A. MATEESCU

A. SALOMAA

Finite degrees of ambiguity in pattern languages

Informatique théorique et applications, tome 28, nº 3-4 (1994), p. 233-253

<http://www.numdam.org/item?id=ITA_1994__28_3-4_233_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol 28, n° 3-4, 1994, p 233 à 253)

FINITE DEGREES OF AMBIGUITY IN PATTERN LANGUAGES (*)

by A. Mateescu ([1]) and A. Salomaa ([1])

Abstract – *The paper investigates nondeterminism and degrees of nondeterminism in representing words according to a pattern given a priori The issues involved belong to the basic combinatorics of words Our main results concern decidability and construction of finite degrees of nondeterminism*

1. INTRODUCTION

There has been much interest recently in *patterns* and *pattern languages*. (*See*, for instance, [3], [5], [6], [8] and their references). Indeed, a natural way of describing a given sample of words is to exhibit a common pattern for the words. Such an approach is especially appropriate if the sample is growing, for instance, through some *learning* process. Finding patterns for a sample sets is, thus, a typical problem of *inductive inference.*

Pattern languages in the sense understood in this paper were introduced in [1]. The essential difference between the two cases, where the empty word λ can or cannot be substituted for the variables, was studied in [5]. It was also observed that many problems in combinatorics of words, ranging from the classical ones discussed in [11] to the more recent ones discussed in [2] and [7], can be expressed in terms of the inclusion problem for pattern languages. The same holds true for certain problems in term rewriting, [8]. From this point of view it is not surprising that the inclusion problem turned out to be undecidable, [6].

Given a terminal word w and a pattern α, it may happen that w "follows" the pattern α in several ways. In other words, there are several assignments for

(*) Research supported by the Academy of Finland, Project 11281 All correspondence to Arto Salomaa
([1]) Academy of Finland and Department of Mathematics University of Turku, SF-20500 Turku, Finland

the variables in α, each of which gives rise to w. This kind of *nondeterminism* or *ambiguity* in patterns will be investigated in this paper. Indeed, the classical language-theoretic notions of *unambiguity, inherent ambiguity* and *degrees of ambiguity* (see [10]) find their natural counterparts in the context of patterns. The proofs make use of various aspects in combinatorics of words. The case of a finite degree of ambiguity greater than one turns out to be rather involved. In decidability issues, modifications of the result by Makanin, [9], can be used.

2. DEFINITIONS AND PRELIMINARY RESULTS

Let Σ be an alphabet (of *terminals*) and V an alphabet (of *variables*) such that $\Sigma \cap V = \varnothing$. Let $H(\Sigma, V)$ (resp. $H_+(\Sigma, V)$) be the set of all morphisms (resp. nonerasing morphisms)

$$h: \quad (\Sigma \cup V)^* \to \Sigma^*$$

such that $h(a) = a$ for all $a \in \Sigma$.

Nonempty words α over $\Sigma \cup V$ are referred to as *patterns*. A pattern $\alpha \in (\Sigma \cup V)^+$ defines the *languages*:

$$L_E(\alpha) = \{w \mid h(\alpha) = w, \quad \text{for some } h \in H(\Sigma, V)\}$$
$$L_{NE}(\alpha) = \{w \mid h(\alpha) = w, \quad \text{for some } h \in H_+(\Sigma, V)\}.$$

The languages $L_E(\alpha)$ and $L_{NE}(\alpha)$ are referred to as *pattern languages*. Sometimes we speak of *E-patterns* and *NE-patterns* ("erasing" and "nonerasing") to indicate which of the languages we are interested in. Also the alphabet Σ may be indicated in the notation: $L_E(\alpha, \Sigma)$ or $L_{NE}(\alpha, \Sigma)$. This is the case especially if Σ is not visible from α, that is, all letters of Σ do not occur in α.

We now come to the central notions of this paper. It may happen that a word w in $L_E(\alpha)$ or $L_{NE}(\alpha)$ has several "representations", that is, there are several morphisms h satisfying $w = h(\alpha)$. For instance, the terminal word $w = a^7 ba^7$ possesses 8 representations in terms of the pattern $\alpha = xyx$. (The number is 7 if α is viewed as an NE-pattern.) We express this by saying that the *degree of ambiguity* of w with respect to α equals 8. Whenever important, we indicate whether we are dealing with the E- or NE-case.

The *degree of ambiguity of a pattern* α equals the maximal degree of ambiguity of words w in the language of α, or infinity (∞) if no such maximal degree exists. More formally, we associate to a pattern α over $\Sigma \cup V$ and a word $w \in \Sigma^+$ the subset $S(\alpha, w, \Sigma)$ of $H(\Sigma, V)$, consisting of morphisms h such that $h(\alpha) = w$. The *cardinality* of this subset is denoted by $card(\alpha, w, \Sigma)$. (We make here the *convention* that morphisms differing only on variables not present in α are not counted as different). The *degree of ambiguity* of α equals $k \geqq 1$ iff

$$card(\alpha, w, \Sigma) \leqq k, \qquad \text{for all } w \in L_E(\alpha),$$

and

$$card(\alpha, w', \Sigma) = k, \qquad \text{for some } w' \in L_E(\alpha).$$

If there is no such k, then the *degree of ambiguity* of α equals ∞. For $k = 1$, α is also termed *unambiguous* and, for $k > 1$, α is termed *ambiguous*.

Remark: The terminals actually appearing in α constitute a subset Σ', maybe empty, of Σ. Indeed, any pattern over $\Sigma' \cup V$ is a pattern also over $\Sigma \cup V$, where $\Sigma' \subseteq \Sigma$. In the definition of the degree of ambiguity we actually specified the pair (α, Σ). However, it is pleasing to observe that, in fact, it suffices to specify only α because the degree is independent of the choice of Σ. The following argument justifies this observation.

If α contains no terminals (that is, Σ' is empty), then the degree of ambiguity of α is 1 or ∞, depending on whether α contains occurrences of one or more than one variable. If Σ' contains at least one terminal a, we denote by $g: \Sigma^* \to \Sigma'^*$ the morphism keeping the letters of Σ' fixed and mapping the letters of $\Sigma - \Sigma'$ into a. Clearly, the degree of ambiguity does not decrease if the terminal alphabet Σ' is replaced by Σ. But it does not increase either because, whenever w has m representations according to α, then $g(w)$ has at least m representations according to α. $\square$

By the above remark, we speak of the degree of ambiguity of a pattern α (without specifying the alphabet). The above definitions were carried out in the E-case. The NE-case is analogous.

The notions are now naturally extended to concern languages. We do this in the E-case. A *pattern language* L is *ambiguous of degree* $k \geqq 1$ if $L = L_E(\alpha)$, for some pattern α ambiguous of degree k, but there is no pattern β of degree less then k such that $L = L_E(\beta)$. Here k is a natural number or ∞. Again, if $k = 1$ we say that L is is *unambigous*. Otherwise, L *(inhenrently) ambiguous.*

It was shown in [1] that two NE-patterns are *equivalent* (in the sence that they generate the same language) exactly in the case they are identical up to a possible renaming of variables. This yields immediately the following result:

THEOREM 1: *The degree of ambiguity of an NE-pattern α equals the degree of ambiguity of the pattern language $L_{NE}(\alpha)$.*

Theorem 1 does not hold for E-patterns: $L_E(X) = L_E(XY)$ but the degrees of ambiguity of the patterns X and XY are 1 and ∞, respectively.

THEOREM 2: *For every E-pattern (containing at least one variable), there is an equivalent E-pattern whose degree of ambiguity is ∞. Conversely, there are E-patterns (for instance, $XYYX$) such that the degree of ambiguity of every equivalent E-pattern, and hence also the degree of ambiguity of the generated language, equals ∞.*

Proof: To prove the first sentence, it suffices to replace all occurrences of a variable X in the pattern with $X_1 X_2$. The second sentence follows by a simple case analysis concerning patterns equivalent to $XYYX$. □

In what follows we do not make any distinction between E- and NE-patterns because the results hold in both cases.

THEOREM 3: *Every pattern containing occurrences of only one variable X is unambiguous. Every pattern containing occurrences of at least two variables X and Y but of at most one terminal a is ambiguous of degree ∞.*

Proof: The first sentence follows by a length argument: for every w, the value $h(X)$ is uniquely determined. To prove the second sentence, we first replace in the given pattern the other variables (if any) with a. The resulting pattern contains $m \geqq 1$ X's, $n \geqq 1$ Y's and $p \geqq 0$ a's. Given any k, we can find a z_k such that

$$mx + ny + p = z_k$$

has more that k positive solutions (x, y). This means that a^{z_k} has more than k representations according to the given pattern. □

3. DETERMINISM AND NONDETERMINISM

We now continue the study begun in the preceding section and characterize some basic cases of ambiguity and nonambiguity.

THEOREM 4: *Every pattern α satisfying the following two conditions is of ∞ degree of ambiguity.* (i) *α contains occurrences of at least two variables.* (ii) *Some variable occurs in α only once.*

Proof: Let Z be the variable that has only one occurrence in α. We'll consider the following two possibilities:

Case 1: The pattern α starts or ends with Z. Assume that α ends with Z (the situation α starts with Z is symmetric). Hence, $\alpha = \beta Z$, where β is a pattern that contains at least one variable, but Z does not occur in β. Let X be the leftmost variable in β, *i. e.* $\beta = X\gamma$, where γ is a pattern. (We assume without loss of generality that β starts with a variable). Therefore, $\alpha = X\gamma Z$ and Z does not occur in $X\gamma$.

Now, assume to the contrary that α has the degree of ambiguity k, where $k < \infty$. Let w be a terminal word that has k different decompositions with respect to α. Let u be a fixed terminal word. Consider the morphism f, defined as: $f(X) = w$ and $f(Y) = u$, for any variable Y, $Y \neq X$. We obtain $f(\alpha) = t = wv$.

The terminal word t has one decomposition with respect to α corresponding to the morphism f and, moreover, t has k other decompositions with respect to α, corresponding to the k possible decompositions of w, each such decomposition being modified as follows: if Z was substituted by r in the originally considered decomposition of w, then Z is substituted by rv, in order to obtain t, and any other variable Y continues to be substituted as in the originally considered decomposition of w. Thus, t has $k+1$ different decompositions with respect to α, contrary to the assumption that α has the degree of ambiguity k.

Case 2: The variable Z has only one occurrence in α and this occurrence is neither the leftmost nor the rightmost occurrence of a variable in α. Hence, $\alpha = X\beta Z\gamma Y$, where β, γ are patterns, X, Y, Z variables (possibly $X = Y$) and Z does not occur in $X\beta\gamma Y$. Again, assume to the contrary that α has the degree of ambiguity k, where $k < \infty$, and let w be a terminal word that can be decomposed in k different ways with respect to α. Hence, there are p_i, q_i, r_i, $i = 1, \ldots, k$, such that $w = p_i r_i q_i$ and Z was substituted by r_i. Consider the morphism f, such that: $f(X) = w$, $f(Y) = w$, and $f(Q) = u$, for any Q, $Q \neq X$, $Q \neq Y$, and u is a fixed, arbitrary terminal word. (Note that, if $X = Y$, then f continues to be well-defined.) Let t be the terminal word $f(\alpha)$. Hence, $t = wsw$, where $s = f(\beta Z\gamma)$. Note that t has one decomposition, with respect to α, corresponding to the morphism f. Moreover, $t = p_i r_i q_i s p_i r_i q_i$, for $i = 1, \ldots, k$. Each such decomposition of t is corresponding to the substitution of Z in α with the terminal word $r_i q_i s p_i r_i$, and the remaining variables are substituted as in the original decomposition of w. Therefore, altogether, t has $k+1$ different decompositions with respect to α, contrary to the assumption that the degree of ambiguity of α is k. $\square$

Theorems 3 and 4 determine the degree of ambiguity of all patterns except patterns α satisfying each of the following three conditions: (i) α contains occurrences of at least two variables. (ii) α contains occurrences of at least two terminals. (iii) Every variable occurs in α at least twice. Indeed, all tricky cases fall among such patterns α. Let us consider patterns with two occurrences of two variables, separated by terminal words. Such patterns belong to one of the three types

$$X\, w_1\, Y\, w_2\, X\, w_3\, Y, \qquad X\, w_1\, Y\, w_2\, Y\, w_3\, X, \qquad X\, w_1\, X\, w_2\, Y\, w_3\, Y.$$

We mention without proof that the first two types are always of degree of ambiguity 1 or ∞, whereas a finite degree $\neq 1$ is possible in the third type. We will return to this matter in Section 5.

The following theorem serves as a basis in many constructions.

THEOREM 5: *The pattern $X\, a\, Y X\, b\, Y$ is unambiguous.*

Proof: Assume that for some terminal words u_1, v_1, u_2, v_2 there is the equality:

$$(1) \qquad u_1\, av_1\, u_1\, bv_1 = u_2\, av_2\, u_2\, bv_2$$

Note that $|u_i\, av_i| = |u_i\, bv_i|$, $i = 1, 2$. Thus, the equality (1) leads to the next two equalities:

$$(2) \qquad u_1\, av_1 = u_2\, av_2 \qquad \text{and} \qquad u_1\, bv_1 = u_2\, bv_2$$

Without loss of generality, we can assume that $|u_1| \leqq |u_2|$ and, consequently, $|v_1| \geqq |v_2|$. Thus, there are terminal words u_3 and v_3 such that $u_2 = u_1\, u_3$ and $v_1 = v_3\, v_2$. Hence, from (2) we deduce that:

$$av_3 = u_3\, a \qquad \text{and} \qquad bv_3 = u_3\, b.$$

The above system of equations has the unique solution $u_3 = v_3 = \lambda$. Therefore, we obtain $u_1 = u_2$ and $v_1 = v_2$ and thus, the pattern is unambiguous. $\square$

Composition can be applied to patterns in the natural fashion: variables are uniformly substituted by patterns. If in the pattern of Theorem 5 the variable

X is replaced by $X_1\, a\, Y_1\, X_1\, b\, Y_1$ (that is, the original pattern with renamed variables) and the variable Y is left unchanged, we obtain the pattern

$$X_1\, a\, Y_1\, X_1\, b\, Y_1\, a\, Y X_1\, a\, Y_1\, X_1\, b\, Y_1\, b\, Y.$$

Clearly, also this pattern is unambiguous, by Theorem 5. In fact, the next theorem is a corollary of Theorem 5.

THEOREM 6: *Compositions of unambiguous patterns are unambiguous. Unambiguous patterns of arbitrarily many variables can be effectively constructed.*

4. DECIDABILITY

Using the general theorem of Makanin, [9], the following results can be obtained quite independently of our other results.

THEOREM 7: *The following problems are decidable, given a pattern α and a natural number k. Is the degree of ambiguity of α equal to k, greater than k or less than k? Consequently, it is decidable whether or not α is unambiguous.*

Proof: It was shown in [4] how Makanin's decidability can be extended to concern systems of equations and *inequalities*. Inequalities $x \neq x'$ are essential in expressing that a given equation possesses two solutions. The details of the argument are left to the reader. □

Theorem 7 does not yield a method of deciding whether or not the degree of ambiguity of α is ∞. Indeed, this is an open decision problem. As regards decision methods for pattern languages, the results of Theorems 1 and 7 can be combined for NE-patterns. The situation is trickier for E-patterns. In fact, even the decidability of the equivalence problem is open for E-patterns.

We mention, finally, that Theorem 6 gives a simple way of going from a system of equations to a single equation. Consider a system of equations

$$(*) \qquad \alpha_i = \beta_i, \qquad i = 1, \ldots, n,$$

where α_i and β_i may contain variables and constants (that is, terminals). Choose an unambiguous pattern $P(X_1, \ldots, X_n)$ of n variables X_i. Then (*) has a solution exactly in the case the equation

$$P(\alpha_1, \ldots, \alpha_n) = P(\beta_1, \ldots, \beta_n)$$

has a solution.

5. FINITE DEGREE OF AMBIGUITY

It is rather difficult to exhibit patterns with the degree of ambiguity $k > 1$, where k is finite. Indeed, it was our conjecture for a long time that 1 and ∞ are the only possible degrees.

Notations: If $t \in \Sigma^*$ then *first(t)* (*last(t)*) denotes the leftmost (rightmost) letter of t. Moreover, *pref(t)* (*suf(t)*) is the set of all proper prefixes (suffixes) of t.

DEFINITION 8: Let $\alpha = \alpha(X_1, \ldots, X_n)$ be a pattern. A *nontrivial solution* of the equation:

$$\alpha(X_1, \ldots, X_n) = \alpha(X'_1, \ldots, X'_n)$$

is a $2n$-tuple of terminal words, $(x_1, \ldots, x_n, x'_1, \ldots, x'_n)$, such that:

$$\alpha(x_1, \ldots, x_n) = \alpha(x'_1, \ldots, x'_n)$$

and, moreover, $(x_1, \cdots, x_n) \neq (x'_1, \ldots, x'_n)$. □

LEMMA 9: *The pattern,*

$$\alpha = X\, uv\, X\, vwu\, Y\, uvw\, Y$$

has the degree of ambiguity 2, if u, v, w are nonempty words over the alphabet $\Sigma = \{a, b, c\}$ such that:

(i) $|u| = |v| = |w|$,

(ii) $first(t) = last(t)$, *for any* $t \in \{u, v, w\}$,

(iii) $pref(uvw) \cap suf(uvw) = \emptyset$ *and* $pref(v) \cap suf(w) = \emptyset$,

(iv) *if* $t_1 \in pref(uv)$ *and* $t_2 \in suf(uv)$, *then* $t_1 t \neq t t_2$ *for any* $t \in \Sigma^*$.

Proof: First, observe that the pattern α is ambiguous. For example:

$$\begin{aligned}\alpha(wu, uvwvw) &= \overbrace{wu}\, uv\, \overbrace{wu}\, vwu\, \overbrace{uvwvw}\, uvw\, \overbrace{uvwvw} \\ &= \underbrace{wuuvw}\, uv\, \underbrace{wuuvw}\, vwu\, \underbrace{vw}\, uvw\, \underbrace{vw} \\ &= \alpha(wuuvw, vw).\end{aligned}$$

It remains to prove that there is no terminal word t that has 3 or more decompositions with respect to α.

CLAIM A_1: *The general form of nontrivial solutions of the equation:*

$$\alpha(X_1, Y_1) = \alpha(X_2, Y_2)$$

is:

$$x_1 = w\,(\nu\, uvw)^i\,\zeta, \qquad y_1 = \mu\, uvw\,(\nu\, vwu)^j\,\eta,$$
$$x_2 = w\,(\nu\, uvw)^i\,\zeta\, uvw\,\nu, \qquad y_2 = (\nu\, vwu)^j\,\eta,$$

where ζ *is a proper prefix of* $\nu\, uvw$ *or* λ, η *is a proper prefix of* $\nu\, vwu$ *or* λ, i, $j \geqq 0$ *and, moreover,*

$$\zeta\, vwu\,\mu = \nu\, uvw\,\zeta \quad \text{and} \quad \eta\, uvw\,\mu = \nu\, vwu\,\eta.$$

Proof of Claim A_1*:* Assume that x_1, y_1, x_2, y_2 are nonempty words over Σ such that $|x_1| < |x_2|$ and

$$x_1\, uvx_1\, vwuy_1\, uvwy_1 = x_2\, uvx_2\, vwuy_2\, uvwy_2. \tag{1}$$

Hence, there are x'_2, $y'_1 \in \Sigma^+$ such that $x_2 = x_1\, x'_2$, $y_1 = y'_1\, y_2$ and $|x'_2| = |y'_1| > 0$. If follows from (1) that

$$uvx_1\, vwuy'_1\, y_2\, uvwy'_1 = x'_2\, uvx_1\, x'_2\, vwuy_2\, uvw. \tag{2}$$

Reading in the above equality a prefix of length $|x'_2\, uv|$ and a suffix of length $|uvwy'_1|$, we obtain

$$uvx_3 = x'_2\, uv \quad \text{and} \quad uvwy'_1 = y_3\, uvw \tag{3}$$

for some x_3, $y_3 \in \Sigma^+$.

Note that x'_2 cannot be a proper prefix of uv. (Otherwise x_3 is a proper suffix of uv and for $t = uv$ this contradicts condition (iv).)

Assume that $x'_2 = uv$. From (i) and (3) (second equality) we deduce that $y'_1 = vw$ and hence $uvw = y_3\, u$. Using (i) it follows that $w = u$, a contradiction.

Now, assume that $x'_2 = uvp_1$, $y'_1 = q_2\, vw$ for some words p_1, q_2, with $0 < |p_1| = |q_2| < |u|$. From the second equality of (3) we obtain

$$uvwq_2 = y_3\, u. \tag{4}$$

Hence, q_2 is a proper suffix of u, *i. e.*, $u = q_1\, q_2$ for some $q_1 \in \Sigma^+$. The equality (4) becomes $uvw = y_3\, q_1$. Thus q_1 is a proper suffix of w and a proper prefix of u. But this contradicts the first part of condition (iii).

Concluding, we deduce that $|x'_2| = |y'_1| \geqq |uvw|$. Hence, $x'_2 = uvg$, $y'_1 = huvw$, for some $g,\ h \in \Sigma^*$ such that $|g| = |h| + |w|$. Therefore, the relation (2) becomes:

$$x_1\, vwuhuvwy_2\, uvwh = guvx_1 uvgvwuy_2.$$

Considering prefixes and suffixes of the same length in the above equality, we obtain:

$$(5) \qquad x_1\, vwuh = guvx_1 \qquad \text{and} \qquad wy_2\, uvwh = gvwuy_2.$$

From the second equality in (5), if follows that g cannot be a proper prefix of w. (Otherwise, after simplification, a proper prefix of v is equal to a proper suffix of w, contrary to the second part of condition (iii).)

Hence, $g = wg'$, for some $g' \in \Sigma^*$, with $|g'| = |h|$. Thus, the relations (5) become:

$$(6) \qquad x_1\, vwuh = wg'\, uvx_1 \qquad \text{and} \qquad y_2\, uvwh = g'\, vwuy_2.$$

By a similar argument, x_1 cannot be a proper prefix of w. Thus, $x_1 = wx''$, for some $x'' \in \Sigma^*$. Denote $y_2 = y''$. From (6), it follows that

$$(7) \qquad x''\, vwuh = g'\, uvwx'' \qquad \text{and} \qquad y''\, uvwh = g'\, vwuy''.$$

Denoting $g' = \nu$ and $h = \mu$, the relations (7) become:

$$x''\,(vwu\,\mu) = (\nu\ uvw)\,x'' \qquad \text{and} \qquad y''\,(uvw\,\mu) = (\nu\ vwu)\,y''.$$

where $|\nu| = |\mu|$.

Using a well-known result often called Lyndon's Theorem, it follows that

$$(8) \qquad x'' = (\nu\, uvw)^i\,\zeta \qquad \text{and} \qquad y'' = (\nu\, vwu)^j\,\eta$$

where $i,\ j \geqq 0$ and:

$$(9) \qquad \nu\, uvw\,\zeta = \zeta\, vwu\,\mu, \qquad \eta\, uvw\,\mu = \nu\, vwu\,\eta.$$

Using (8) and (9) we obtain the general form of $x_1,\ x_2,\ y_1,\ y_2$ as in the Claim A_1.

End of the proof of Claim A_1.

Assume now the existence of an intermediate solution, (x', y'), of the equation:

$$\alpha(X_1, Y_1) = \alpha(X_2, Y_2).$$

i. e., $|x_1| < |x'| < |x_2|$, $|y_2| < |y'| < |y_1|$ and

$$\alpha(x_1, y_1) = \alpha(x', y') = \alpha(x_2, y_2).$$

CLAIM A_2: *If x_1, x', x_2, y_1, y', y_2 are terminal words with the above properties, then:*

$$\begin{aligned} x_1 &= w\,\zeta, & y_1 &= \rho\, uvw\, \delta\, uvw\, \eta, \\ x' &= w\,\zeta\, uvw\,\theta, & y' &= \delta\, uvw\, \eta, \\ x_2 &= w\,\zeta\, uvw\,\theta\, uvw\,\pi, & y_2 &= \eta, \end{aligned}$$

where θ, δ, η, ρ, ζ, π are terminal words that satisfy the system of equations $E = \{e_1, e_2, e_3, e_4, e_5\}$ with:

(e_1) $\theta\, vwu\, \delta = \delta\, uvw\, \theta$

(e_2) $\eta\, uvw\, \rho = \theta\, uvw\, \eta$

(e_3) $\zeta\, vwu\, \rho = \theta\, uvw\, \zeta$

(e_4) $\zeta\, uvw\, \delta = \pi\, uvw\, \zeta$

(e_5) $\eta\, uvw\, \delta = \pi\, vwu\, \eta$

and, moreover, $|\pi| = |\delta|$ and $|\theta| = |\rho|$.

Proof of Claim A_2: From Claim A_1 we obtain that

$$x' = w\,(\nu\, uvw)^i\, \zeta\varphi \qquad \text{and} \qquad y' = \psi\,(\nu\, vwu)^j\, \eta, \tag{10}$$

for some nonempty terminal words φ and ψ, such that φ is a proper prefix of $uvw\,\nu$, ψ is a proper suffix of $\mu\, uvw$ and $|\varphi| + |\psi| = |\nu| + |uvw| = |\mu| + |uvw|$.

Moreover, we obtain that:

$$\begin{aligned} \alpha(x', y') = w\,(\nu\, uvw)^i\, \zeta\varphi\, uvw\,(\nu\, uvw)^i \\ \zeta\varphi\, vwu\,\psi\,(\nu\, vwu)^j\, \eta\, uvw\,\psi\,(\nu\, vwu)^j\, \eta. \end{aligned} \tag{11}$$

From the equality $\alpha(x_1, y_1) = \alpha(x', y')$ and (11) it follows that:

$$\begin{aligned} uvw\,(\nu\, uvw)^i\, \zeta\, vwu\, \mu\, uvw\,(\nu\, vwu)^j\, \eta\, uvw\, \mu\, uvw \\ = \varphi\, uvw\,(\nu\, uvw)^i\, \zeta\varphi\, vwu\,\psi\,(\nu\, vwu)^j\, \eta\, uvw\,\psi. \end{aligned} \tag{12}$$

Note that φ cannot be a proper prefix of uvw. (Otherwise, after simplification of φ, it follows that a proper suffix of uvw must be a proper prefix of uvw, contrary to the condition (iii).)

Hence, $\varphi = uvw\,\theta$ and $\psi = \delta\,uvw$, for some words θ, $\delta \in \Sigma^*$ such that θ is a prefix of ν with $|\theta| < |\nu|$, and δ is a suffix of μ with $|\delta| < |\mu|$.

The equality (12) becomes:

$$\begin{aligned}(13)\qquad & (\nu\,uvw)^i\,\zeta\,vwu\,\mu\,uvw\,(\nu\,vwu)^j\,\eta\,uvw\,\mu \\ & = \theta\,uvw\,(\nu\,uvw)^i\,\zeta\,uvw\,\theta\,vwu\,\delta\,uvw\,(\nu\,vwu)^j\,\eta\,uvw\,\delta\end{aligned}$$

where:

$$(14)\qquad |\theta| + |\delta| + 3\,|u| = |\mu| = |\nu| \qquad \text{and} \qquad |\theta| < |\mu|$$

Note that $|\mu| \geqq |\delta| + 3\,|u| = |uvw\,\delta|$.

Reading the suffixes of the equality (13), we deduce that:

$$(15)\qquad \mu = \rho\,uvw\,\delta$$

for some word ρ, with $|\rho| = |\theta|$. If follows from (13) (reading prefixes/suffixes of the same length) that

$$(16)\qquad (\nu\,uvw)^i\,\zeta\,vwu\,\rho = \theta\,uvw\,(\nu\,uvw)^i\,\zeta$$

and

$$\delta\,uvw\,(\nu\,vwu)^j\,\eta\,uvw\,\rho = \theta\,vwu\,\delta\,uvw\,(\nu\,vwu)^j\,\eta.$$

Now, assuming that $i \geqq 1$ and using (16), we have:

$$(\nu\,uvw)^{i-1}\,\nu\,uvw\,\zeta\,vwu\,\rho = \theta\,uvw\,(\nu\,uvw)^{i-1}\,\nu\,uvw\,\zeta.$$

Using the equality $\zeta\,vwu\,\mu = \nu\,uvw\,\zeta$ (*see* Claim A_1) and using (15), the above equality becomes:

$$\begin{aligned}& (\nu\,uvw)^{i-1}\,\zeta\,vwu\,\rho\,uvw\,\delta\,vwu\,\rho \\ & = \theta\,uvw\,(\nu\,uvw)^{i-1}\,\zeta\,vwu\,\rho\,uvw\,\delta.\end{aligned}$$

Note that $|\delta\,vwu\,\rho| = |\rho\,uvw\,\delta|$. From the suffixes of the above equality, we obtain:

$$uvw\,\delta\,vwu\,\rho = vwu\,\rho\,uvw\,\delta.$$

But, this is contradicts $u \neq v$.

Therefore, the only possibility is $i = 0$. By a similar proof, we obtain that also $j = 0$.

Note that, from (14), $|\nu| \geqq |\theta| + 3|u| = |\theta\, uvw|$ and from prefixes of the equality $\alpha(x', y') = \alpha(x_2, y_2)$ we obtain that $\nu = \theta\, uvw\, \pi$, for some word π, with $|\pi| = |\delta|$.

Thus, from Claim A_1, (10) and (13), we obtain the first part of Claim A_2, *i. e.*:

$$\begin{aligned} x_1 &= w\zeta, & y_1 &= \rho\, uvw\, \delta\, uvw\, \eta \\ x' &= w\, \zeta\, uvw\, \theta & y' &= \delta\, uvw\, \eta, \\ x_2 &= w\, \zeta\, uvw\, \theta\, uvw\, \pi, & y_2 &= \eta. \end{aligned}$$

It remains to verify the conditions satisfied by the terminal words θ, δ, η, ρ, ζ, π, *i. e.*, the second part of Claim A_2.

Reading prefixes and suffixes of the same length in the equalities

$$\alpha(x_1, y_1) = \alpha(x', y') = \alpha(x_2, y_2)$$

we obtain the following 5 equalities:

$$\zeta\, vwu\, \rho = \theta\, uvw\, \zeta \tag{17}$$

(from prefixes in $\alpha(x_1, y_1) = \alpha(x', y')$),

$$\zeta\, uvw\, \theta\, vwu\, \delta = \pi\, uvw\, \zeta\, uvw\, \theta \tag{18}$$

(from prefixes in $\alpha(x', y') = \alpha(x_2, y_2)$),

$$\zeta\, vwu\, \rho\; uvw\, \delta = \theta\, uvw\, \pi\, uvw\, \zeta \tag{19}$$

(from prefixes in $\alpha(x_1, y_2) = \alpha(x_2, y_2)$),

$$\eta\, uvw\, \rho\; uvw\, \delta = \theta\, uvw\, \pi\, vwu\, \eta \tag{20}$$

(from suffixes in $\alpha(x_1, y_1) = \alpha(x_2, y_2)$),

$$\eta\, uvw\, \delta = \pi\, vwu\, \eta \tag{21}$$

(from suffixes in $\alpha(x', y') = \alpha(x_2, y_2)$).

From (17) and (19) we obtain:

$$\zeta\, uvw\, \delta = \pi\, vwu\, \zeta. \tag{22}$$

From (20) and (21) it follows that

$$\eta\, uvw\, \rho = \theta\, uvw\, \eta. \tag{23}$$

From (22) and (18) we obtain:

$$\theta\, vwu\, \delta = \delta\, uvw\, \theta. \tag{24}$$

Therefore, from (24), (23), (17), (22) and (21) we can conclude that the terminal words θ, δ, η, ζ, π, ρ should satisfy the following set E of equations:

$$\begin{aligned}
(e_1)\quad & \theta\, vwu\, \delta = \delta\, uvw\, \theta \\
(e_2)\quad & \eta\, uvw\, \rho = \theta\, uvw\, \eta \\
(e_3)\quad & \zeta\, vwu\, \rho = \theta\, uvw\, \zeta \\
(e_4)\quad & \zeta\, uvw\, \delta = \pi\, uvw\, \zeta \\
(e_5)\quad & \eta\, uvw\, \delta = \pi\, vwu\, \eta
\end{aligned}$$

with the supplementary conditions $|\theta| = |\rho|$ and $|\pi| = |\delta|$.

End of the proof of CLaim A_2.

We will complete the proof of Lemma 9, proving the following:

CLAIM A_3: *The system E from Claim A_2, with the supplementary conditions $|\theta| = |\rho|$ and $|\pi| = |\delta|$, does not have solutions.*

Proof of Claim A_3: First, it is easy to observe that from conditions (ii) and (iii) of Lemma A, it follows that for any two different words $t,\ t' \in \{u,\ v,\ w\}$, $first\,(t) \neq first\,(t')$ and $last\,(t) \neq last\,(t')$. Hence, without loss of generality, we can assume that: $first\,(u) = last\,(u) = a$, $first\,(v) = last\,(v) = b$, $first\,(w) = last\,(w) = c$.

Consider now the following two sets:

$$A = \{\theta,\ \delta,\ \eta,\ \zeta,\ \pi\}$$

and

$$B = \{\theta,\ \delta,\ \eta,\ \zeta,\ \rho\}.$$

From the system E, we can deduce that the words from the set A have the following important property: If $t \in A$, then for any $t' \in A$, $first\,(t) = first\,(t')$. Similarly, if $t \in B$, then for any $t' \in B$, $last\,(t) = last\,(t')$. The above property can be extended for prefixes of words from A (for suffixes of words from B).

Thus, we'll start a discussion concerning the possible shortest word among the words: θ, δ, η, ζ, π, ρ.

Case I_1 *:* The shortest word is θ and, moreover,

$$|\theta| < |\delta| = |\pi|,\ |\theta| < |\eta|,\ |\theta| < |\zeta|.$$

Note that $\theta, \rho \in B$, $|\theta| = |\rho|$ and therefore $\theta = \rho$. Let's denote: $\theta = \rho = \xi_1$. Hence, $\delta = \delta_1' \, \xi_1$, $\zeta = \zeta_1' \, \xi_1$, for some nonempty words, δ_1', ζ_1'. The equalities (e_1), (e_3) and (e_4) from E become:

(e_{11}) $\quad \xi_1 \, vwu \, \delta_1' \ = \ \delta_1' \, \xi_1 \, uvw$

(e_{31}) $\quad \zeta_1' \, \xi_1 \, vwu \ = \ \xi_1 \, uvw \, \zeta_1'$

(e_{41}) $\ = \ \zeta_1' \, \xi_1 \, uvw \, \delta_1' \ = \ \pi \, uvw \, \zeta_1'$

From (e_{11}) if follows that $last\,(\delta_1') = last\,(w) = c$. From (e_{31}) we obtain that $last\,(\zeta_1') = last\,(u) = a$ and from (e_{41}) it follows that $last\,(\delta_1') = last\,(\zeta_1')$, *i. e.* $a = c$, a contradiction.

Case I_2 *:* The shortest word is δ and, moreover,

$$|\delta| < |\theta| = |\rho|,\ |\delta| < |\eta|,\ |\delta| < |\zeta|.$$

Note that $\delta, \pi \in A$, $|\delta| = |\pi|$ and therefore $\delta = \pi$. Let's denote: $\delta = \pi = \xi_2$. Hence, $\theta = \xi_2 \, \theta_2'$, $\eta = \xi_2 \, \eta_2'$, for some nonempty words, θ_2', η_2'. The equalities (e_1), (e_2) and (e_5) from E become:

(e_{12}) $\quad \theta_2' \, vwu \, \xi_2 \ = \ uvw \, \xi_2 \, \theta_2'$

(e_{22}) $\quad \eta_2' \, uvw \, \rho \ = \ \theta_2' \, uvw \, \eta_2'$

(e_{52}) $\quad \eta_2' \, uvw \, \xi_2 \ = \ vwu \, \xi_2 \, \eta_2'$

From (e_{12}) it follows that $first\,(\theta_2') = first\,(u) = a$. From (e_{52}) we obtain that $first\,(\eta_2') = first\,(v) = b$ and from (e_{22}) it follows that $first\,(\eta_2') = first\,(\theta_2')$, *i. e.* $a = b$, a contradiction.

Case I_3 *:* The shortest word is η.

Denote $\eta = \xi_3$ and observe that $\delta = \delta_3' \, \xi_3$, $\rho = \rho_3' \, \xi_3$, $\zeta = \zeta_3' \, \xi_3$, for some nonempty words, δ_3', ρ_3', ζ_3'. The equalities (e_2), (e_3), (e_4) and (e_5) from E become:

(e_{23}) $\quad \xi_3 \, uvw \, \rho_3' \ = \ \theta \, \xi_3 \, uvw$

(e_{33}) $\quad \zeta_3' \, \xi_3 \, vwu \, \rho_3' \ = \ \theta \, \xi_3 \, uvw \, \zeta_3'$

(e_{43}) $\quad \zeta_3' \, \xi_3 \, uvw \, \delta_3' \ = \ \pi \, uvw \, \zeta_3'$

(e_{53}) $\quad \xi_3 \, uvw \, \delta_3' \ = \ \pi \, vwu$

From (e_{33}) and (e_{43}) it follows that $last\,(\rho_3') = last\,(\zeta_3') = last\,(\delta_3')$. From (e_{23}) we obtain that $last\,(\rho_3') = c$ and from (e_{53}) it follows that $last\,(\delta_3') = a$. Therefore, we obtain $a = c$, a contradiction.

Case I_4 *:* The shortest word is ζ.

Denote $\zeta = \xi_4$ and observe that $\theta = \xi_4 \theta'_4$, $\eta = \xi_4 \eta'_4$, $\pi = \xi_4 \pi'_4$, for some nonempty words, θ'_4, η'_4, π'_4. The equalities (e_2), (e_3), (e_4) and (e_5) from E become:

(e_{24}) $\eta'_4\, uvw\, \rho = \theta'_4\, uvw\, \xi_4 \eta'_4$

(e_{34}) $vwu\, \rho = \theta'_4\, uvw\, \xi_4$

(e_{44}) $uvw\, \xi_4\, \delta = \pi'_4\, uvw\, \xi_4$

(e_{54}) $\eta'_4\, uvw\, \xi_4\, \delta = \pi'_4\, vwu\, \xi'_4$

From (e_{24}) and (e_{54}) it follows that $first\,(\eta'_4) = first\,(\theta'_4) = first\,(\pi'_4)$. From (e_{34}) we obtain that $first\,(\eta'_4) = b$ and from (e_{44}) it follows that $first\,(\pi'_4) = a$. Therefore, we obtain $a = b$, a contradiction.

Now we consider the situation when the shortest length of a word in the set $C = \{\zeta, \eta, \theta, \delta\}$ is reached by exactly 2 words from this set. There are 6 possible combinations as follows:

Case J_1 *:* The shortest words are ζ and η, *i. e.* $|\zeta| = |\eta|$ and all others words from the set C are longer.

Because $\zeta, \eta \in A$, it follows that $\zeta = \eta = \xi_5$ and $\theta = \xi_5\, \theta'_5$ for nonempty word θ'_5.

The equalities (e_2) and (e_3) from E become:

(e_{25}) $uvw\, \rho = \theta'_5\, uvw\, \xi_5$

(e_{35}) $vwu\, \rho = \theta'_5\, uvw\, \xi_5$

Hence, $first\,(\theta'_5) = a$ and $first\,(\theta'_5) = b$, a contradiction.

Case J_2 *:* The shortest words from C are ζ and θ.

The equality (e_3) from E becomes:

(e_{36}) $vwu\, \rho = uvw\, \zeta$

Thus, $first\,(v) = first\,(u)$, a contradiction.

Case J_3 *:* The shortest words from C are ζ and δ.

Because $\zeta, \delta \in A$, and, moreover, $|\delta| = |\pi|$ it follows that $\zeta = \delta = \pi = \xi_7$ and $\theta = \xi_7\, \theta'_7$ for some nonempty word θ'_7.

The equalities (e_1) and (e_3) from E become:

(e_{17}) $\theta'_7\, vwu\, \xi_7 = uvw\, \xi_7\, \theta'_7$

(e_{37}) $vwu\, \rho = \theta'_7\, uvw\, \xi_7$

Hence, $first\,(\theta'_7) = a$ and $first\,(\theta'_7) = b$, a contradiction.

Case J_4 *:* The shortest words from C are η and θ.

Because η, $\theta \in B$, and, moreover, $|\theta| = |\rho|$ it follows that $\eta = \theta = \rho = \xi_8$ and $\delta = \delta_8' \xi_8$ for some nonempty word δ_8'.

The equalities (e_1) and (e_5) from E become:

(e_{18}) $\xi_8\, vwu\, \delta_8' = \delta_8' \xi_8\, uvw$

(e_{58}) $\xi_8\, uvw\, \delta_8' = \pi\, vwu$

Hence, $last\,(\delta_8') = c$ and $last\,(\delta_8') = a$, a contradiction.

Case J_5 *:* The shortest words from C are η and δ.

The equality (e_5) from E becomes:

(e_{59}) $\eta\, uvw = \pi\, vwu$

Thus, $last\,(w) = last\,(u)$, a contradiction.

Case J_6 *:* The shortest words from C are θ and δ.

The equality (e_1) from E becomes:

$(\bar{e}_{110})$ $vwu = uvw$

Thus, we obtain again a contradiction.

All the remaining cases lead to contradictions, because one can argue: $\eta \neq \pi$ (*see* (e_5)), $\delta \neq \theta$ (*see* (e_1)), $\zeta \neq \rho$ (*see* (e_3)) and also, $\zeta \neq \theta$ (*see* (e_3)).

This completes the proof of Claim A_3 and the proof of Lemma 9, too. □

COROLLARY 10: *The pattern,*

$$\beta = X\, ab\, X\, bca\, Y\, abc\, Y$$

has the degree of ambiguity 2.

Proof: Take in Lemma 9, $u = a$, $v = b$ and $w = c$ and note that conditions (i)-(iv) are satisfied. □

Notations: Let γ be the unambiguous pattern from Theorem 5 and let α_1 be the pattern α from Lemma 9, for:

$$u = a^4\, ba^8\, ba^4, \qquad v = ba^7\, bcacaabcab, \qquad w = caabcbcabcabcbcabc.$$

LEMMA 11: *The pattern,*

$$\sigma = \gamma\,(\alpha_1\,(X,\, Y),\, X\,\beta\,(Q,\, R))$$

with the variables X, Y, Q, R *has the degree of ambiguity 3.*

Proof: Let α_0 be the pattern from Lemma 9 for some fixed values u_0, v_0, w_0 of the terminal words, u, v, w. Consider the pattern:

$$\sigma' = \gamma(\alpha_0(X, Y), X\beta(Q, R))$$

where γ is the unambiguous pattern from Theorem 5 and β is the pattern from Corollary 10.

CLAIM C_1 : *The pattern σ' has the degree of ambiguity at most 3.*

Proof of Claim C_1 : Observe that the equality

$$\sigma'(X, Y, Q, R) = t,$$

for some terminal word $t \in \Sigma^*$, leads to the equalities:

$$\alpha_0(X, Y) = t_1, \qquad X\beta(Q, R) = t_2,$$

for some unique words t_1, $t_2 \in \Sigma^*$ (γ is an unambiguous pattern).

Consider now the following two possible situations:

Case K_1 : The equation $\alpha_0(X, Y) = t_1$ has at most one solution, (x_0, y_0).

It follows that the equation $x_0\beta(Q, R) = t_2$ leads to at most one value of $\beta(Q, R)$. Consequently, because β has the degree of ambiguity 2, there are at most two pairs of terminal words, (q', r'), (q'', r'') that satisfy the equality $x_0\beta(Q, R) = t_2$. Hence, in this case, t has at most 2 possible decompositions with respect to σ'.

Case K_2 : The equation $\alpha_0(X, Y) = t_1$ has 2 solutions, (x_1, y_1), (x_2, y_2).

If follows from Claim A_1 (*see* the proof of Lemma 9) that

$$x_1 = w_0(\nu u_0 v_0 w_0)^i \zeta, \qquad x_2 = w_0(\nu u_0 v_0 w_0)^i \zeta u_0 v_0 w_0 \nu,$$

for some terminal words ν, ζ, and $i \geqq 0$.

From the equality $x_2\beta(Q, R) = t_2$, we obtain

$$\text{(I)} \qquad \beta(Q, R) = t'_2$$

and from the equality $x_1\beta(Q, R) = t_2$ it follows that

$$\text{(II)} \qquad \beta(Q, R) = u_0 v_0 w_0 \nu t'_2.$$

The equation (I) has at most 2 solutions, say (q_1, r_1) and (q_2, r_2). The equation (II) has at most 1 solution (q_3, r_3), because the value of $\beta(Q, R)$ has a prefix that starts with $u_0 v_0 w_0$. (*See* the proof of Lemma 9, Claim A_1.)

Therefore, altogether, t has at most 3 decompositions with respect to σ', corresponding to:

$$(x_1, y_1, q_3, r_3), \qquad (x_2, y_2, q_1, r_1), \qquad (x_2, y_2, q_2, r_2).$$

End of the proof of Claim C_1.

CLAIM C_2: *For the values:*

$$u_0 = a^4\, ba^8\, ba^4, \quad v_0 = ba^7\, bcacaabcab, \quad w_0 = caabcbcabcabcbcabc,$$

the pattern σ' $(= \sigma)$ *has the degree of ambiguity 3.*

Proof of Claim C_2*:* From Lemma 9, we have that:

$$\beta\,(ca,\, abcbc) = \beta\,(caabc,\, bc) = \Psi,$$

where $\Psi = caabcabcaabcbcabcabcbc$.

In a similar way, we obtain that:

$$\alpha_0\,(wu,\, uvwvw) = \alpha_0\,(wuuvw,\, vw) = \Psi',$$

where $\Psi' = h\,(\Psi)$ and the morphism h is defined by: $h\,(a) = u$, $h\,(b) = v$, $h\,(c) = w$.

Notations:

$$\begin{array}{ll} x_1 = wu, & y_1 = uvwvw, \\ x_2 = wuuvw, & y_2 = vw, \\ q_1 = ca, & r_1 = abcbc, \\ q_2 = caabc, & r_2 = bc. \end{array}$$

Note that in all the above solutions we have $\nu = \lambda$.

Now we'll define the values of u, v, w in such a way that the equation

$$(*) \qquad \beta\,(Q,\, R) = uvw\,\beta\,(q_1,\, r_1) = uvw\,\Psi$$

will have a solution (q_3, r_3). (Note that this equation has at most 1 solution, because u is starting with a.) Moreover, the values of u, v, w will satisfy the conditions (i)-(iv) from Lemma 9.

The equation (*) becomes:

$$(**) \qquad Q\,ab\,Q\,bca\,R\,abc\,R = uvw\,\Psi.$$

For simplicity, we can consider that $R = \Psi$. Hence, we obtain from (**):

$$(***) \qquad Q\,ab\,Q\,bca\,\Psi\,abc = uvw.$$

It is easy to check that there is no value of Q with $|Q| \leqq 11$ such that the corresponding values of u, v, w are satisfying the conditions (i)-(iv) of Lemma 9. Assume now that $|Q| = 12$. It follows from (***) that $|u| = |v| = |w| = 18$. The resulting value of w is $w_0 = caabcbcabcabcbcabc$. Using (***), it is easy to observe that we can choose the value of v as being $v_0 = ba^7\, bcacaabcab$, and the value of u as being $u_0 = a^4\, ba^8\, ba^4$.

Note that u_0, v_0, w_0 satisfy the requirements (i)-(iv) of Lemma 9.

End of the proof of Claim C_2. □

THEOREM 12: *Explicit examples of patterns with degrees of ambiguity 2 and 3 can be given.*

Comment: Our example of a pattern of degree 3 has length 324 and the shortest word that actually has 3 different decompositions with respect to this pattern has length 1018. On the other hand, our example of a pattern of degree 2 given in Corollary 10 is rather simple: $X\, ab\, X\, bca\, Y\, abc\, Y$.

By forming compositions and using Theorem 6, our last results is obtained as a corollary of Theorem 12.

THEOREM 13: *For any* $m \geqq 0$ *and* $n \geqq 0$, *a pattern with the degree of ambiguity* $2^m\, 3^n$ *can be effectively constructed.*

It is worth mentioning that we have not been able to find any inductive way of going from the degree of ambiguity k to the degree $k+1$. Thus, we cannot exhibit patterns with an arbitrarily given finite degree of ambiguity, although we can do so for patterns whose degree of ambiguity is arbitrarily high.

6. CONCLUSION. OPEN PROBLEMS.

Our results deal with patterns and pattern languages and, thus, are interconnected with all related areas, already indicated in the Introduction. However, the results can also be viewed to concern the basic theory of word equations as follows.

Let $P(X_1, \ldots, X_n)$ be a pattern of n variables X_i. The pattern P defines infinitely many individual equations

$$(*) \qquad P(X_1, \ldots, X_n) = Z,$$

where Z ranges over Σ^+. For given P and Z, we denote by $N(P, Z)$ the number of solutions of (*), that is, the number of n-tuples of words $(w_1, \ldots, w_n)$ over Σ^* satisfying (*). For each pair (P, Z), $N(P, Z)$ is a nonnegative integer. For a *fixed* P, there are three possibilities.

(i) $N(P, Z) \leqq 1$, for all Z.

(ii) There is a Z' such that $N(P, Z') > 1$ but the numbers $N(P, Z)$ possess an upper bound, that is, for some k, $N(P, Z) \leqq k$ for all Z.

(iii) The numbers $N(P, Z)$ possess no upper bound, that is, for every k, $N(P, Z') > k$ holds for some Z'.

We have been able to exhibit extensive classes of patterns for which (i) or (iii) holds. For instance, (i) holds if the number of variables $n = 1$, and (iii) holds if $n > 1$ and P is "linear" with respect to some variable (*see* Theorems 3 and 4). According to our main results (Theorems 12 and 13), also (ii) is possible. However, it is an open problem, and in our estimation a very fundamental one in the theory of word equations, whether *all finite degrees* of ambiguity can actually by constructed. By theorem 6, it suffices to carry out the construction for prime degrees. We conjecture that such a construction is possible. Since arbitrarily large degrees can be obtained (Theorem 13), it would seem rather strange if some degrees were "missing".

The most interesting open decision problem is the decidability status of (iii). "Almost all" patterns seem to satisfy (iii), and yet Makanin's Theorem is not directly applicable to this case.

REFERENCES

1. D. Angluin, Finding patterns common to a set of strings, *Journal of Computer and System Sciences*, 1980, *21*, pp. 46-62.
2. J. Bean, A. Ehrenfeucht, and G. McNulty, Avoidable patterns in strings of symbols, *Pacific Journal of Mathematics*, 1979, *85*, pp. 261-294.
3. J. Cassaigne, Unavoidable binary patterns, *Acta Informatica*, 1993, *30*, pp. 385-395.
4. K. Culik II and J. Karhumäki, Systems of equations over a free monoid and Ehrenfeucht's conjecture, *Discrete Mathematics*, 1983, *43*, pp. 139-153.
5. T. Jiang, E. Kinber, A. Salomaa, K. Salomaa and S. Yu, Pattern languages with and without erasing, to appear in *International Journal of Computer Mathematics.*
6. T. Jiang, A. Salomaa, K. Salomaa and S. Yu, Inclusion is undecidable for pattern languages, ICALP-93 Proceedings, *Springer Lecture Notes in Computer Science*, 1993, *700*, pp. 301-312.
7. V. Keränen, Abelian squares can be avoided on four letters, *Springer Lecture Notes in Computer Science*, 1992, *623*, pp. 41-52.
8. G. Kucherov, M. Rusinowitch, *On ground reducibility problem for word rewriting systems with variables*, Centre de Recherche en Informatique de Nancy, Report CRIN 93-R-012.
9. G. S. Makanin, The problem of solvability of equations in a free semigroup (in Russian), *Matematiceskij Sbornik*, 1977, *103*, 145, pp. 148-236.
10. A. Salomaa, *Formal Languages*, Academic Press, 1973.
11. A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Kristiania*, 1906, *7*, pp. 1-22.