'; 'How are they related and defined?', and so on. The answers given here to these questions may require revision in the light of further study; at present, however, the list of relationships is evidently small, perhaps 10-20, and they form a hierarchy in that more powerful relations are defined in terms of simpler ones (*see* Postscript).

It should now be evident, therefore, that a characterisation of Form is possible along these (relational) lines, and accordingly we can claim to have met the objection formulated in §4.5 concerning the status of observations about position which rely upon a grasp of Form for their statement. Specifically we may ask 'What is a straight line?'

6. SHAPE

A straight line is a Form having a Boundary whose underlying positional relationships are specified by a single *CBND*; that is, it is not a compound form. The *SEDGE* set comprising this *CBND* is further characterised by having a pair of *SEDGE*s between which the relationship *parallel* obtains and whose *p*s enjoy a *near* relationship. That is if *A*, *B* (in (35)) are the *coord*s in question, then p_1 is *near* p_3 and p_2 *near* p_4.

(35)



Thus we see that 'straight line' (and any other line for that matter) involves *a judgement of positional relationships upon specified elements of a Form.* We have seen (§ 4.2) that the accuracy of positional judgements is based upon the assignment of an *axis* which acts as a scale determiner. If we take this axis to be either A or B then we see that we are saying that a Form will appear line-like if, in addition to the requirements set out above, p_1 is *near* p_3 and p_2 *near* p_4, taking A or B as axis. This will be the case if the form is, as it were, much 'narrower' than it is 'long'.

The crucial concept underlying this formulation of 'line' is that it 'involves a judgement of positional relationships upon specified elements of a Form'. We take this to be a general definition of Shape.

In this way we see that Shape involves Form but involves in 'addition' the 'recovery' of further position relationships whose accuracy is, of course, subject to the limitations discussed in § 4.2. We may think of these additional relationships as the *metrical aspect of Shape.* In judging the similarity of Shape of two or more forms, e.g. as in (7), we first recover the *CBND*s underlying the 'raw data' – and note we attempt to recover the same *CBND*s – then we evaluate metrical relationships between the various *p*s.

Where there are alternative sets of metrical judgements of an essentially different kind, e.g. *Relpos* as against *parallel*, we may see that a single Form has two or more alternative Shapes. This is the case in (6).

We noted in § 5 that 'the raw data is a very large number of possible relationships between positions (sampled by a scanner) having distinguishable colours'. The characterisation now given of Form and Shape suggests that the central problem in the assignment of structure to a picture (i.e., in Picture Interpretation) is the 'decision' as to what varieties of relationship are to be recovered, since as we have seen there is no '*S*' (in the normal grammatical sense) from which all well-formed pictorial sentences are derived. This is a more or less direct consequence of espousing a wholly relational and transformational syntax. For the simple (whatever that means) pictures we have discussed thus far it appears counter-intuitive to suggest that there are many possible Forms and Shapes which are recoverable, i.e., visible in it. When faced, however, with a wholly novel picture, for example that produced in some esoteric experiment in physics, we may find that it takes some considerable time to adjust our view so as to recover the significant elements of Form and reject the insignificant. That we have a strong predisposition to see certain Forms and Shapes and not others is of course familiar to the psychologist. Boring's 'Young Girl/Mother in Law' (1930) is a classic example.

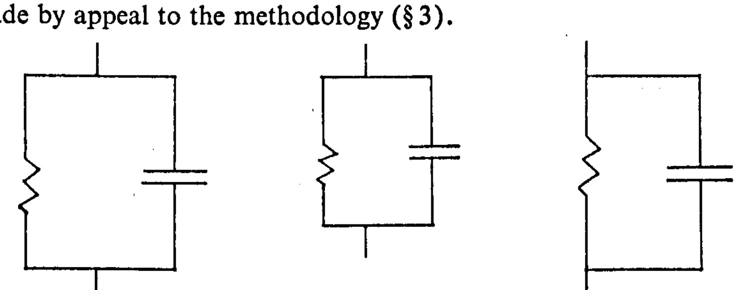
The conclusion we would draw from this is that the structure we assign to a picture is determined not solely by the 'raw data' of that picture but also by *a priori* decisions as to the varieties of relationship we expect to find there. The question therefore becomes 'can we formalise these *a priori* decisions?'

7. THE SEMANTICS OF PICTURES

We may summarise the foregoing argument as 'People see what they expect to see'. The essential rider is that *what* they want to see is *things* not pictorial relationships, that is, the *a priori* decisions reflect assumptions about the things and events which we expect to see exhibited in the picture. We shall argue that it is necessary and indeed possible to give a structural characterisation of things and events which is a mapping of the relational structure of the picture. This characterisation we call the *semantics* of the picture.

The case for something beyond the recovery of pictorial relationships is readily made by appeal to the methodology (§ 3).

(36)



(a) (b) (c)

Thus, while (36a) and (36b) are pictorial paraphrases, there is a variety of paraphrase, evident between both of them and (36c) which cannot be established on pictorial grounds. It is the underlying electrical relationships which are the same in all three pictures. We might expect that, corresponding to *Relpos*, we have *potential difference* and also *phase difference*. These relationships are, of course, purely abstract, that is, they cannot be recovered except *via* some sensory manifestation, usually pictorial. Just what the syntactic structure of a circuit might look like, utilising electrical relationships in the metalanguage, is not yet clear. The example parallels closely those given by Chomsky (1965, pp. 160–3) in discussing 'additional problems in semantic theory'. He suggests that anomalies such as '* the cut has a finger' (*vs.* 'the finger has a cut') are to be accounted for not in terms of language use but in terms of 'language independent constraints . . . in traditional terms, the system of possible concepts'. He remarks 'it is surely our ignorance of the relevant psychological and physiological facts that makes possible the widely held belief that there is little or no *a priori* structure to the system of "attainable concepts"'. The semantic structure we have been arguing for here is in our view identifiable with Chomsky's 'structured system of attainable concepts'.

We may note that developments in question-answering programs are placing increasing weight on the structure of the data base. From the standpoint espoused here we would regard the data base as exhibiting the relationships between *events*, e.g. games in Baseball (see Green *et al.* 1963), and entities, e.g. teams, places, scores. These relationships characterise our knowledge of league games in the same way that *Relpos* characterises our pictorial knowledge.

8. CONCLUSION

In this paper an approach to picture interpretation has been outlined. This views the process as one of assigning to the 'raw data' sampled by a picture scanner (equivalently the retina) a structure which makes explicit the varieties of pictorial relationship *visible* in that picture. These relationships are specified in a metalanguage (regarding the 'raw data' as an object language) having a strong similarity to that deployed in transformational grammar. In choosing to characterise *visible* structure we adopt a methodology which is intended to expose just those relationships which mediate *our* grasp of Form and Shape.

Among the many problems thrown up by this work we would single out the characterisation of the semantics of pictures in relational syntactic terms as crucial. The parallels drawn with current work in question-answering suggests that it may be profitable to consider not only event structures having a pictorial manifestation but events readily characterised in English too. Indeed we might consider situations like particle physics where we have bubble chamber photographs, English descriptions of particle interactions, and an algebraic representation , e.g. $\mu^{-}+p-\varepsilon^{\circ}+K^{\circ}$, as ideally suited to determine the semantic relationships.

A machine (or program) capable of mediating translations between these various languages would utilise the underlying semantic structure as the 'pivot' of the translational process. We could describe such a machine as 'informed' – 'informed', that is, about the varieties of relationship applicable in these various representations of an event. It would not, however, be intelligent. Such an appellation should be reserved for a machine (like us) capable of formulating and testing hypotheses about *new* relationships and ultimately about new systems of attainable concepts manifesting these relationships.

Acknowledgements

The approach to picture interpretation outlined here has emerged from the Verbigraphics Project. It is a pleasure to acknowledge my indebtedness to my colleagues D. J. Langridge and R. B. Stanton. I am grateful to Dr G. N. Lance for encouraging us and supporting us in this work.

REFERENCES

Boring, E.G. (1930) A new ambiguous figure. *Am. J. of Psych.*, **42**, 444–5.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.

Clowes, M.B. (1967) Perception, picture-processing and computers. *Machine Intelligence 1*, pp. 181–98 (eds Collins, N.L. & Michie, D.). Edinburgh: Oliver & Boyd.

Clowes, M.B. (1968) Transformational grammars and the organisation of pictures. Seminar Paper No. 11. CSIRO Div Computing Research, Canberra. Presented to the Nato Summer School on Automatic Interpretation and Classification of Images, Pisa, Italy.

Evans, T.G. (1968) Paper presented to the Nato Summer School on Automatic Interpretation and Classification of Images, Pisa, Italy.

Greanias, E.C. & Meagher, P.F. (1963) The recognition of handwritten numerals by contour analysis. *IBM J. Res. & Dev.* **17**, 1.

Green, B.F., Wolf, A.R., Chomsky, C. & Laughery, K. (1963) Baseball: an automatic question answer. *Computers and thought*, pp. 207–16 (eds Feigenbaum, E.A. & Feldman, J.). New York: McGraw-Hill.

Guzman, A. (1968) Paper presented to the Nato Summer School on Automatic Interpretation and Classification of Images, Pisa, Italy.

Kirsch, R.A. (1964) Computer interpretation of English text and picture patterns. *I.E.E. Trans. on Electronic Computers*, **13**, 363–76.

Klemmer, E. T. & Frick, F.C. (1953) Assimilation of information from dot and matrix patterns. *J. Exp. Psychol.*, **45**, 15–19.

Ledley, R.S. (1964) High speed automatic analysis of biomedical pictures. *Science*, **146**, 216–23.

Miller, G.A. (1956) The magical number seven; plus or minus two: some limits on our capacity for processing information. *Psych. Rev.*, **63**, 81–97.

Minsky, M. (1961) Steps toward artificial intelligence. *Computers and thought*, pp. 406–50 (eds Feigenbaum, E.A. & Feldman, J.). New York: McGraw-Hill.

Narasimhan, R. (1966) Syntax-directed interpretation of classes of pictures. *Communs Ass. comput. mach.*, **9**, 166–73.

Shaw, A.C. (1967) A proposed language for the formal description of pictures. *GSG Memo* **28**, Stanford Linear Accelerator Centre.

Sutherland, I.E. (1963) Sketchpad – a man-machine graphical communication system. MIT Lincoln Lab. Tech. Rept. No. 296.

Woodward, P.M. (1966) LIST programming. *Advances in programming and non-numerical computation*, pp. 29–48 (ed. Fox, L.). Oxford: Pergamon Press.

POSTSCRIPT

During the process of putting the manuscript of this paper into typescript several points have emerged which help to place this work in better perspective.

Coord: There are two aspects of this entity which make it clear that it should be regarded as an object (cf. *SEDGE*) rather than a relationship: (i) The *Relpos* included in the L.H.S. of (23) references the elements suffixed 1,2. This is completely atypical of relation definitions but quite characteristic of object definitions. (ii) We have not utilised *coord* as a relationship in the formulation of any other relationship or object nor does it seem likely that we would want to. We therefore conclude that (23) is incorrect and that (23a) is more likely.

(23a) $COORD\langle \underset{1}{p},\underset{2}{p}\rangle[Relpos\langle axis\langle 1,p\rangle[bint],2\rangle[bint,bint]]$

Objects and Relations: In defining *SEDGE, CBND* and (informally) *LINE* we have deployed only those relationships defined earlier in the paper (we could hardly have utilised any others). These relationships are characterised as being defined on, and in terms of, *positions*. If we regard position as an object (as suggested in § 5.1) then the relationships defined in this paper are those which involve just position(s) as the objects in terms of which they are defined. The informal discussion of compound forms (§ 5.3) makes frequent reference to *SEDGE*. Underlying every *SEDGE* we may presume a *COORD*, and it may be plausible, therefore, to consider *COORD* as the object necessary to the formulation of these 'higher' order relations 'join', 'touch', etc.

In terms of the theories of generative grammar espoused by Chomsky, it is tempting to identify these higher-order relations as those involved in co-ordinate constructions (Chomsky 1965, p. 134 and Note 7, p. 224), with the consequence that the lower-order relations (those defined on positions) might be identified with intra-sentential structure. Thus the inter-word and inter-phrase relationships, reflected in the use made of lexical substitution, might be regarded as *relations* defined on words and phrases as objects. The question which then arises is 'why the distinction between the notational system for inter-sentential co-ordination (generalised T-rules) and intra-sentential co-ordination (strict subcategorization, selectional rules, and lexical substitution)?'

I would argue that the Aspects answer (which makes a distinction) obscures a real uniformity (and therefore an economy) in the metalinguistic apparatus, and may well be obscuring our formal grasp of the linguistic significance of the word. Thus the analogy between picture points (positions) and words breaks down precisely because a word has a very complicated definition (its lexical entry) which from our standpoint would brand it as an *OBJECT*. To grasp an Object is to grasp the relationships which underly it. To regard words as sentences (complex objects) would perhaps be one way to tackle the anomaly underlying say '*phonophone' (*see* Chomsky 1965, p. 187).

These speculations may prove empty; what they point to, however – as does the rest of this paper – is the necessity for a clearer grasp of the distinction between relations and objects in the formulation of syntactic theories.