Push and Pop on Pictures: Generalizing the Augmented Transition Network Formalism to Capture the Structure and Meaning of Images

Heinz Breu and Alan K. Mackworth

Department of Computer Science
University of British Columbia
Vancouver, B.C., Canada V6T 1W5

Abstract

This paper presents a generalization of the Augmented Transition Network formalism to allow the writing of picture grammars. The generalized ATN is used to write a picture grammar for a subclass of Heraldic Shields. This application serves to illustrate such (usually linguistic) terms as "anomaly", "ambiguity" and "paraphrase" as applied to pictures. The paper also suggests two obstacles to progress in picture grammars and concludes that the proposed system overcomes them.

1. Introduction

1.1 Overview

The linguistic approach to image understanding pushed the analogy between sentences and images to the point of designing picture grammars as generalizations of string grammars (Miller and Shaw 1968, Fu 1974, Ledley 1964, Evans 1969). Such attempts were successful in highly circumscribed two-dimensional scene domains, but the approach has not made much further progress. Two reasons for the lack of progress are, first, the failure to allow for the expression and exploitation of graphical relationships and, second, the weak expressive power of conventional grammars as a programming language in which to write effective recognition procedures.

We propose a substantial generalization of the Augmented Transition Network formalism for string grammars (Woods, 1970) to allow the writing of picture grammars. This modification required the solution of the following problems:

- the ATN must look at suitable picture primitives rather than words
- the ATN must work with many graphical relations not just concatenation and composition
- the notion of "get next token" must be generalized for pictures
- the notion of "end of sring" must be extended to pictures

The solutions to these problems are presented in detail. They required the generalization both of the ATN formalism for grammars and its associated parser.

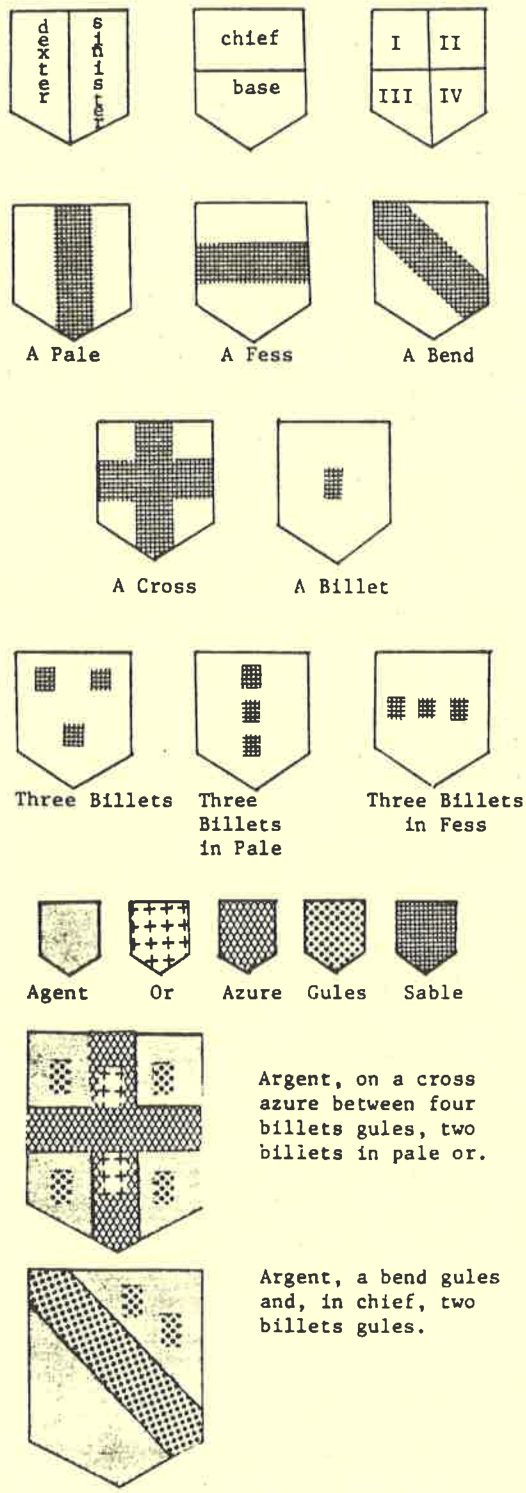
The modified ATN was used to write a picture grammar for a subclass of Heraldic Shields. This picture grammar takes, as input, an image of the shield represented as an array of pixels with colour values. It then outputs a linguistic description of the shield called an Heraldic Blazon.

1.2 Heraldic Shields

We will restrict ourselves to the class of shields that can be described by the class of blazons generated by the grammar of (Baker, 1977). The grammar for this class is presented below.

```
<shield> ::= <field>.
<field> ::= <colours>,<charges> |
            per pale: on the dexter <field>; and
                      on the sinister <field> |
            per fess: in chief <field>;
                      and in base <field> |
            quarterly: I <field>; II <field>;
                       III <field>; IV <field> |
            quarterly: I and IV <field>;
                       II <field>; III <field> |
            quarterly: I and IV <field>;
                       II <field> and III <field>
<colours> ::= <tincture> |
              <partition> <tincture> and
                          <tincture>
<tincture> ::= argent | or | azure |
               gules | sable
<partition> ::= per pale | per fess | quarterly
<charges> ::= <centred charge> |
              <minor charges> |
              on <centred charge>,
                 <minor charges> |
              <palewise ordinary> and,
                   <fesswise side>,
                   <minor charges> |
              <palewise ordinary> charged with
                   <minor charges> and,
                   <fesswise side>,
                   <minor charges> |
              <fesswise ordinary> and,
                   <palewise side>,
                   <minor charges> |
              <fesswise ordinary> charged with
                   <minor charges> and,
                   <palewise side>,
                   <minor charges> |
<centred charge> ::= <ordinary> |
                     <ordinary> between
                     <minor charges>
<ordinary> ::= <palewise ordinary> |
               <fesswise ordinary> | <cross>
<palewise ordinary> ::= a pale |
                          a pale <colours>
<fesswise ordinary> ::= a fess |
                        a fess <colours> |
                        a bend |
                        a bend <tincture>
<cross> ::= a cross | a cross <tincture>
<fesswise side> ::= on the dexter |
                    on the sinister
<palewise side> ::= in chief | in base
<minor charges> ::= a billet <tincture> |
                    <number> billets <direction>
                                     <tincture>
<number> ::= two | three | four
<direction> ::= in pale | in fess | <empty>
<empty> ::=
```

The terminology of heraldic blazons is bound to be unfamiliar to most readers. To aid in the reading of this paper, a description of these terms is provided in the following figure.



2. Generalizing the ATN

The ATN is generally regarded as an "off-the-shelf" tool for writing string grammars. The nature of the words in the dictionary and on the WRD arcs are left to the grammar writer. (The WRD arc is an extension to Wood's original proposal that requires the presence of a particular word next in the string.) Similarly, the nature of the function MORPH is also left to the user. (MORPH is the morpher that, given a derived word not in the dictionary, finds its root form in the dictionary and adds the required inflectional features.)

Our picture grammar ATN is to be treated this way also. We found however, that discussion of the ATN concepts is facilitated by the use of an example application. With this end in mind, let us first examine the heraldic shields domain.

2.1 Domain Specific Concepts

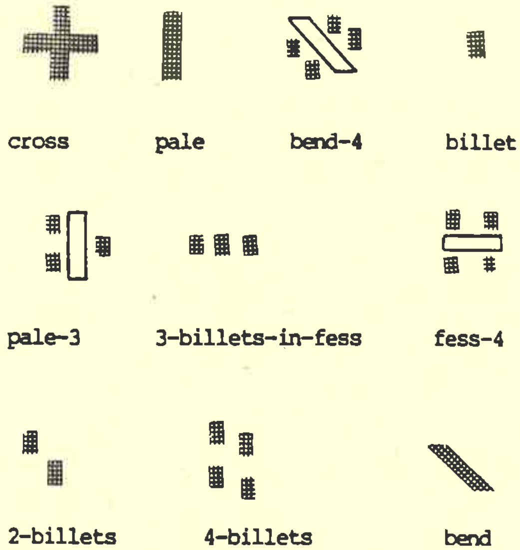
We would like to solve the following problems:

- define the picture primitives
- identify the picture primitives
- choose the relations needed to describe picture structure
- write a grammar and define the associated semantics

These problems will now be examined in more detail.

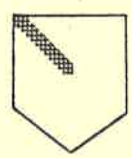
Primitives.

A subset of the primitives defined is shown below. The names underneath are the "words" on the WRD arcs of the grammar.

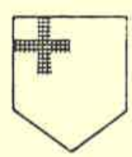


There are 24 primitives in all. The unshaded objects above are for illustration only and are not part of the primitives.

Each primitive is qualified by its coords which uniquely determines a primitive's size, shape, and coordinates. For example, (BEND COORDS) where COORDS = (Q-I SHIELDS), means only the primitive:



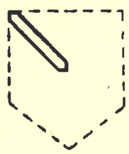
Note that "coords" is used in a slightly non-standard sense. This word here refers to the region in which the primitive is enclosed. For example, (CROSS (Q-I SHIELD)) is the shaded region below.



A billet in the middle of this region is indicated by (BILLET COORDS).

Identification

These primitives are detected by the WRD arcs of the grammar. The WRD arc requires the COORDS as an argument. A simple pattern matcher checks the occurrence of the primitive and returns its colour. For example, suppose the word is BEND and the coords are (Q-I SHIELD). The following pattern would then be generated:



Note that only the boundary of the primitive is generated as the pattern. This is to allow charged charges (e.g. a billet on a fess). The inside of the pattern is an "I don't care" region. The boundary must be a uniform colour (which is returned), and the pixels adjacent to it on the outside must be a different colour.

A possible matcher is shown below

```
(DEFUN MATCH (OBJECT COORDS)
  (COND ((APPLY1 '(LAMBDA (PATTERN)
                    (DIFFERENT (RIM PATTERN)
                               PATTERN))
                    (GENERATE OBJECT COORDS))
         (ERASE OBJECT COORDS))
        (T NIL)))
```

where

GENERATE returns a list of x,y coordinates making up the pattern i.e. $((X_1\ Y_1)\ (X_2\ Y_2)\ \ldots\ (X_n\ Y_n))$

RIM returns a list of x,y coordinates making up the rim of the pattern (i.e. the outside boundary)

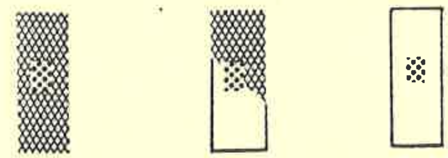
DIFFERENT returns T if PATTERN is a unique colour in the picture and (RIM PATTERN) consists of any colours different from those of PATTERN

ERASE SETQ2's every point of the object to a deletion figure (the object is all points within PATTERN of PATTERN's colour) and returns the colour of the object

MATCH therefore returns the colour of the object or NIL if the conditions are not met

PATTERN is a list of x,y coords

Upon recognition, the object is deleted. This is best done with a "bug" (or several, if need be) placed on the object, which "eats" away the coloured object. For example,



Note that the differently coloured billet in the centre is left untouched.

In our system, this deletion should be done using SETQ2. This is an undoable function in LISP/MTS and is chosen here to allow backtracking in the grammar. This notion of deletion corresponds to that of string advance in the regular ATN parser.

Relations

The relation among words in string grammars is that of concatenation. Another way of thinking of a sentence is as a series of slots into which the words fit. Along these lines, we consider a picture as a collection of regions into which the prmitives fit. These regions are the ones used in describing the shields with blazons. Namely,

DEXTER	Q-I
SINISTER	Q-II
CHIEF	Q-III
BASE	Q-IV
CROSS	FESS
PALE	BEND

Colours, as well, are considered to be relations. This holds not so much for primitives, as for entire regions. This is tested by looking at several points in the region where charges could not be. All the points must be the colour being tested for.

2.2 Generalization of ATN Concepts

The ATN parser was designed for string grammars, consequently some modifications were necessary to enable use with picture grammars.

String Advance

The area of interest is restricted on PUSH arcs by sending (via SENDR) the desired COORDS to the net of the subpicture being PUSHed for. For example, suppose we were looking for a cross in quadrant I of the shield and that we were presently at the top level. We would

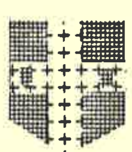
(SENDR COORDS (Q-I (GETR COORDS)))

to the CENTRED-CHARGE/ net. Here COORDS is the register containing the current region and Q-I is a function which returns the region which is the first quadrant of the region defined by its argument.

The notion of "advancing the input picture" is realized by the following actions:

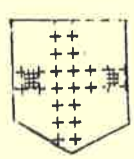
(1) Upon recognition of a primitive, that primitive is deleted from the picture (replaced by deletion characters). This occurs on WRD arcs.

(2) Upon detection of the colour of a region, that colour is deleted for that region. This occurs on TST arcs. For example,



detection of

(SABLE SHIELD)



Note that the two billets in fess sable are not deleted due to the nature of the bug method of deletion.

End of String

The notion of, "a string grammar parse is successful when the entire string is accepted", is generalized to that of, "a picture grammar parse is successful when the entire picture is accepted". This condition is detected in our system by a function (EMPTY COORDS) which tests if all the objects within the region specified by COORDS have been deleted. "End of Picture" is determined by calling (EMPTY SHIELD/).

WRD arcs

These are now used to detect picture primitives, rather than words. This arc now takes COORDS as an argument. This is necessary to completely specify the primitive as explained earlier in the section on primitives. An internal variable, PROPERTY, is set to (MATCH OBJECT COORDS). This could be anything the user desires, in this case, the colour of the object. Following the conventions of (Reiter, 1978), the WRD arc definition appears below:

(WRD <word> [coords] [test] [action]*)

If "test" evaluates to non-NIL, and if the primitive "word" is found at "coords" by MATCH (user defined function), perform the sequence of actions. The last action must be (TO <state>). * is bound to the tuple (word PROPERTY), where PROPERTY is returned by (MATCH word coords). WRD "advances the input picture".

MEM arcs

As for WRD, only MEM allows a list of primitives (in only one region, COORDS) all of which are checked. The format for MEM arcs is:

(MEM <words> [coords] [test] [action]*)

where <words> is a list of possible

alternatives.

TST arcs

These are used to test relations. In the SHIELDS example, they are used to test for the colours of regions. As such, TST still "advances the input picture".

PARSE

Since no sentence is being parsed, the function PARSE need not accept a sentence as an argument. The new format for calling PARSE is:

(PARSE <state>)

All string advance calls have been eliminated from the parser, as have the CAT arc and the calls to MORPH.

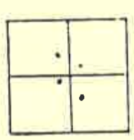
3. The Heraldic Shields Grammar

3.1 Semantic Tests

A few words need to be said about the semantic tests used. At present, these tests are only simulated but their designs are presented in the following discussion.

Q-TEST
PALE-TEST
FESS-TEST

These tests return T if the region specified by the argument is divided quarterly, vertically (per pale), or horizontally (per fess) respectively. These tests would work as follows:



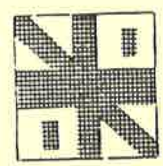
Tests are made at the points (.) indicated above. These points are chosen so as to avoid charges. The numerals in the tests below are to be interpreted as

<NUMERAL> = colour of <NUMERAL> quadrant

Q-TEST (I $\neq$ II) AND (II $\neq$ IV) AND
(I $\neq$ III) AND (III $\neq$ IV)
PALE-TEST (I = III) OR (II = IV)
FESS-TEST (I = II) OR (III = IV)

Q-COLOURED
P-COLOURED
F-COLOURED

These are tests to determine if a given region is coloured by quarterly, palewise or fesswise division respectively. The tests themselves are the same as those above, but the tests are made on all the unshaded points below:



Points of one quadrant _must_ be of uniform colour.

3.2 EMPTY

This test determines if the region specified by its argument has had all objects deleted, which is the case if the entire region is made up of deletion characters. It is used on POP arcs to ensure that a given region has been correctly parsed before going on to the next region. For example, we would like to know that quadrant I has been correctly parsed before attempting to process quadrant II. It is also used on some PUSH arcs to prevent looking for something in an empty region.

Note also the use of SENDR in restriction of the area of interest by narrowing down of the COORDS.

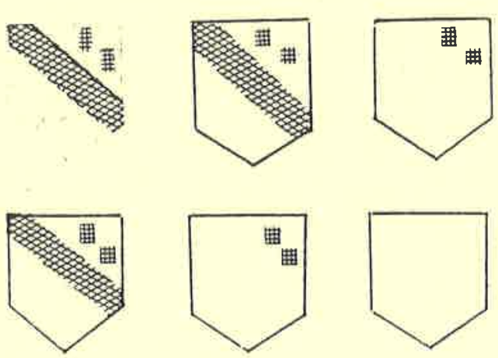
3.3 Successful Parses

A simulation of the proposed system has been implemented. This includes:

a) a modified ATN parser
b) the grammar
c) simulated semantic tests

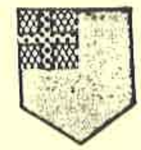
That is, the entire system has been implemented on the AMDAHL/470 at UBC, except the semantic tests, which are not automatic, but require interaction with a human operator. This implementation correctly translates shields into blazons for those shields described by Baker's heraldic blazon grammar. For example, when given the two example shields of Figure 1, the program yielded exactly those blazons corresponding to the shields in Figure 1. Shields not describable by that grammar are said to be anomalies. Anomalies are correctly rejected by our system.

The figures below show how the picture is processed.



3.4 Ambiguities

An interesting point that arose in this work is the existence of ambiguous shields. A shield is said to be ambiguous if it can be described by more than one blazon. An example of such a shield along with two possible blazons is shown below.



(1) Per pale: on the dexter per fess: in chief azure, a cross gules; and in base argent; and on the sinister argent

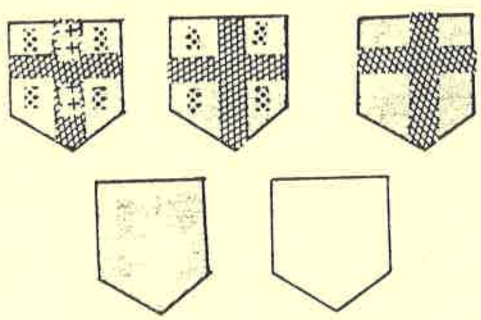
(2) Per fess: in chief per pale: on the dexter azure, a cross gules; and on the sinister argent; and in base argent.

Note that the above shield could also be described as divided quarterly, but this would not be a valid parse in our system due to the semantic Q-TEST. The ambiguity is resolved in our system by arc ordering. The ATN is thought of as being a parallel process, but this is, of course, not the case in any sequential machine implementation. In our case, the arcs are tried in the order in which they appear. Hence, if we put the arc for PER PALE division first, then it will be tried first. This is equivalent to saying that if a shield can be divided both PER PALE and PER FESS, then divide it PER PALE. Blazon (1) is therefore the correct description as generated by our parser.

4. Directions for Further Work

Due to the nature of our pattern matcher, multi-coloured primitives are not permitted (e.g. a two tone bend). The matcher could be extended to allow such objects, which are permitted in Baker's grammar.

The present system's recognition of region colours is unsatisfactory. It must avoid all charges that might be on the region in its detection and must be careful to delete around them. A re-ordering of nodes and arcs could have the system look first for charges, then for the colours of the region. Deletion would now mean extending the colours over the deleted charge. The semantic test EMPTY would require modification to consider a region of pure colour as "empty". As well, the grammar could no longer rely on knowledge of the region before checking for charges (e.g. at present, if a cross is detected, charges on the cross are looked for. The new design would find the charges on the cross first (and then realize that they are on a cross later, then look for the cross). An example of how the new system would process shields is diagrammed below.



Since our pattern matcher effectively generates each primitive, we could, by modifying this matcher into a generator, generate shields by running the grammar "in reverse". That is, we (or some program) could decide on precisely what arcs to take (i.e. the parse is pre-selected).

This, together with our notion of shield ambiguity, gives us a method of paraphrasing blazons. If we consider our arc ordering as being "standard", then our parse can be considered to be a "standard blazon". In particular then, we have a method for the standardization of non-standard blazons:

- use the non-standard blazon to generate a shield
- parse the shield
- the output will be the standardized blazon

5. Conclusions

This paper has shown that ATN's are modifiable to allow the writing of picture grammars. We have demonstrated this for a picture grammar for heraldic shields exhibiting context-sensitive properties, such as identity of quadrants.

We have found that the syntactic method (in our grammar) yields useful information (blazons) about the structure of heraldic shields. Anomalous shields are rejected and a notion of "standard blazon" eliminates ambiguity.

We feel that generation of shields is possible by running the grammar "in reverse", and that heraldic blazons may be paraphrased by considering shields to be the semantics of the blazon.

We see that relational richness is provided in our ATN by our use of "layered" regions. That is, we look for things on crosses, etc. An alternative way of thinking of this, perhaps more indicative of its power, is as "restricted" regions. That is, when we PUSH for (Q-I coords), we are actually focusing our attention on this restricted area of coords.

Finally, we conclude that our system has inherited the ease-of-programming associated with string grammar ATNs.

REFERENCES

Baker, J. Grammar for Heraldic Blazons. In UBC CPSC 215 Notes, 1977.

Evans, T.G. Descriptive pattern analysis techniques. In (Grasselli 1969), 79-96

Fu, K.S. Syntactic Methods in Pattern Recognition. Academic Press, New York, 1974.

Graselli, A.(Ed.) Automatic Interpretation and Classification of Images. NATO Advanced Study Institute, Pisa, 1968. Ledley, R.S. High-speed automatic analysis of biomedical pictures. Science 146, 9 (1964), 216-223

Miller, W.F. and Shaw, A.C. Linguistic methods in picture processing—a survey. Proc. AFIPS Fall (1968) Joint Computer Conference, 279-290.

Reiter, R. The Woods Augmented Transition Network parser. Technical Note 78-3, Department of CPSC, UBC, 1978.

Woods, W.A. Transition network grammars for natural language analysis. Comm. ACM 13, 10 (1970), 591-606