Environment and Planning B, 1976, volume 3, pages 187-210

Two exercises in formal composition

G Stiny
628 Hauser Boulevard, Los Angeles, California 90036, USA
Received 16 July 1976

Abstract. Spatial relations and shape grammars are defined. A spatial relation is specified by a finite number of shapes arranged in a certain way. A shape grammar provides for the recursive generation of arrangements of shapes in terms of given spatial relations. The possible spatial relations between two edge-connected polygons are characterised. The use of these spatial relations to define shape grammars generating distinct languages of shapes is discussed. The construction of shape grammars to generate infinite languages of shapes containing a given shape made up of polygons is also considered.

1 Introduction

It has been argued recently that the study of form for itself should again become a central enterprise in architectural research and education. As March and Matela (1974, page 212) remark, "... a designer with a well-understood and structured vocabulary of form is more likely to find suitable matchings with functional requirements than one who attempts to let form follow function in some supposedly self-generative way". Of course, this view is justified only to the extent that knowledge about form can be obtained and taught in a scientific way.

The possibility of a full-fledged science of form presupposes a set of basic tools for the systematic and uniform treatment of a host of varied problems in formal composition. Formal composition consists of arranging or combining or putting together certain spatial elements in accordance with some system of rules. A notorious example is *Beaux Arts* composition(1). Here the spatial elements are the functional volumes of buildings—rooms, vestibules, exits, and staircases—and the system of rules determines the symmetrical disposition of these volumes about one or more axes. The nature of problems in formal composition depends on the spatial elements and the systems of rules involved. In this paper, formal composition is regarded in its widest possible sense. Spatial elements are considered generally as shapes that may be two-dimensional or three-dimensional, rectilinear or curvilinear, or structural or functional components of rooms, buildings, or towns. Systems of rules are considered generally as classes of spatial relations that may fix the disposition of shapes in any spatially conceivable way.

The basic tools for the treatment of problems in formal composition should be independent of the spatial elements and the systems of rules involved. That is, these tools should apply equally well to any collection of shapes and to any class of spatial relations. Further, for these tools to be fully productive they should be easily usable and understandable by architects and designers with strong visual (formal in the architectural sense) intuitions, and at the same time suitable for the rigorous mathematical (formal in the logical sense) investigation of form and spatial organization. In this paper, shape grammars are proposed as one of these basic tools.

Shape grammars are defined in section 2; their use in two general exercises in formal composition is illustrated in section 3. An effort has been made to keep the

(1) An interesting discussion of the classical academic tradition of composition in terms of present work in computer-aided design in architecture is given in Mitchell (1973).

presentation of the ideas in this paper visually compelling but always mathematically precise.

2 Shape, spatial relations, and shape grammars

2.1 *Shape*

A *shape* is a finite arrangement of lines. A two-dimensional shape can be drawn in a finite amount of time, for example on a piece of paper with a limited number of pencil strokes. A three-dimensional shape can be constructed in a finite volume in a finite amount of time. A shape can contain occurrences of straight or curved lines, connected or disconnected lines, or open or closed lines.

In this paper, only two-dimensional rectilinear shapes(2) are considered. These shapes are specified by drawing them in a fixed, two-dimensional Cartesian coordinate system. This coordinate system is usually not given explicitly, its origin, axes, and units being understood.

Two important relations are defined for shapes:

One shape is a *subshape* of a second shape if and only if every part of the first shape is also a part of the second shape. That is, the first shape coincides point for point with some part of the second shape in the coordinate system in which they are drawn.

Two shapes are *pictorially equivalent* (identical) if and only if the first shape is a subshape of the second, and the second shape is a subshape of the first. That is, the two shapes coincide point for point in the coordinate system in which they are drawn.

The following operations allow for the manipulation of shapes:

The shape produced by combining two or more shapes is called the *shape union* of the shapes. The shape union of shapes is specified by superimposing their drawings so that the coordinate systems in which they are drawn coincide. Sometimes the shape union of shapes is called an arrangement of shapes.

The *Euclidean transformations* are translation, rotation, scale, and mirror image, or finite compositions of these. A Euclidean transformation G of a shape s is the shape denoted by G(s). The shapes s and G(s) are *similar*, that is, they may differ only in location, orientation, size, or reflection.

A shape produced by taking the shape union of shapes or Euclidean transformations of shapes in a set of shapes S is said to be *made up* of shapes in S. For example, all rectilinear shapes are made up of a shape consisting of exactly one straight line.

2.2 *Spatial relations*

When two or more shapes are combined to form a new shape, they have a certain *spatial relation*. This spatial relation is usually talked about in terms of what we see when we look at the arrangement of these shapes. For example the shapes s_1 and s_2 drawn in figure 1(a) have the spatial relation shown in figure 1(b). The longest edge x of the triangle s_1 and the next to longest edge x of the triangle s_2 coincide. Together these triangles form a trapezoid. The shapes t_1 and t_2 drawn in figure 1(c) have the spatial relation shown in figure 1(d). The longest edge x of the triangle t_1 and the next to longest edge x of the triangle t_2 coincide. But together these triangles form a quadrilateral.

The easiest way to specify a spatial relation is to draw it, that is, to give the shapes that have the spatial relation and to show their shape union. Any Euclidean transformation of an arrangement of shapes is considered to have the same spatial relation (to 'look' the same) as the arrangement.

(2) A two-dimensional rectilinear shape is taken to be any finite arrangement of straight lines in the plane. For a more traditionally mathematical treatment of these shapes see Stiny (1975, pages 117-170). Although only two-dimensional rectilinear shapes are considered, the methods and results presented apply equally well to three-dimensional or curvilinear shapes.

More precisely, a spatial relation is specified by a set of shapes in this way. Let an equivalence relation R be defined on finite sets of shapes S and S' by:
S R S' if and only if S and S' contain the same number of shapes and there is a Euclidean transformation G that makes every shape in S pictorially equivalent to some shape in S'.
Each equivalence class determined by R is called a spatial relation and is specified by one of its elements.

In some cases it may be more proper to define the equivalence relation R in terms of only some of the Euclidean transformations or in terms of additional transformations. For example if mechanical equilibrium is taken to be a compositional factor, then R would more properly be defined in terms of translation and scale or finite compositions of these. Otherwise, two sets of shapes could be in the same equivalence class, but the shape union of all the shapes in one set could have mechanical equilibrium, and the shape union of all the shapes in the other set not. If only topological properties are taken to be compositional factors, then R would more properly be defined in terms of the Euclidean transformations and shape deformations of several types.

The shapes in a set S' are said to have the spatial relation specified by a set of shapes S if and only if S and S' are members of the same equivalence class. For example, the shapes u_1 and u_2 drawn in figure 1(e) have the spatial relation specified by the set of shapes $\{s_1, s_2\}$, where s_1 and s_2 are the shapes drawn in figure 1(a). Sets $\{s_1, s_2\}$ and $\{u_1, u_2\}$ both contain two shapes and there is a Euclidean transformation that makes s_1 pictorially equivalent to u_1 at the same time as it makes s_2 pictorially equivalent to u_2. The shapes t_1 and t_2 drawn in figure 1(c) do not have this spatial relation. Even though sets $\{s_1, s_2\}$ and $\{t_1, t_2\}$ both contain two shapes, there is no Euclidean transformation that makes s_1 pictorially equivalent to t_1 or t_2 at the same time as it makes s_2 pictorially equivalent to t_1 or t_2.

If the shapes in two sets have the same spatial relation, then the shape union of all the shapes in the first set is similar to the shape union of all the shapes in the second set. The converse of this statement is not true. For example, as shown in figure 1(h), the shape union of the shapes v_1 and v_2 drawn in figure 1(f) is similar to the shape union of the shapes v_3 and v_4 drawn in figure 1(g). But the sets of shapes $\{v_1, v_2\}$ and $\{v_3, v_4\}$ specify different spatial relations.

Any finite set of shapes specifies a spatial relation. In this paper, spatial relations specified by sets of shapes in which a shape is similar to another shape or to part of another shape are of special interest. The spatial relations specified by the sets $\{s_1, s_2\}$, $\{t_1, t_2\}$, $\{v_1, v_2\}$, and $\{v_3, v_4\}$, where the shapes in these sets are the shapes of figure 1, are of this type.

Once a spatial relation has been specified, it can be used as a pattern or template for the recursive definition[3] of a class of shapes. For a spatial relation specified by

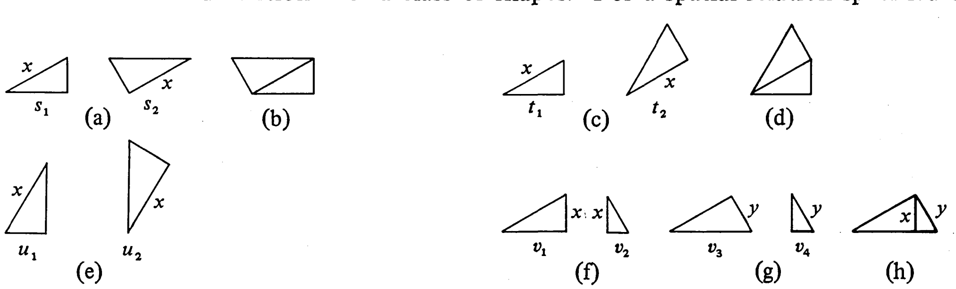


Figure 1. Spatial relations.

[3] For an extensive but elementary treatment of the important mathematical method of recursive definition, see Minsky (1967). A discussion of some other recursive techniques for defining shapes is given in Stiny (in press).

a set of shapes, S, a class of shapes is defined recursively in this way:
(1) If s is a shape in S, then s is a shape in the class.
(2) If t is a shape in the class and there is a shape s in S and a Euclidean transformation G such that G(s) is a subshape of t, then the shape produced by replacing the occurrence of G(s) in t with the shape resulting from the application of G to the shape union of all the shapes in S is a shape in the class.
(3) No other shapes are in the class.
This recursive definition gives a method for constructing shapes in terms of the shapes in the set S. The construction of a shape begins with some shape in S. If a shape s is similar to some part of a shape t under construction, then shapes similar to each of the remaining shapes in S can be added to t so that the part of t similar to s and the added shapes have the spatial relation specified by S. That is, the part of t similar to s is a couple or link between t and shapes similar to shapes in S. Each shape in S provides for this coupling or linking in a different way.

The application of the recursive definition is made clear by a simple example. Suppose S is the set of shapes $\{s_1, s_2\}$, where s_1 and s_2 are the triangles drawn in figure 1(a), and s_1 and s_2 have the spatial relation shown in figure 1(b).

By virtue of statement 1 of the recursive definition, the triangles s_1 and s_2 (redrawn in figure 2) are shapes in the class.

Additional shapes in the class are produced by applying statement (2) of the recursive definition to shapes already known to be in the class. Statement (2) allows for a triangle to be added to a shape in two ways. If a triangle occurs in a shape and it is similar to s_1 (or alternatively to s_2), then another triangle similar to s_2 (or alternatively to s_1) can be added to the shape so that the triangles have the spatial relation specified by the set of shapes $\{s_1, s_2\}$. Because s_1 and s_2 are similar triangles, statement (2) can be applied to any shape produced by a previous application of statement (2) to produce another shape. Every shape produced by an application of statement (2) to a shape in the class is also a shape in the class.

Figure 2 shows some additional shapes in the class. The shapes s_3, s_4, s_5, and s_6 result from the application of statement (2) to the shapes s_1, s_3, s_4, and s_5 respectively. Each triangle added in this potentially infinite series is progressively larger. The shapes s_3, s_7, s_8, and s_9 result from the application of statement (2) to the shapes s_2, s_3, s_7, and s_8 respectively. Each triangle added in this potentially infinite series is progressively smaller. Notice that each shape in this series can be inscribed in a triangle similar to s_2 with a longest edge equal to $c/[1-(\sin B/\sin A)^2]$. The shape s_{10} results from the application of statement (2) to either the shape s_4 or s_7. A smaller triangle is added to s_4 or a larger triangle is added to s_7 to obtain s_{10}.

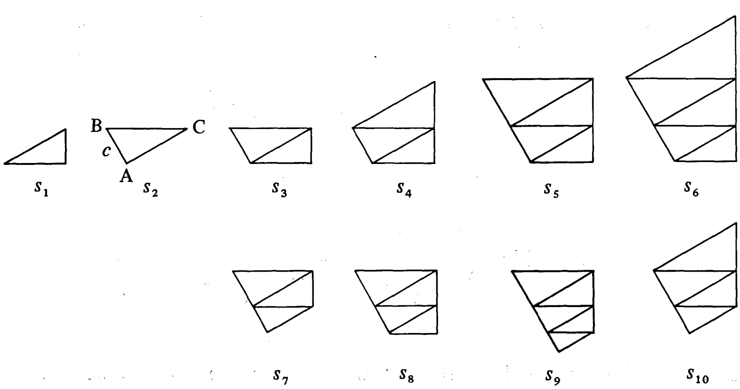


Figure 2. Some shapes in the class defined recursively using the spatial relation specified by the set of shapes $\{s_1, s_2\}$.

Any two triangles occurring in a shape in this class and that share an edge have the spatial relation specified by the set of shapes $\{s_1, s_2\}$. The class contains infinitely many different shapes. In general, an infinite class of shapes is determined by the recursive definition when it is used in conjunction with a spatial relation specified by a set of shapes in which a shape is similar to another shape or to part of another shape.

The recursive definition determines different classes of shapes when used in conjunction with different spatial relations. For example, figure 3 shows some of the shapes in the class determined when S is the set of shapes $\{t_1, t_2\}$, where t_1 and t_2 are the triangles drawn in figure 1(c), and t_1 and t_2 have the spatial relation shown in figure 1(d). Shapes in this class are 'spirals' made up of triangles. Shapes in the potentially infinite series t_3, t_4, t_5, t_6 are produced by adding progressively larger triangles to t_1, t_3, t_4, and t_5 respectively, in the potentially infinite series t_3, t_7, t_8, t_9 by adding progressively smaller triangles to t_2, t_3, t_7, and t_8 respectively. The shape t_{10} is obtained by adding a smaller triangle to t_4 or a larger triangle to t_7. Any two triangles occurring in a shape in the class and that share an edge have the spatial relation specified by the set of shapes $\{t_1, t_2\}$.

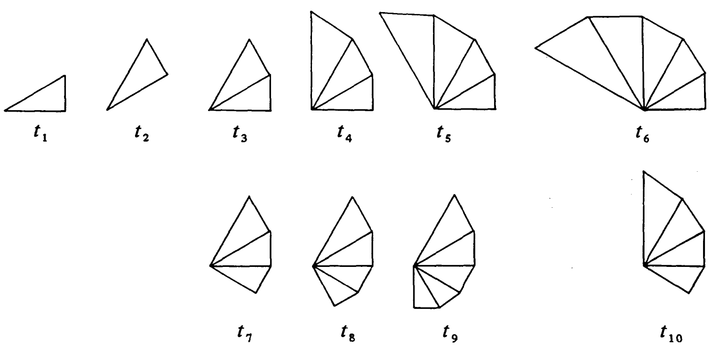


Figure 3. Some shapes in the class defined recursively using the spatial relation specified by the set of shapes $\{t_1, t_2\}$.

2.3 *Shape grammars*

The method given in section 2.2 for the recursive definition of classes of shapes in terms of spatial relations is generalized and refined in the definition of shape grammars. Shape grammars are like phrase structure grammars (Chomsky, 1963). A phrase structure grammar is defined over an alphabet of symbols, and maps strings of symbols into strings of symbols to generate a language of strings of symbols; a shape grammar is defined over a set of shapes, and maps shapes into shapes to generate a language of shapes.

A *shape grammar* (Stiny and Gips, 1972) has four parts:

(1) V_T is a finite set of shapes.

(2) V_M is a finite set of shapes.

(3) R is a finite set of *shape rules* of the form $u \rightarrow v$, where u and v are shapes made up of shapes in V_T or shapes in V_M. The shape u must have a subshape made up of shapes in V_M.

(4) I is a shape made up of shapes in V_T or shapes in V_M. The shape I must have a subshape made up of shapes in V_M.

A shape grammar S is given by the 4-tuple: $S = \langle V_T, V_M, R, I \rangle$. Shapes made up of shapes in V_T are called *terminals*. Shapes made up of shapes in V_M are called *markers*. No subshape of a terminal is a subshape of a marker. This condition ensures that terminals and markers in shapes can be distinguished uniquely. For a shape rule $u \rightarrow v$ in the set R, u is called the *left side* of the shape rule; v is called the *right* *side* of the shape rule. Together the left side and the right side of a shape rule specify a spatial relation. The correspondence between the two sides of the shape rule that determines this spatial relation is established by the coincidence of the coordinate systems in which the shapes of the two sides are drawn. The shape I is called the *initial shape*.

A shape is generated by a shape grammar $S = \langle V_T, V_M, R, I \rangle$ by beginning with the initial shape I and applying the shape rules in the set R until no shape rule can be applied. A shape rule $u \rightarrow v$ applies to a shape s if and only if there is a Euclidean transformation G such that G(u) is a subshape of s. The result of applying the shape rule $u \rightarrow v$ to the shape s under the Euclidean transformation G is the shape produced by replacing the occurrence of G(u) in s with G(v)[4]. The shape generation process terminates when no shape rule in the set R can be applied. The *language* defined by a shape grammar is the set of shapes generated by the shape grammar that are made up of terminals or parts of terminals only.

Figure 4(a) shows a shape grammar that generates some of the shapes in the class determined by the recursive definition, given in section 2.2, when S is the set of shapes $\{s_1, s_2\}$, where s_1 and s_2 are the triangles drawn in figure 1(a), and s_1 and s_2 have the spatial relation shown in figure 1(b). In this shape grammar, the set V_T contains the triangle s_1. All shapes in the language defined by the shape grammar are made up of triangles similar to s_1. The set V_M contains a circle. Shapes made up of shapes in V_T (terminals) and shapes made up of shapes in V_M (markers) are distinguishable. The set R contains two shape rules. The first shape rule incorporates the spatial relation specified by the set of shapes $\{s_1, s_2\}$. The left side of this shape rule consists of the triangle s_1 with a circle inscribed in it; the right side consists of the triangle s_1 and the similar triangle s_2 with a circle inscribed in it. Essentially the shape rule states that if a triangle, similar to s_1, with a circle inscribed in it occurs in a shape, then the circle can be erased and a larger triangle, similar to s_1, with a circle inscribed in it can be added to the shape so that the triangles have the spatial relation specified by the set of shapes $\{s_1, s_2\}$. Markers are used in this shape rule to restrict its possible application and to guide the shape generation process. In this shape grammar, the terminals and the markers in the shape rule prevent it from applying at any place to which it has already been applied. The triangle with a circle inscribed in it in the left side of the shape rule is similar to the larger triangle with a circle inscribed in it in the right side. Hence, under the conditions for shape-rule application, this shape rule can be applied to any shape to which it has been previously applied.

$S = \langle V_T, V_M, R, I \rangle$

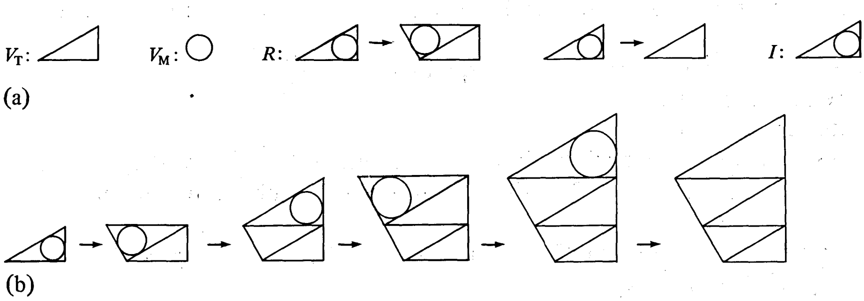


Figure 4. A shape grammar and the generation of a shape.

[4] An algorithm for shape-rule application is given in Stiny (1975, pages 212–225). Given a shape s and a shape rule $u \rightarrow v$, the algorithm determines a finite set of Euclidean transformations that make u a subshape of s. The shape rule $u \rightarrow v$ applies to the shape s under a Euclidean transformation G if and only if G is equivalent to some Euclidean transformation in the set.

The second shape rule is required to terminate the shape generation process. The left side of this shape rule consists of the triangle s_1 with a circle inscribed in it; the right side consists of s_1 only. Essentially the shape rule states that if a triangle, similar to s_1, with a circle inscribed in it occurs in a shape, then the circle can be erased. In this shape grammar, once this shape rule has been applied, no other shape rule can be subsequently applied. The initial shape consists of the triangle s_1 with a circle inscribed in it.

The generation of the shape s_6, drawn in figure 2, by means of this shape grammar is shown in figure 4(b). The first shape rule is applied to the initial shape in step 1. In step 2, the first shape rule is applied to the triangle with a circle inscribed in it added in step 1; in step 3, to the triangle with a circle inscribed in it added in step 2; and in step 4, to the triangle with a circle inscribed in it added in step 3. The second shape rule is applied in step 5 to erase the circle in the shape in step 4. The shape generation process terminates with the shape produced in step 5. This shape is made up of terminals only and hence is a shape in the language defined by the shape grammar.

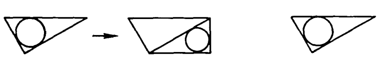
The shapes s_1, s_3, s_4, s_5, and s_6 drawn in figure 2 are in the language defined by the shape grammar given in figure 4(a). Any shape in this potentially infinite series can be generated by this shape grammar.

By replacing the first shape rule and the initial shape in the shape grammar given in figure 4(a) with the shape rule and initial shape shown in figures 5(a) and (b) respectively, a new shape grammar is obtained that generates the shapes s_2, s_3, s_7, s_8, and s_9 drawn in figure 2. The left side of this new shape rule consists of the triangle s_2 with a circle inscribed in it; the right side consists of the triangle s_2 and the similar triangle s_1 with a circle inscribed in it. Essentially the shape rule states that if a triangle, similar to s_2, with a circle inscribed in it occurs in a shape, then the circle can be erased and a smaller triangle, similar to s_2, with a circle inscribed in it can be added to the shape so that the triangles have the spatial relation specified by the set of shapes $\{s_1, s_2\}$. The new initial shape consists of the triangle s_2 with a circle inscribed in it. This shape grammar also defines a language containing infinitely many different shapes.

By adding the shape rule given in figure 5(a) to the shape grammar given in figure 4(a), a new shape grammar is obtained. This shape grammar generates all the shapes except s_2 in the class determined by the recursive definition of section 2.2, when S is the set of shapes $\{s_1, s_2\}$. By replacing the initial shape of this new shape grammar with the initial shape given in figure 5(b), another shape grammar is obtained. This shape grammar generates all the shapes except s_1 in the class.

Another shape grammar is shown in figure 6(a). This shape grammar generates all the shapes except t_2 in the class determined by the recursive definition of section 2.2, when S is the set of shapes $\{t_1, t_2\}$, where t_1 and t_2 are the triangles drawn in figure 1(c), and t_1 and t_2 have the spatial relation shown in figure 1(d). The generation of the shape t_{10}, drawn in figure 3, using this shape grammar is shown in figure 6(b). Additional shapes in the language defined by this shape grammar are drawn in figure 3.

Shape grammars can be defined to generate curvilinear as well as rectilinear shapes, and three-dimensional as well as two-dimensional shapes. A variety of other shape grammars, and techniques for defining shape grammars, are given in Stiny (1975).



(a) (b)

Figure 5. A shape rule and an initial shape.

Shape grammars are examined as a means for performing a three-dimensional perceptual task in Gips (1975), and in terms of their algebraic and computational properties in Stiny (1975).

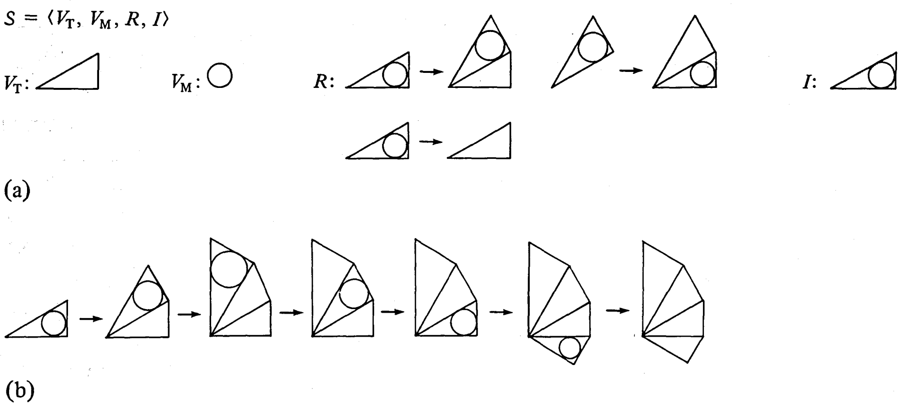


Figure 6. A shape grammar and the generation of a shape.

3 Two exercises in formal composition

In this section, the use of shape grammars as the basic tool for the treatment of two general exercises in formal composition is sketched. The presentation is intended to be suggestive rather than definitive.

3.1 *Exercise 1*

A beguilingly simple and architecturally interesting exercise in formal composition consists of exploring the possibilities for arranging certain spatial elements in some ordered way. To cite a classical example,

> "Leonardo wanted to know in a general way what forms he could give to the central-plan church, and set about systematically to find the answer. He realized that if he began with the simplest spatial forms (square, octagon, circle, or dodecagon), he would arrive at every conceivable central-plan church ... by the mechanical addition of circular, semicircular, square, rectangular, or octagonal ancillary spaces to the principal and cross axes of his basic figures. A complete series of related central-plan churches could be developed from a basic schema. For example, he could begin with a Greek cross (four square arms added to the sides of a central square), and then either replace the square space by an octagon, a circle, or a dodecagon, or replace the square arms with rectangles, octagons, circles, semicircles, or dodecagons" (Frankl, 1914, pages 5–6).

A generative approach to problems of this kind is illustrated by considering the possibilities for constructing shapes consisting of edge-connected polygons similar to polygons in a given finite vocabulary (collection). Special attention is given to vocabularies containing exactly one polygon and hence to shapes made up of similar, edge-connected polygons.

This exercise may be regarded as a generalization of questions about the enumeration and geometric classification of cell configurations, for example, of polyominoes. Indeed, when the vocabulary contains a unit square, the exercise consists exactly of exploring the population of polyominoes. The architectural relevance of this special exercise and, by natural extension, of the general exercise is demonstrated, for example, by March and Matela (1974), Mitchell and Dillon (1972), and Rittle (1968).

In general terms, the approach has three stages. First, a finite vocabulary of primitive shapes is given. This vocabulary fixes the spatial elements that are to be used to make other shapes. Second, the distinct spatial relations that govern the joint occurrence of these shapes in a legally constructed shape are enumerated. These spatial relations are specified using the shapes in the vocabulary. Third, the vocabulary and spatial relations are used to define shape grammars. These shape grammars generate shapes made up of shapes in the vocabulary in accordance with the spatial relations.

Consider a finite vocabulary of polygons. Shapes are to be constructed by linking polygons via their edges in accordance with this proviso:

The construction of a shape begins with a polygon in the vocabulary. A polygon can be added to a shape under construction when the polygon is similar to a polygon in the vocabulary and shares at least one of its edges with a polygon already added to the shape.

The occurrence of polygons in a legally constructed shape is governed by spatial relations specified by sets $\{s_i, s_j\}$, where s_i and s_j are polygons similar to polygons in the vocabulary, and s_i and s_j have at least one edge in common. Polygons in these sets can be nonsimilar or similar (but not pictorially equivalent). Hence, if the vocabulary contains n nonsimilar polygons, there are $n(n+1)/2$ pairs of polygons that can be used to specify spatial relations of this type. Each pair of polygons can be used to specify only a finite number of distinct spatial relations.

In general, four distinct spatial relations can be specified using two nonsimilar polygons that have a fixed edge in common. Imagine two nonsimilar polygons that share an edge x in a certain way. In addition to this way, the polygons share the edge x when the reflection of one about the edge x is taken, when the reflection of the reflection of one about the edge x about the perpendicular bisector of the edge x is taken, and when the reflection of one about the perpendicular bisector of the edge x is taken. These additional cases are illustrated in figures 7(c)–(e) respectively for the quadrilateral and triangle drawn in figure 7(a) that have a common edge x as shown in figure 7(b).

Only two distinct spatial relations can be specified using two nonsimilar polygons that have a fixed edge in common, when either of these polygons has bilateral symmetry with respect to the perpendicular bisector of the edge. These cases are illustrated in figures 7(g) and (h) for the quadrilateral and triangle drawn in figure 7(f).

Let an edge of one polygon be *equivalent* to an edge of a second, similar polygon if and only if these edges coincide when the polygons are made pictorially equivalent.

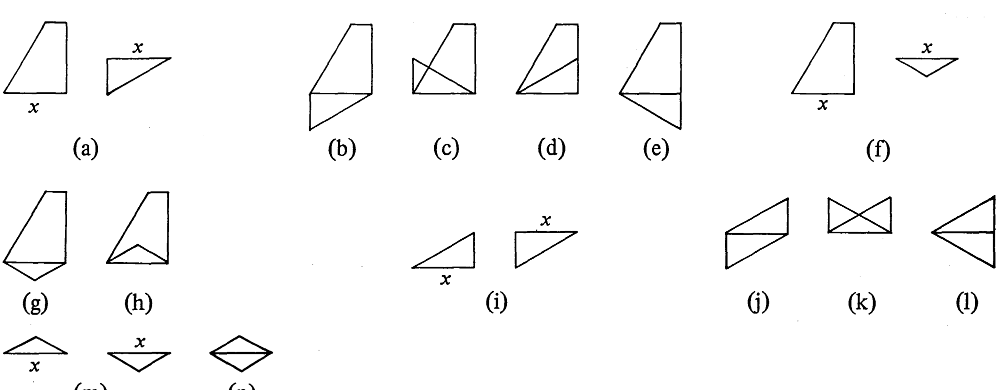


Figure 7. Spatial relations specified using polygons that share a fixed edge.

In general there are three distinct spatial relations that can be specified using two similar but not pictorially equivalent polygons that have a fixed equivalent edge in common. These cases are illustrated in figures 7(j)–(l) for the similar triangles drawn in figure 7(i).

Only one distinct spatial relation can be specified using two similar polygons that are not pictorially equivalent but have a fixed, equivalent edge in common when each of these polygons has bilateral symmetry with respect to the perpendicular bisector of the edge. This case is illustrated in figure 7(n) for the similar triangles drawn in figure 7(m).

It is easy to see that if two polygons are nonsimilar, then they can be used to specify at most $4mn$ distinct spatial relations, where m is the number of edges in the first polygon and n is the number of edges in the second polygon. If two polygons are similar, then they can be used to specify at most $(2n^2+n)$ distinct spatial relations, where n is the number of edges in each of the polygons. In those cases where the polygons share more than one edge at a time, or either of the polygons has edges that can be mapped into one another by a Euclidean transformation so that the polygon coincides with itself, less than these numbers of distinct spatial relations can be specified.

There are forty-eight distinct spatial relations that can be specified using the quadrilateral and triangle drawn in figure 7(a) so that they have at least one edge in common. The twenty-one distinct spatial relations specified using the two similar triangles drawn in figure 7(i) so that they have at least one edge in common are shown in figure 8.

A given finite vocabulary of shapes and a given class of distinct spatial relations specified using the shapes in the vocabulary can be used to define shape grammars[5]. A shape grammar $S = \langle V_T, V_M, R, I \rangle$ can be defined in this way:

(1) Let V_T be the given vocabulary of shapes.

(2) Let V_M be a finite set of shapes such that shapes made up of shapes in V_T and shapes made up of shapes in V_M are distinguishable.

(3) Let R be a finite set of shape rules that are formed using some of the spatial relations in the given class. These spatial relations are specified in terms of shapes in the given vocabulary and hence in terms of terminals in the shape grammar. For a spatial relation specified by the set of shapes S, a shape rule can be defined by treating some of the shapes in S as terminals in the left side of the shape rule and associating markers with them, and treating others as terminals in the right side of the shape rule and associating markers with them. The terminals in the left side and the right side of the shape rule incorporate the spatial relation specified by S, and hence provide for the replacement of terminals in a shape in accordance with this spatial relation. Markers are associated with the terminals in the left side and the right side of the shape rule in order to control, guide, or disambiguate the shape generation process. Markers may be thought of as encoding computational details important for the shape generation process. The markers in the left side restrict the immediate application of the shape rule; the markers in the right side restrict the subsequent application of other shape rules. In general, the occurrence of these markers depends on the actual shapes used to specify the spatial relation incorporated in the shape rule. More precisely, the left side of the shape rule is the shape union of some of the shapes in S, that is, shapes in V_T, and a shape made up of shapes in V_M; the right side is the shape union of some of the shapes in S, that is, shapes in V_T, and a shape made up of shapes in V_M. All the shapes in S occur in the left side or the

(5) An important empirical component of a science of formal composition would be a catalogue of spatial relations, and heuristics for specifying spatial relations, that can be used to define shape grammars that generate interesting shapes. Several types of spatial relations different from those discussed in this paper, and the shape grammars that can be defined in terms of them, are characterized in Stiny (1975, pages 66–105).

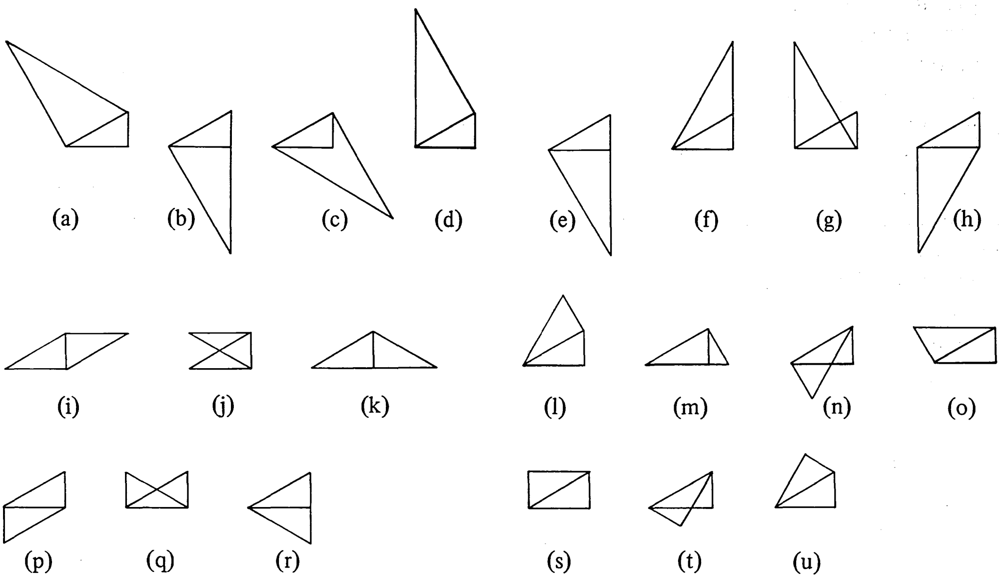


Figure 8. The twenty-one distinct spatial relations specified using the similar triangles drawn in figure 7(i).

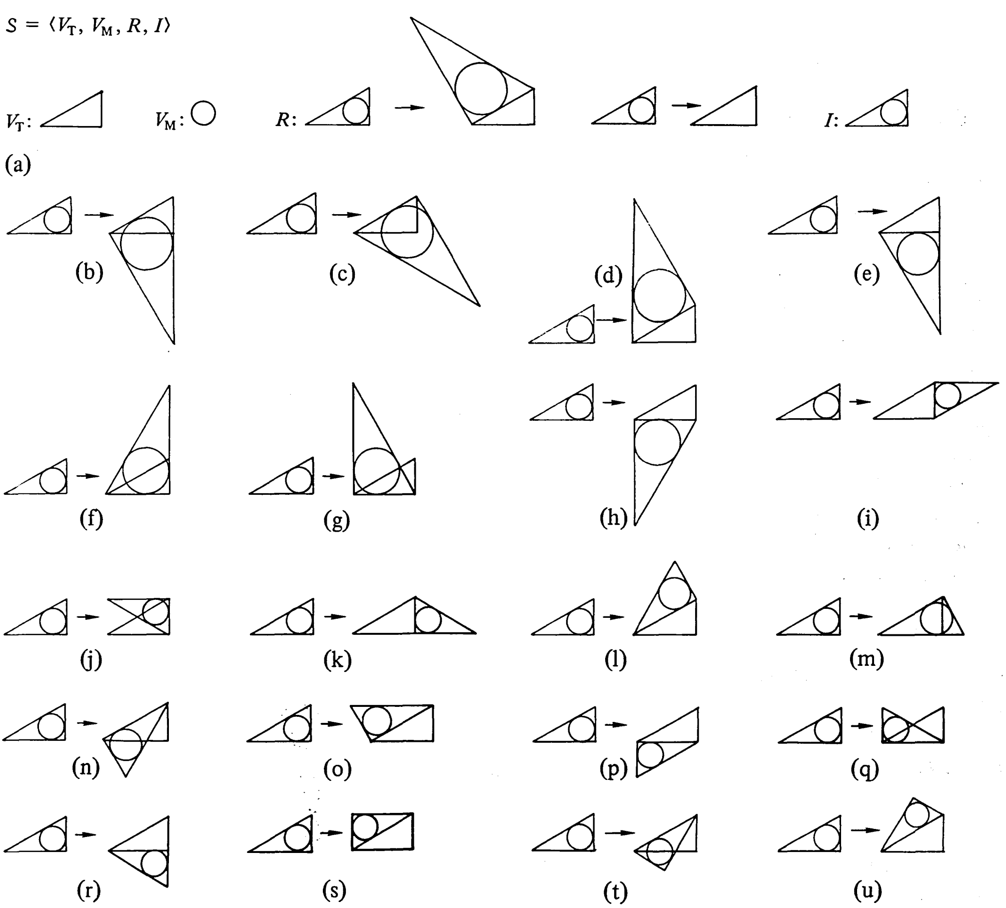


Figure 9. A shape grammar and shape rules defined in terms of the spatial relations of figure 8.

right side of the shape rule. Any spatial relation in the class may be used to define several different shape rules in this way. In addition to these shape rules, let R contain shape rules that allow for markers to be erased. The left sides of these shape rules are either left sides or right sides of shape rules defined in the above way; the right sides consist of only the terminals in their left sides.

(4) Let I be the shape union of some of the left sides of shape rules in R.

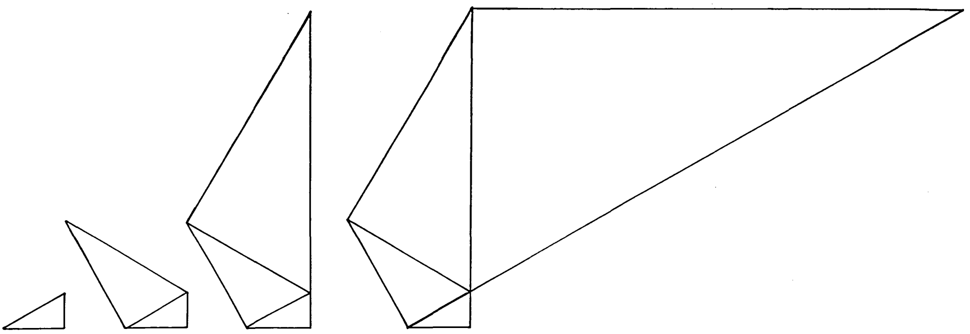
A shape grammar defined in this way provides for the generation of shapes made up of shapes in the given vocabulary in accordance with some of the spatial relations in the given class. The language defined by this shape grammar is never empty, as markers in the initial shape can always be erased to produce a shape in the language. In general, if the left sides of some of the shape rules in the shape grammar are similar to parts of the right sides of other shape rules, the shape grammar can define an infinite language of shapes.

This method for defining shape grammars is made clear by an example: Consider the class of twenty-one distinct spatial relations shown in figure 8. These spatial relations are specified in terms of a vocabulary containing a single triangle. A shape grammar can be defined for each one of the spatial relations in this class.

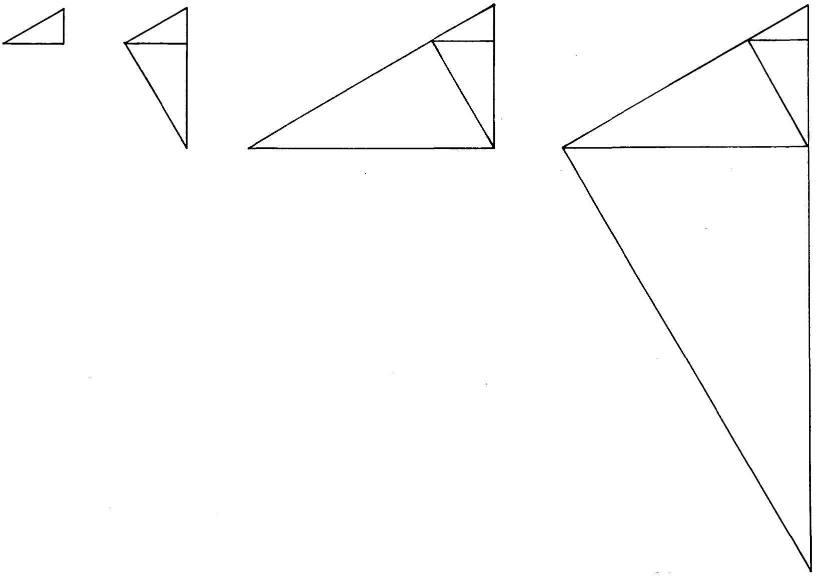
The shape grammar defined using the spatial relation of figure 8(a) is shown in figure 9(a). In this shape grammar, the set V_T contains a triangle, that is, V_T is the given vocabulary. The set V_M contains a circle. The set R contains two shape rules. The left side of the first shape rule consists of the smaller triangle used to specify the spatial relation with a circle inscribed in it. The left side is the shape union of a shape used to specify the spatial relation and a shape in V_M. The right side of the first shape rule consists of both triangles used to specify the spatial relation with a circle inscribed in the larger triangle. The right side is the shape union of both shapes used to specify the spatial relation, and a shape in V_M. Both shapes used to specify the spatial relation occur in the left side or the right side of the first shape rule. The left side of the second shape rule is the left side of the first shape rule; the right side consists of the triangle (terminal) in the left side. The initial shape I is the left side of the first shape rule. This shape grammar has the same general properties as the shape grammar given in figure 4(a). [By replacing the first shape rule of this shape grammar with the shape rule of figure 9(o), the shape grammar of figure 4(a) is obtained.] Some shapes in the language defined by the shape grammar are drawn in figure 10(a).

Twenty other shape grammars of this kind can be defined using the twenty other spatial relations shown in figure 8. These shape grammars are obtained by replacing the first shape rule in the shape grammar of figure 9(a) with one of the shape rules shown in figures 9(b)–(u). The letter under each shape rule corresponds to the spatial relation of figure 8 used to define it. The left side of each of these shape rules consists of the smaller triangle used to specify the spatial relation incorporated in the shape rule with a circle inscribed in that triangle; the right side consists of both triangles used to specify the spatial relation with a circle inscribed in the larger triangle[6]. Notice that in those cases where the two similar triangles in the right side of a shape rule defined in this way form a third similar triangle, shape rules (b), (e), and (m), the circles in the left side and the right side of the shape rule determine unambiguously the actual triangles used. Some shapes in the languages defined by the shape grammars corresponding to the shape rules of figures 9(b)–(u) are drawn in figures 10(b)–(u) respectively.

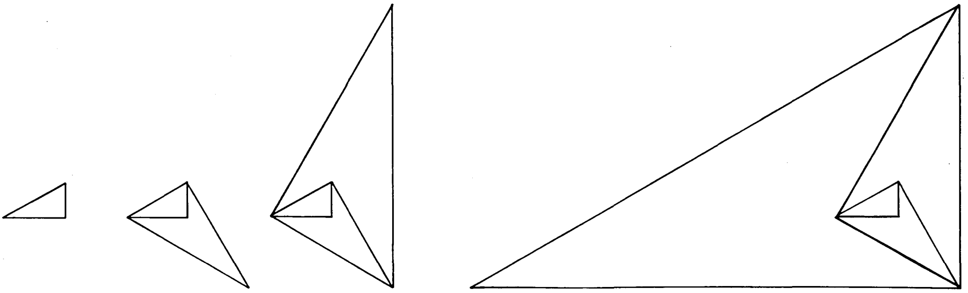
[6] The spatial relations of figures 8(i)–(k) and 8(p)–(u) are specified using triangles of the same size. The triangle with a circle inscribed in it in the right sides of the shape rules corresponding to these spatial relations is the triangle not in the left sides.



(a)

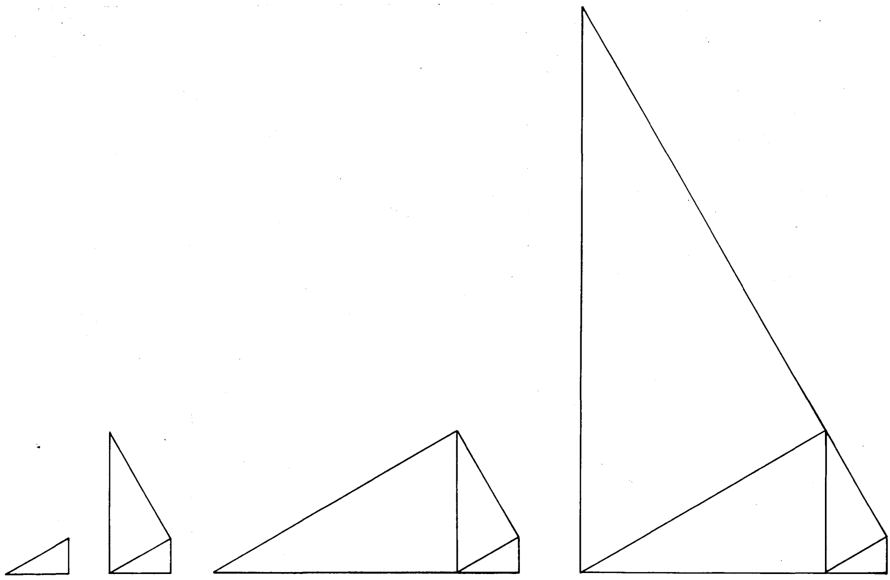


(b)

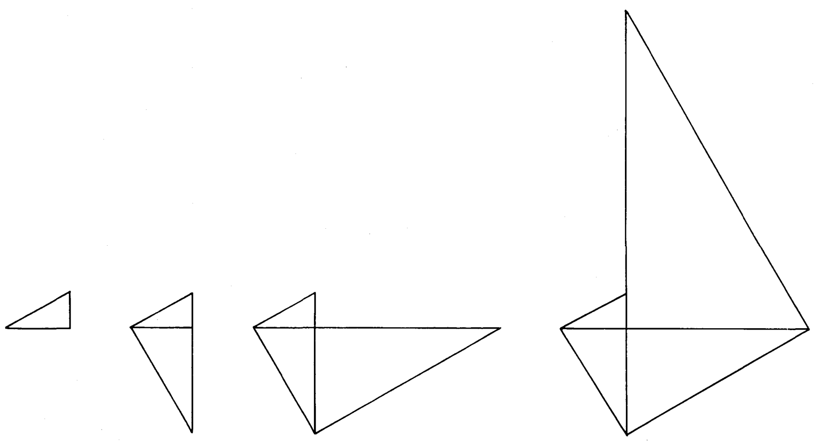


(c)

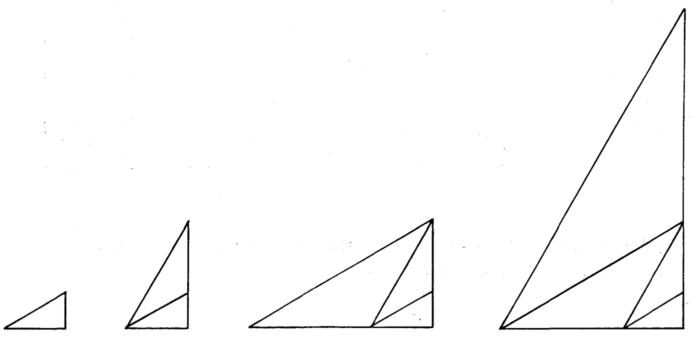
Figure 10. Some shapes in the languages defined by the shape grammar of figure 9(a) and the shape grammars corresponding to the shape rules of figures 9(b)–(u).



(d)



(e)



(f)

Figure 10 (continued)

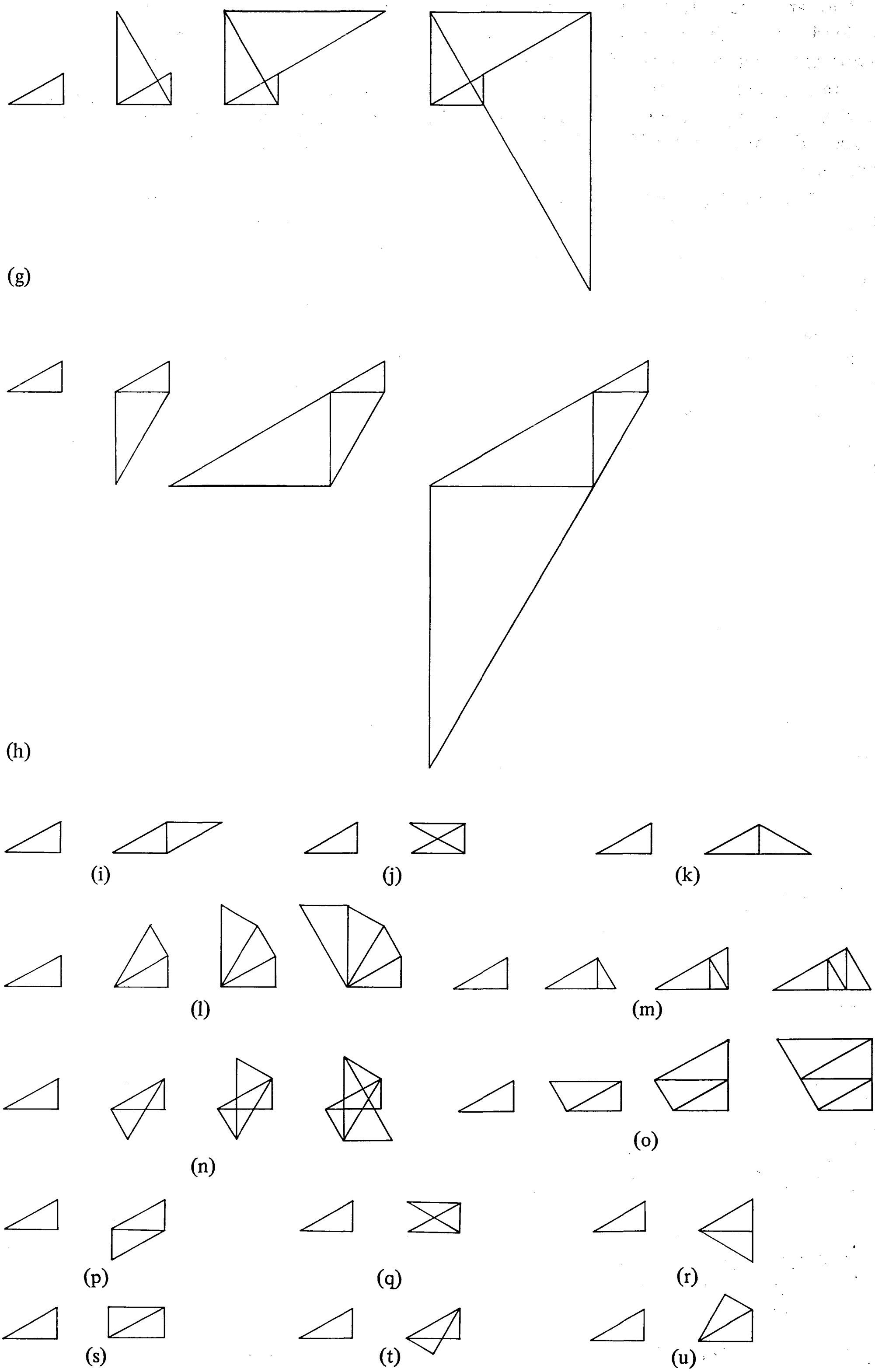


Figure 10 (continued)

The language defined by the shape grammar of figure 9(a) and the languages defined by the shape grammars obtained by replacing the first shape rule of this shape grammar with one of the shape rules of figures 9(b)–(u) are all distinct. In general, shape grammars defined using distinct spatial relations define distinct languages. Notice that only the languages defined by the shape grammars corresponding to the shape rules of figures 10(i)–(k) and (p)–(u) are finite. All others define infinite languages.

Let two shape grammars be *similar* if and only if every shape in the language defined by the first is similar to some shape in the language defined by the second, and every shape in the language defined by the second is similar to some shape in the language defined by the first. For each shape grammar defined above using a spatial relation of figure 8, there is a similar shape grammar defined using the spatial relation. This shape grammar contains two shape rules. The left side of the first shape rule consists of the larger triangle used to specify the spatial relation with a circle inscribed in it; the right side consists of both triangles used to specify the spatial relation with a circle inscribed in the smaller triangle[7]. The left side of the second shape rule is the left side of the first shape rule; the right side consists of the triangle in the left side. The initial shape is the left side of the first shape rule.

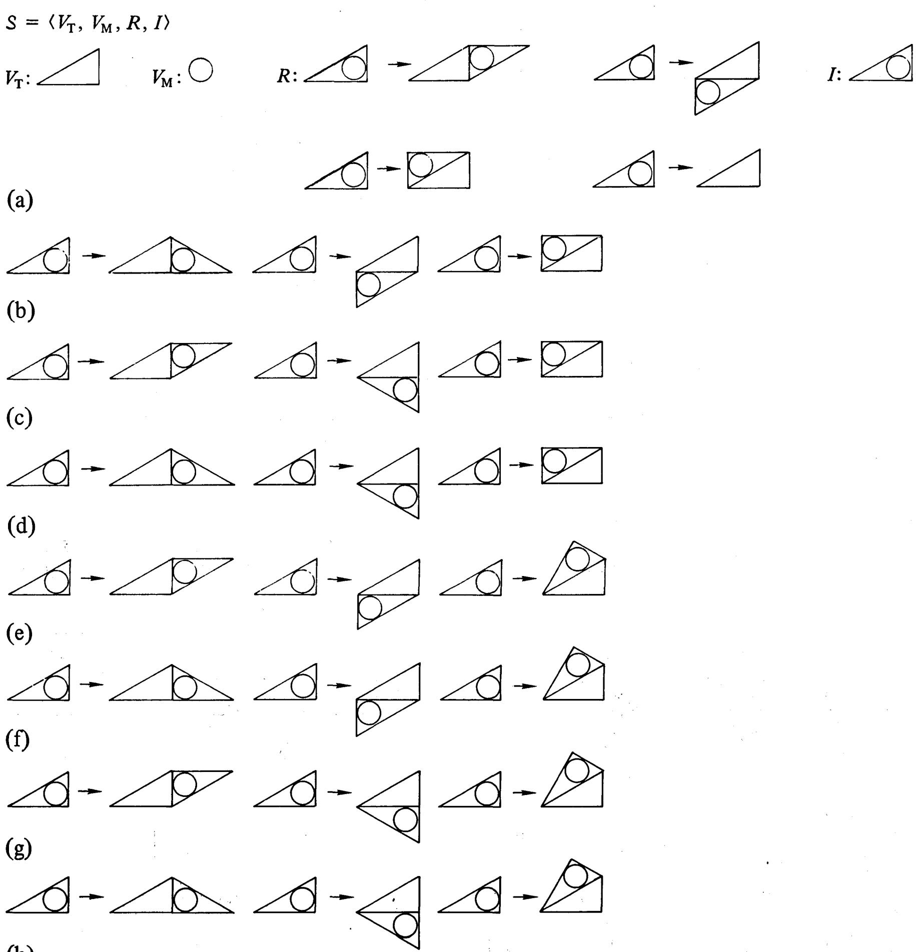


Figure 11. A shape grammar and shape rules generating triangular tessellations.

[7] See footnote (6).

Shape grammars containing more than two shape rules can also be defined using the spatial relations shown in figure 8. Of special interest are eight distinct shape grammars each containing four shape rules: either the shape rule of figure 9(i) or (k), either the shape rule of figure 9(p) or (r), either the shape rule of figure 9(s) or (u), and the shape rule having the left side of any of these shape rules as its left side and a right side consisting of only the triangle in the left side. One of these shape grammars is shown in figure 11(a). The other seven shape grammars of this kind are obtained by replacing the first three shape rules of the shape grammar of figure 11(a) with one of the seven shape rule combinations shown in figures 11(b)–(h). The six possible edge-connected, triangular tessellations drawn in figures 12(a)–(f) are generated by the shape grammar of figure 11(a) and the shape grammars obtained from this shape grammar using the shape rule combinations of figures 11(b)–(e) and (h) respectively. This outcome is not too surprising as the spatial relations of figures 8(i) and (k), (p) and (r), and (s) and (u) used to define the shape rules of figures 9(i) and (k), (p) and (r), and (s) and (u) respectively govern the way two similar, scalene triangles of the same size can share an edge without overlapping. The reader is invited to explore the more difficult-to-characterize languages defined by the two shape grammars obtained from the shape grammar of figure 11(a) using the shape rule combinations of figures 11(f) and (g).

Of course, a large variety of other shape grammars can be defined using the spatial relations of figure 8. In fact there are $2^{21}-1$ (= 2097151) distinct shape grammars that can be defined using the first shape rule in the shape grammar of figure 9(a) and the shape rules of figures 9(b)–(u). All of these shape grammars define distinct languages of shapes that are legally constructed in the sense of the proviso given at the beginning of this section.

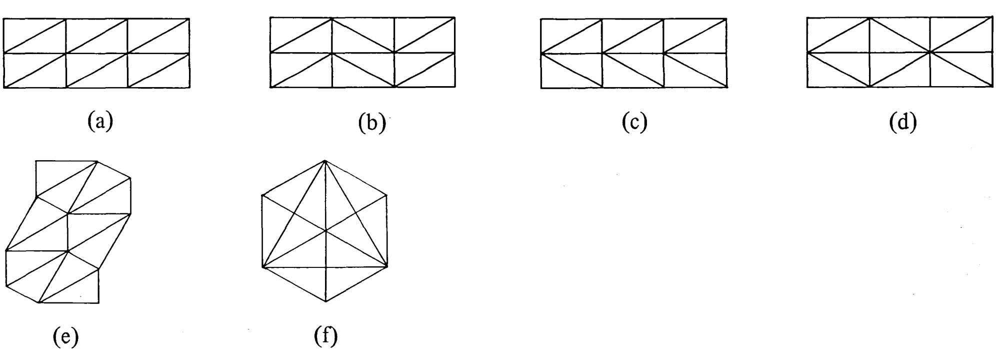


Figure 12. Six triangular tessellations.

3.2 *Exercise 2*

Another exercise in formal composition consists of beginning with a given arrangement or arrangements of certain spatial elements and constructing or identifying additional arrangements of these elements that are in the same *style*. To adapt the classical example of section 3.1, suppose we are given one of Leonardo's forms for a central-plan church and asked to complete the series of related central-plan churches. We could first decompose the given form into its principal and ancillary elements and then recombine the ancillary elements with the principal one in as many new ways as possible. Our completion of the series would depend on what we inferred the principal and ancillary elements of the given form to be and what we inferred legal combinations of these elements to be. For example, if a Greek cross occurs in the given form, we could take it as the principal element, or its central square as the principal element and its arms as some of the ancillary ones.

A generative approach to problems of this kind is illustrated by considering the possibilities for defining simple shape grammars that generate a given shape made up of polygons, and additional shapes made up of polygons as well[8]. The assumption here is that two shapes are in the same style when they are generated by the same shape grammar. More intuitively, two shapes are in the same style when they are made up of shapes in a common vocabulary, and the combinations of these shapes are governed by the same spatial relations. In general this notion of style is probably too strict, but in the present context it seems to be a natural, formal counterpart of the standard, informal notions of style used by historians of architecture, such as Ackerman (1962) and Summerson (1963).

In general terms the approach has four stages. First, a shape (or shapes) is given. The style of this shape is to be inferred. Second, the vocabulary of primitive shapes occurring in the shape is determined by decomposing it. A decomposition of the shape specifies the primitive shapes of which it is seen to be made up. Third, the spatial relations that govern the joint occurrence of any two of these primitive shapes in the shape are specified. These spatial relations are specified in terms of the decomposition of the shape. Fourth, the vocabulary of primitive shapes and the spatial relations are used to define simple shape grammars that generate the given shape and other shapes as well. These shape grammars are defined using the ideas of section 3.1.

Consider a shape made up of polygons. Let a *decomposition* of this shape be any finite set of polygons such that each polygon in the set is a subshape of the shape, and the shape union of all the polygons in the set is pictorially equivalent to the shape. In general a shape made up of polygons has more than one decomposition. Each decomposition corresponds to a different way of looking at (parsing) the shape. For example, the shape drawn in figure 13(a) can be decomposed as shown in figures 13(b) and (c). In terms of the first decomposition, the shape is seen to be made up of five squares; in terms of the second, the shape is seen to be made up of two similar Ls.

For a shape made up of polygons, there is a finite number of decompositions. The decompositions of the shape can be enumerated by generating the set of all polygons in the shape and then checking to see which subsets of this set are decompositions. The set of all polygons in the shape can be generated using various obvious procedures. The twenty polygons occurring in the shape drawn in figure 13(a) are shown in figure 14.

Not all of the decompositions of a shape made up of polygons are of equal interest. The appropriateness of a decomposition depends on the purpose for which it is intended. Here a decomposition of a shape is to be used as the basis for defining the terminal vocabularies and the shape rules in shape grammars that generate the shape and other shapes as well. In order for a shape grammar to generate several (infinitely many) shapes, it is necessary that the left sides of some of its shape rules be similar

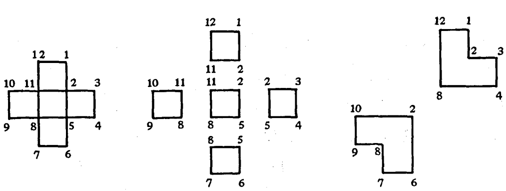


Figure 13. A shape and two of its possible decompositions.

[8] In this sense, the exercise corresponds, more or less, to the grammatical inference problem which can be stated as follows: Given a finite corpus of sentences in a potentially infinite language, construct or find the simplest grammar that defines (generates) the language. A review of the grammatical inference problem and its relevance to syntactic pattern recognition is given in Fu (1974, pages 193–229).

to parts of the right sides of some other of its shape rules. Since the sides of shape rules will be made up of shapes in the terminal vocabulary of the shape grammar, the following heuristic is suggested:

(1) Each polygon in a decomposition should be similar to as many other polygons in the decomposition as possible. At the very least the decomposition should contain two polygons that are similar.

In addition to this basic heuristic for selecting the most appropriate decompositions of a shape made up of polygons, these further heuristics are sometimes useful:

(2) No two polygons in the decomposition should overlap.

(3) The decomposition should contain the fewest number of nonsimilar polygons with the fewest number of edges.

Heuristic (2) is apropos when the shape is considered to be a floor plan; heuristic (3) is apropos when the shape is considered perceptually.

The decompositions of the shape drawn in figure 13(a) that are most appropriate in the sense of heuristics (1) and (2) are shown in figures 15(a) and (b), in the sense of heuristics (1) and (3), in figures 15(a)–(d).

Once a decomposition of a shape is given, it can always be used in conjunction with the recursive definition of section 2.2 to determine a class of shapes that contains, at the very least, the given shape. That is, the decomposition itself is considered to specify a spatial relation that in turn provides the basis for the recursive definition of a class of shapes. For example, suppose the decomposition of figure 15(b) is given for the shape drawn in figure 13(a). Using the recursive definition of section 2.2, let the set S contain the five squares in this decomposition. Clearly the shape drawn in figure 13(a) is in the class determined by the recursive definition. The class contains infinitely many other shapes; three representative shapes in the class are shown in figure 16.

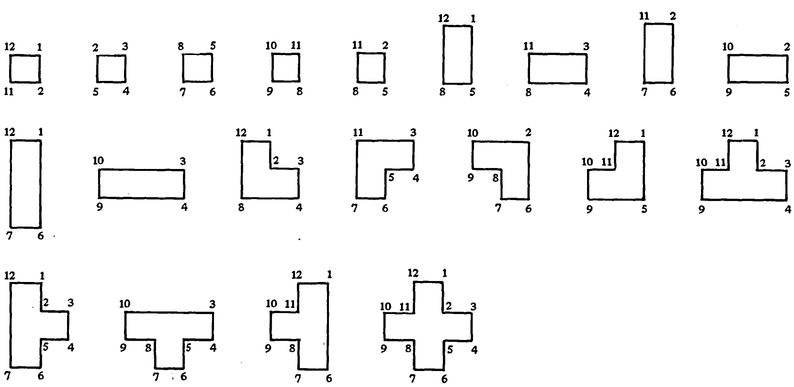
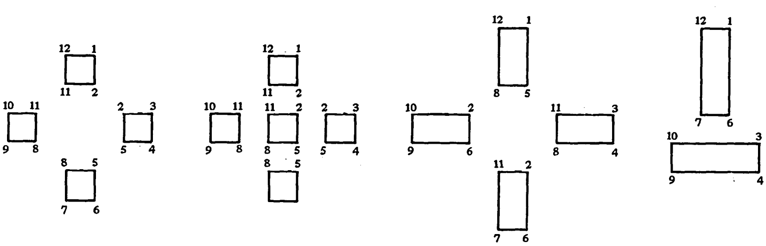


Figure 14. The twenty polygons in the shape drawn in figure 13(a).



(a) (b) (c) (d)

Figure 15. Decompositions of the shape drawn in figure 13(a).

A more subtle approach is to use the decomposition given for a shape made up of polygons to define the terminal vocabularies and shape rules of shape grammars.

Given a decomposition of a shape, the vocabulary of which the shape is made up is the least set of polygons such that every polygon in the decomposition is similar to some polygon in the set. The vocabularies corresponding to the decompositions of figures 15(a) and (b) both contain a square; the vocabularies corresponding to the decompositions of figures 15(c) and (d) contain a 1×2 rectangle and a 1×3 rectangle respectively.

Given a decomposition of a shape, the spatial relations that govern the joint occurrence of any two polygons in the shape are specified by subsets of the decomposition that contain exactly two polygons. There are two distinct spatial relations specified in this way using the decomposition shown in figure 15(a) of the shape drawn in figure 13(a). These spatial relations are shown in figures 17(a) and (b). There are three distinct spatial relations specified in this way using the decomposition shown in figure 15(b) of the shape drawn in figure 13(a). These spatial relations are illustrated in figures 17(a)–(c).

Once the vocabulary of polygons of which a shape is made up and the spatial relations that govern the joint occurrence of any two of these polygons in the shape are determined, the ideas of section 3.1 can be used to define shape grammars that generate the shape and other shapes as well. For example, the shape drawn in figure 13(a) may be considered to be made up of four squares so that any two squares in the shape have one of the two spatial relations shown in figures 17(a) and (b). This vocabulary and these spatial relations are used to define the shape grammar given in figure 18(a). Shapes in the language defined by this shape grammar are made up of squares. The first two shape rules of the shape grammar incorporate the spatial relations shown in figures 17(a) and (b); they provide for the generation of shapes in accordance with these spatial relations. The left side of the first shape rule consists of one square used to specify the spatial relation of figure 17(a) with a circle inscribed in it; the right side consists of both squares used to specify the spatial relation with a circle inscribed in the square not in the left side. Essentially the shape rule states that if a square with a circle inscribed in it occurs in a shape, then the circle can be erased and another square of the same size with a circle inscribed in it can be added to the shape so that the squares share exactly one vertex and their diagonals are colinear. The left side of the second shape rule consists of one square used to specify the spatial relation of figure 17(b) with a circle inscribed in it; the right side consists of both squares used to specify the spatial relation with a circle inscribed in the square not in the left side. Essentially the shape rule states that if a square with a circle inscribed in it occurs in a shape, then the circle can be erased and another

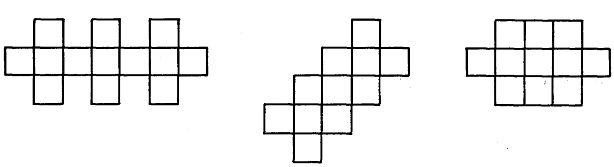
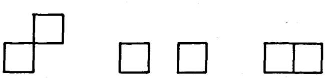


Figure 16. Three shapes in the class determined by the recursive definition of section 2.2, when used in conjunction with the spatial relation specified by the decomposition of figure 15(b).



(a) (b) (c)

Figure 17. Spatial relations specified by subsets of the decompositions of figures 15(a) and (b).

square of the same size with a circle inscribed in it can be added to the shape, such that the centers of the squares are separated by a distance equal to twice the length of their edges, and two edges of one square are colinear with two edges of the other square. The third shape rule of the shape grammar provides for the termination of the shape generation process by erasing markers (circles). The left side of this shape rule consists of a square with a circle inscribed in it; the right side consists of the square in the left side. Two possible generations of the shape drawn in figure 13(a) using this shape grammar are shown in figure 18(b). In the first generation of the shape, all three shape rules are applied; in the second generation, only the first and third shape rules are applied. Three additional shapes in the language defined by the shape grammar are drawn in figure 18(c). It is easy to see that this language contains infinitely many different shapes.

Alternatively the shape drawn in figure 13(a) may be considered to be made up of five squares so that any two squares in the shape have one of the three spatial relations shown in figures 17(a)–(c). This vocabulary and these spatial relations are used to define the shape grammar given in figure 19(a). Shapes in the language defined by this shape grammar are made up of squares. The shape grammar contains a shape rule incorporating the spatial relation of figure 17(c) in addition to the three shape rules in the shape grammar of figure 18(a). The left side of this new shape rule consists of one of the squares used to specify the spatial relation of figure 17(c) with a circle inscribed in it; the right side consists of both squares used to specify the spatial relation with a circle inscribed in the square not in the left side. Essentially the shape rule states that if a square with a circle inscribed in it occurs in a shape, then the circle can be erased and another square of the same size with a circle inscribed in it can be added to the shape so that the squares share an edge. Two possible generations of the shape drawn in figure 13(a) using this shape grammar are shown in figure 19(b). In the first generation of the shape, all four shape rules are applied; in the second generation, only the first and fourth shape rules are applied. Three additional shapes in the language defined by the shape grammar are drawn in figure 19(c).

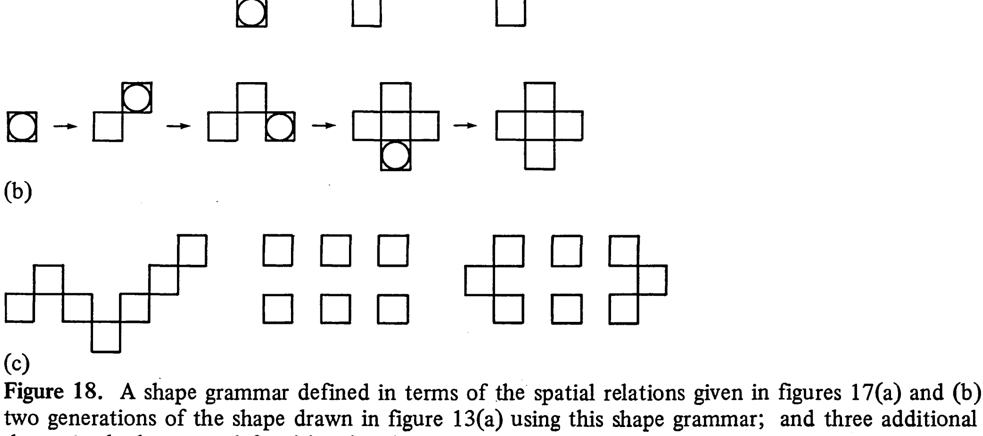


Figure 18. A shape grammar defined in terms of the spatial relations given in figures 17(a) and (b); two generations of the shape drawn in figure 13(a) using this shape grammar; and three additional shapes in the language defined by this shape grammar.

This language contains infinitely many different shapes. Notice that the language defined by the shape grammar of figure 18(a) is a proper subset of this language.

A shape grammar defined using only the spatial relation shown in figure 17(c) is given in figure 20(a). Shapes in the language defined by this shape grammar are made up of squares. The shape grammar contains the first and fourth shape rules in the shape grammar of figure 19(a), and an additional shape rule incorporating the spatial relation of figure 17(c). The left side of this shape rule consists of one of the squares used to specify the spatial relation with a circle inscribed in it; the right side consists of the other square used to specify the spatial relation with a circle inscribed in it. This shape rule provides for the translation of squares occurring in a shape. Essentially the shape rule states that if a square with a circle inscribed in it occurs in a shape, then the square and circle can be erased and another square of the same size with a

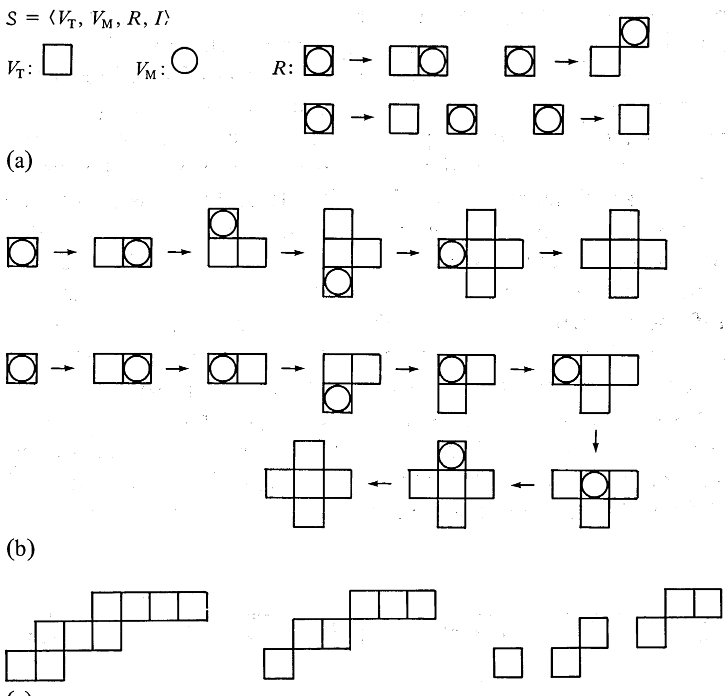


Figure 19. A shape grammar defined in terms of the spatial relations given in figures 17(a)-(c); two generations of the shape drawn in figure 13(a) using this shape grammar; and three additional shapes in the language defined by this shape grammar.

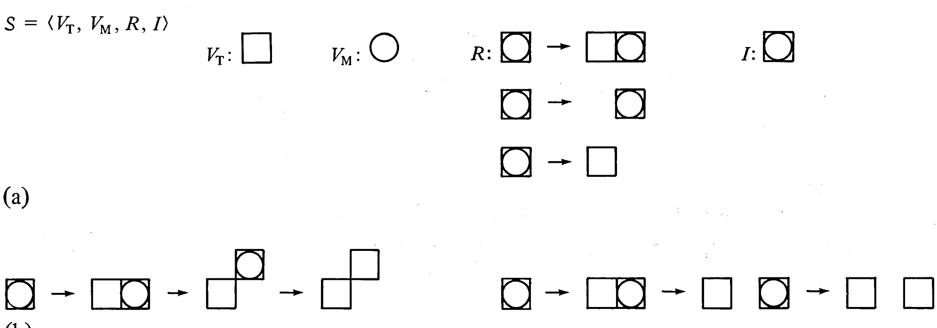


Figure 20. A shape grammar defined in terms of the spatial relation given in figure 17(c) and the generation of two shapes using this shape grammar.

circle inscribed in it can be added to the shape so that the squares would have shared an edge. The generation of two shapes in the language defined by this shape grammar is shown in figure 20(b). The squares in the first shape have the spatial relation shown in figure 17(a); the squares in the second shape have the spatial relation shown in figure 17(b). The languages defined by the shape grammars of figures 18(a) and 19(a) and the class of shapes determined by the recursive definition of section 2.2, when used in conjunction with the spatial relation specified by the decomposition of figure 15(b), are proper subsets of this language. The language defined by the shape grammar containing only the first and third shape rules in the shape grammar of figure 20(a) is the least set containing all possible polyominoes.

There are several other shape grammars that can be defined using the spatial relations of figure 17. The languages defined by these shape grammars are proper subsets of the language defined by the shape grammar of figure 20(a).

4 Conclusions

Shape grammars seem ideally suited as a basic research tool for the development of a science of form. Because shape grammars are defined in terms of shapes and not in terms of a fixed collection of primitives out of which all shapes must be tediously specified, they provide for the straightforward treatment of problems in formal composition in terms of the spatial elements that seem most natural to them. In particular, two-dimensional shape grammars can be used for floor plan composition, and three-dimensional shape grammars can be used for characterizing component building systems. More generally, shape grammars provide a foundation for a theory of architectural composition, as attempted by Durand (1823-1831) and Guadet (1902).

The educational value of shape grammars should be clear. Shape grammars supply inexhaustible material for teaching the six areas outlined by March and Matela (1974) and for teaching 'recursion' to visually oriented students.

Acknowledgement. I wish to thank Professor William J Mitchell, School of Architecture and Urban Planning, University of California, Los Angeles, for his helpful comments and suggestions.

References

Ackerman J S, 1962 "A theory of style" in *Aesthetic Inquiry* Eds M C Beardsley, H M Schueller (Dickenson, Belmont, Calif.)
Chomsky N, 1963 "Formal properties of grammars" in *Handbook of Mathematical Psychology, Volume II* Eds R D Luce, R R Bush (John Wiley, New York)
Durand J N L, 1823-1831 *Précis des Leçons d'Architecture* 3 volumes (École Royale Polytechnique, Paris)
Frankl P, 1914 *Principles of Architectural History* (MIT Press, Cambridge, Mass.)
Fu K S, 1974 *Syntactic Methods in Pattern Recognition* (Academic Press, New York)
Gips J, 1975 *Shape Grammars and Their Uses* (Birkhauser Verlag, Basel, Switzerland). Also PhD dissertation, Computer Science Department, Stanford University, Stanford, Calif.
Guadet J, 1902 *Éléments et Théories de l'Architecture* (Paris)
March L, Matela R, 1974 "The animals of architecture: some census results on N-omino populations for $N = 6, 7, 8$" *Environment and Planning B* **1** 193-216
Minsky M, 1967 *Computation: Finite and Infinite* (Prentice-Hall, Englewood Cliffs, NJ)
Mitchell W J, 1973 "Vitruvius computatus" in *Proceedings, LUBFS Conference on Models and Systems in Architecture and Building* (Medical and Technical Press, Lancaster)
Mitchell W J, Dillon R L, 1972 "A polyomino assembly procedure for architectural floor planning" in *Proceedings of the EDRA3/AR 8 Conference* School of Architecture and Urban Planning, University of California, Los Angeles
Rittle H, 1968 "Theories of cell configurations" in *Emerging Methods in Environmental Design and Planning* Ed. G Moore (MIT Press, Cambridge, Mass.)
Stiny G, 1975 *Pictorial and Formal Aspects of Shape and Shape Grammars* (Birkhauser Verlag, Basel, Switzerland). Also PhD dissertation, System Science Department, University of California, Los Angeles

Stiny G, in press "Generating and measuring aesthetic forms" in *Handbook of Perception, Volume X* Eds E C Carterette, M P Friedman (Academic Press, New York)
Stiny G, Gips J, 1972 "Shape grammars and the generative specification of painting and sculpture" in *Information Processing 71* Ed. C V Freiman (North-Holland, Amsterdam)
Summerson J, 1963 *The Classical Language of Architecture* (MIT Press, Cambridge, Mass)