GRAPH CONGRUENCES AND PAIR TESTING

J.A. Brzozowski
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Robert Knast
Institute of Mathematics
Polish Academy of Sciences
Poznan, Poland

CS-84-25
September, 1984

This research was supported in part by the Natural Sciences and Engineering Research Council under grant No. A-1617.

Abstract

This paper considers the congruence $_2\sim$ on a free monoid where $u \; _2\sim \; v$ iff u and v have the same letters and the same ordered pairs of letters. The motivation for this comes from the study of bi-locally testable languages defined by testing pairs of words. As in the case of locally testable languages, a theorem on graph congruences is used in order to obtain a characterization of the family of bi-locally testable languages. Such a theorem on graph congruences is developed in this paper.

1. Introduction

The family of locally testable languages plays a key role in the study of star-free languages. It is defined as follows: The membership of a word w in a language L is uniquely determined by the prefix of length $k - 1$ of w, the suffix of length $k - 1$ of w, and the set of all segments of length k appearing in w, where $k \geq 1$ is an integer depending on L. The syntactic semigroup S that corresponds to a locally testable language L satisfies the condition that for each idempotent $e \in S$, the monoid eSe is idempotent and commutative. Conversely if S is the syntatic semigroup of L and S is finite and satisfies the above-mentioned conditions on eSe, then L is locally testable. The proof of this last statement is quite difficult. One of the key steps in this proof is a theorem on graphs. This theorem, due to Simon, appeared originally in [2], though it was not formulated as a separate result on graphs. The treatment of the theorem as a theorem on directed graphs is due to Eilenberg [3]. The theorem involves a congruence $_1\sim$ that corresponds to $k = 1$ in the test described above. More precisely, the prefix and suffix are not tested (since $k - 1 = 0$), and only segments of length one (i.e. letters) are considered.

The next family in the hierarchy of languages of depth one [1], after the locally testable family, is that of bi-locally testable languages. Membership of a word w in a bi-locally testable language is determined by the prefix and suffix of length $k - 1$ of w, and by the set of ordered pairs of segments of length k that appear in w. The characterization of syntactic semigroups of bi-locally testable languages is due to Knast [4], and uses the theorem on graphs presented in this paper as one of the basic steps. The theorem involves the congruence $_2\sim$ that again corresponds to $k = 1$. This time, however, ordered pairs of letters are used.

2. The Main Theorem

We first briefly recall Eilenberg's notation for graphs [3].

A directed graph G consists of two possibly infinite sets V (vertices) and E (edges) along with two functions

$$\alpha, \omega : E \to V$$

If e is an edge, $e\alpha$ and $e\omega$ are the initial and final vertices of e. Two edges e_1 and e_2 are consecutive iff $e_2\alpha = e_1\omega$. Let E^+ (E^*) be the free semigroup (free monoid) generated by E, and let $C \subseteq E^2$ be the set of words e_1e_2 such that e_1 and e_2 are non-consecutive. The set of (non-empty) paths of G is then

$$P = E^+ - E^* C E^*$$

If $p = e_1 \ldots e_n$ is a path, define $p\alpha = e_1\alpha$ and $p\omega = e_n\omega$. The length of the path is $|p| = n$, where $n \geq 1$. A path p is a loop about vertex v iff $v = p\alpha = p\omega$. If $p = e_1 \ldots e_n$, $q = e_1' \ldots e_m'$, and $p\omega = q\alpha$ then p and q are consecutive and $pq = e_1 \ldots e_n\, e_1' \ldots e_m'$

is a path. For any vertex v, 1_v is a loop of length 0 about v, i.e. $1_v\alpha = 1_v\omega = v$. For technical reasons we assume that the set $\{1_v \mid v \in V\}$ of trivial paths is adjoined to P. Two paths p and p' are coterminal iff $p\alpha = p'\alpha$ and $p\omega = p'\omega$. An equivalence relation $\sim$ on P is a congruence iff

(i) $p \sim p'$ implies p and p' are coterminal.

(ii) If $p \sim p'$, $q \sim q'$ and p and q are consecutive, then $pq \sim p'q'$.

Let $\tau\colon\ E^* \to 2^E$ be the function that associates with each word w in E^* the set of edges (letters) appearing in w:

$$w\tau = \{e \in E \mid w = w_1 e w_2 \quad \text{for some } w_1, w_2 \in E^*\}.$$

Similarly let $w\tau_2$ be the set of ordered pairs of edges in w:

$$w\tau_2 = \{(e_1,e_2) \in E \times E \mid w = w_0 e_1 w_1 e_2 w_2,\ w_0, w_1, w_2 \in E^*\}.$$

We define the following congruence on E^*. Given $x, y \in E^*$

$$x \;{}_2\!\sim y \quad \text{iff} \quad x\tau_2 = y\tau_2 \ \text{ and } \ x\tau = y\tau.$$

If p is a path of length > 0, then $p\tau$ and $p\tau_2$ are defined as above. If $p = 1_v$ for some $v \in V$ then $p\tau = p\tau_2 = \phi$.

<u>Theorem</u> Let $\sim$ be the smallest congruence on P satisfying

$$z_1(pq)^2pzr(sr)^2z_2 \sim z_1(pq)^2z'(sr)^2z_2 \tag{1}$$

for all $p,q,r,s,z_1,z_2,z,z' \in P$ such that

$$z\tau \subseteq z_1\tau \cap z_2\tau \quad \text{and} \quad z'\tau \subseteq z_1\tau \cap z_2\tau$$

Then for any two coterminal paths x and y the conditions $x \sim y$ and $x \ {}_2{\sim} \ y$ are equivalent.

The proof of this result is the subject of the rest of this paper. Before proceeding with the proof we make the following comments. The congruence ${}_2{\sim}$ involves testing the set $w\tau_2$ of pairs of letters appearing in a word w (or the set $w\tau$ in case $w\tau_2 = \phi$, i.e. $|w| \leq 1$), and is defined on E^*. The theorem states that the equivalence of any two coterminal paths with respect to ${}_2{\sim}$ can always be demonstrated by coterminal path transformations of the form (1). It is easily verified that

$$x \sim y \text{ implies } x \ {}_2{\sim} \ y \tag{2}$$

The converse of (2) constitutes the problem.

Rule (1) is quite complex as compared to the rules in Simon's theorem, where the rules corresponding to (1) are

$$x \sim x^2 \text{ and } xy \sim yx$$

for any two coterminal loops x and y. We were unable to simplify Rule (1) or to replace it by a set of equivalent or weaker rules. The graph of Fig. 1 provides an example of the difficulty involved. Consider the coterminal paths

$$x = c'd_1cd_2(a_1a_2)^2a_1cb_1(b_2b_1)^2e_1ce_2c'$$

and $$y = c'd_1cd_2(a_1a_2)^2c'(b_2b_1)^2e_1ce_2c'$$

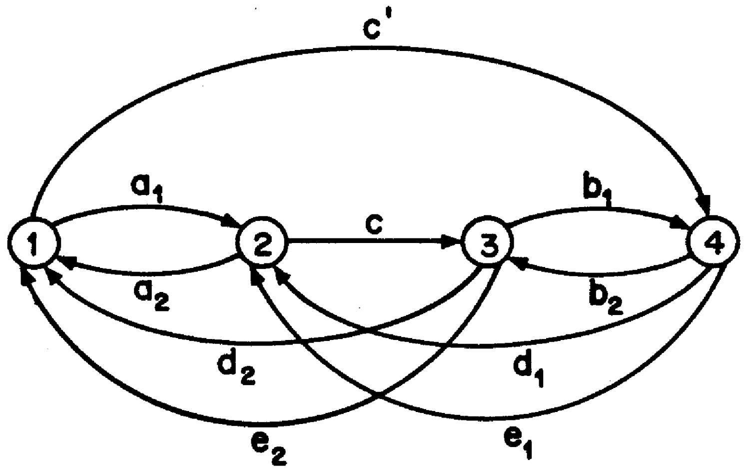

Figure 1

One easily verifies that $x \underset{2}{\sim} y$. If we let $z_1 = c'd_1cd_2$ and $z_2 = e_1ce_2c'$, we have an instance where Rule (1) applies. We were unable to find a simpler set of rules for this example.

In a number of cases Rule (1) degenerates to considerably simpler rules. It will be convenient to identify them distinctly, even though they are covered by (1). If $z\tau,\ z'\tau \subset z_1\tau \cap z_2\tau$ then:

$$z_1zz_2 \sim z_1z'z_2 \tag{1a}$$

$$z_1(pq)^2pzz_2 \sim z_1(pq)^2z'z_2 \tag{1b}$$

$$z_1zr(sr)^2z_2 \sim z_1z'(sr)^2z_2 \tag{1c}$$

3. Singularities

Let A be a finite alphabet and $x \in A^*$. If $x = x_1 a x_2$, $a \in A$ and $a \notin (x_1 x_2)\tau$ then a is a singular letter of x. If $x = x_0 a x_1 b x_2$ where a and b are not singular letters of x and $(b, a) \notin x\tau_2$, then (a, b) is a singular pair of x. Singular letters and singular pairs are called singularities of x. If $x = x_0 a x_1 b x_2$, this factorization is an occurrence of (a, b). An occurrence is inner if $a \notin x_1\tau$, $b \notin x_1\tau$. Clearly every singular pair (a, b) has a unique inner occurrence consisting of the rightmost a of x and the leftmost b. An occurrence $x_0 a x_1 b x_2$ is proper if ax_1 and $x_1 b$ have no singularities of x; note that every proper occurrence is necessarily inner. A singular pair need not necessarily have a proper occurrence. For example, let $x = aebbacdfdfc$. Then e is the only singular letter of x and $(a, c), (a, d), (a, f), (b, c), (b, d), (b, f)$ are the singular pairs of x. The factorization $(aeb)b(ac)d(fdfc)$ shows the inner occurrence of (b, d). Only (a, c) has a proper occurrence, namely $(aebb)a(1)c(dfdfc)$.

Proposition 1. Let (a, b) be a singular pair of x.

(a) Let $x = x_0 a x_1 b x_2$ be an inner occurrence. Then

$$a \in x_0\tau - (x_1 b x_2)\tau \qquad b \in x_2\tau - (x_0 a x_1)\tau$$

(b) Let $x = x_0 a x_1 b x_2$ be a proper occurrence. Then

$$x_1\tau \subset x_0\tau \cap x_2\tau$$

(c) Let $x \;{}_2\!\sim y$ and let $x = x_0 a x_1 b x_2$ and $y = y_0 a y_1 b y_2$ be inner occurrences. Then

$$x_0\tau = y_0\tau \qquad x_2\tau = y_2\tau$$

(d) Let $x \underset{2}{\sim} y$ and let $x = x_0ax_1bx_2$ be proper and $y = y_0ay_1by_2$ be inner. Then y_1 has no singular letters of x.

Proof:

(a) If $a \in x_2\tau$ then $(b, a) \in x\tau_2$ contradicting that (a, b) is singular. If $a \in x_1\tau$ then the occurrence shown is not inner. If $a \notin x_0\tau$ then a is a singular letter of x, contradicting that (a, b) is a singular pair. The same arguments apply to the claim about b.

(b) Let $c \in x_1\tau$; then $(a, c) \in x\tau_2$. The pair (a, c) cannot be singular because the occurrence of (a, b) as shown is proper. Hence $(c, a) \in x\tau_2$. Since $a \notin (x_1bx_2)\tau$, we must have $c \in x_0\tau$. Thus $x_1\tau \subset x_0\tau$, and $x_1\tau \subset x_2\tau$ follows similarly.

(c) $c \in x_0\tau$ implies $(c, a) \in x\tau_2 = y\tau_2$. Hence $c \in y_0\tau$, and $x_0\tau \subset y_0\tau$. Similarly $y_0\tau \subset x_0\tau$ and the claim follows. By symmetry $x_2\tau = y_2\tau$.

(d) If $c \in y_1\tau$ is singular then $(c, a),(b, c) \notin y\tau_2$. Since $x\tau_2 = y\tau_2$, c must occur exactly once in x_1, to satisfy these conditions and the condition that c is a singular letter of x. But this contradicts the assumption that $x_0ax_1bx_2$ is proper.

Proposition 2. Proper occurrences of singular pairs do not overlap, i.e. suppose $x = x_0ax_1bx_2$ and $x = y_0cy_1dy_2$ where the occurrences are proper; then either $|x_0| \geq |y_0cy_1d|$ or $|y_0| \geq |x_0ax_1b|$, and a, b, c, d are all distinct.

Proof: Without loss of generality, assume that $|x_0| \leq |y_0|$. Then cy_1d is to the right of x_0. Suppose first the overlap has the form $b = c$ and $x = x_0ax_1by_1dy_2$. Then $b \notin (x_0ax_1)\tau$ because (a, b) is inner as shown and $b \notin (y_1dy_2)$ because $(b, d) = (c, d)$ is inner as shown. Hence b is a singular letter, contradicting that (a, b) is a singular pair. Thus this type of overlap cannot occur. Next suppose $x = x_0ax_{11}cx_{12}by_{12}dy_2$. We know $a \neq b$ and $c \neq d$. If $c = b$ then the occurrence $(x_0)a(x_{11}cx_{12})b(y_{12}dy_2)$ of (a, b) is not inner; hence $c \neq b$. Now $b \notin (x_0ax_{11}cx_{12})\tau$ because (a, b) is inner. Also $c \notin (x_{12}by_{12}dy_2)\tau$ because (c, d) is inner. Hence (c, b) is a singular pair of x, contradicting that the occurrence of (a, b) is proper. Again, this type of overlap cannot occur. Thirdly, if $a = c$, then $x = x_0ax_1by_{12}dy_2$ and the occurrence of (a, d) cannot be proper. This is a contradiction. Similarly we can't have $b = d$. Finally, we can't have (c, d) occur in x_1 because the occurrence of (a, b) is proper. Hence, no overlap can occur.

We already know that $a \neq b$, $a \neq c$, $b \neq c$, $b \neq d$, and $c \neq d$. One verifies also that $a \neq d$.

4. Alignment of Singularities

We introduce the following notation to reduce the number of cases that have to be considered. Let

$$u\underline{awb}v$$

represent the usual word uawbv, with $a,b \in A$, or the word uav. The latter case occurs when $w = 1$ and $a = b$. Frequently it is possible to handle both cases by the same arguments, and this notation permits this.

<u>Proposition 3</u>. Let $x = x_0ax_1bx_2$ be a proper occurrence of (a, b). Suppose $y \,{}_2\!\sim x$ and $y = y_0ay_1by_2$ where the occurrence of (a, b) is inner. Then either the occurrence of (a, b) in y is proper or ay_1b contains exactly one proper occurrence of a singular pair of x.

<u>Proof</u>: Suppose (a, b) in y is not proper. By Proposition 1(d) y_1 has no singular letters; hence it must have at least one singular pair. Suppose it has two proper occurrences of singular pairs. By Proposition 2 they do not overlap, so y has the form

$$y = y_0a\underline{y_{10}c}y_{11}dy_{12}ey_{13}\underline{fy_{14}}by_2$$

where (c, d) and (e, f) are the two proper occurrences. Now

$(d, e) \in y\tau_2 = x\tau_2$

$(b, e) \notin y\tau_2$ because b is leftmost and e is rightmost

$(d, a) \notin y\tau_2$ because a is rightmost and d is leftmost.

Thus (e, b) and (a, d) are singular pairs of x. Therefore $d \notin x_0\tau$, and $d \notin x_1\tau$ because $x_0ax_1bx_2$ shows a proper pair (a, b). Similarly $e \notin x_2\tau$ and $e \notin x_1\tau$. Hence (d, e) cannot occur in x. This is a contradiction, showing that exactly one singular pair can be proper in y_1.

Proposition 4. Let $x = x_0ax_1bx_2$ be a proper occurrence of (a, b) in x. Suppose that $x \,_2\!\sim y$ but y has no proper occurrence of (a, b). By Proposition 3 y has the form $y = y_0\underline{ay_{10}c}y_{11}\underline{dy_{12}b}y_2$ where the occurrence of (a, b) is inner, either $a \neq c$ or $b \neq d$, and the occurrence of (c, d) is proper. Then

$$x = x_{01}\underline{cx_{02}a}x_1\underline{bx_{21}d}x_{22}$$

where the occurrence of (c, d) is inner.

Proof: Observe that $(a, d) \in y\tau_2$ but $(d, a) \notin y\tau_2$ because a is rightmost and d is leftmost. Hence $(a, d) \in x\tau_2$ and $(d, a) \notin x\tau_2$. Thus $d \notin x_0\tau$. Also $d \notin x_1\tau$ because the singular pair (a, d) would appear in ax_1b and the latter is assumed to be proper. Thus $d \in (bx_2)\tau$ and $x = x_0ax_1\underline{bx_{21}d}x_{22}$, where $d \notin x_{21}\tau$. Similarly, $(c, b) \in x\tau_2$, $(b, c) \notin x\tau_2$ and $x_0a = x_{01}\underline{cx_{02}a}$, giving the desired form for x.

Lemma 1. Let $x \,_2\!\sim y$, where x and y are coterminal paths in a graph. Then there exists $y' \sim y$ such that a proper occurrence of a singularity exists in x iff it exists in y'. Further, if $x = x_0ax_1bx_2$ where (a, b) is proper, then $y' = y_0'ax_1by_2'$.

Proof:

(i) If $x = x_1ex_2$ where e is a singular letter, we must have $y = y_1ey_2$, since the occurrence of a singular letter is always proper.

(ii) Suppose $x = x_0 a x_1 b x_2$ and $y = y_0 a y_1 b y_2$ where both occurrences are proper. By Proposition 1(b), $x_1\tau \subset x_0\tau \cap x_2\tau$ and $x_0\tau = y_0\tau$, $x_2\tau = y_2\tau$ by Proposition 1(c). Thus $x_1\tau \subset y_0\tau \cap y_2\tau$. Also $y_1\tau \subset y_0\tau \cap y_2\tau$. Since x_1 and y_1 are coterminal paths, we can apply Rule (1a):

$$y = (y_0 a) y_1 (b y_2) \sim (y_0 a) x_1 (b y_2) = y' \ .$$

(iii) Suppose y is as above, but the occurrence of (a, b) is not proper. Then, by Proposition 3,

$$y = y_0 a \underline{y_{10}} c y_{11} \underline{d y_{12}} b y_2 \tag{3}$$

where (c, d) is proper and (a, b) is inner and either $a \neq c$ or $d \neq b$ or both. Then, by Proposition 4,

$$x = x_{01} \underline{c x_{02}} a x_1 \underline{b x_{21}} d x_{22} \tag{4}$$

where (a, b) is proper, (c, d) is inner and either $a \neq c$ or $b \neq d$ or both.

<u>Case 1: $a \neq c$, $b = d$</u>

We have the following factorizations:

$$x = x_{01} c x_{02} a x_1 b x_2,$$
$$y = y_0 a y_{10} c y_{11} b y_2.$$

Let $u = y_{10} c y_{11} b y_2$, so that $y = y_0 a u$ where a is rightmost. Then $a \notin u\tau$ and $(x_{02}a)\tau \not\subset u\tau$. However, $(x_{02}a)\tau \subset y\tau$ because $x \underset{2}{\sim} y$ implies $x\tau = y\tau$. Therefore there must exist precisely one suffix $w = e\underline{y_{02}}au$ of y such that $(x_{02}a)\tau \subset w\tau$ but $(x_{02}a)\tau \not\subset (\underline{y_{02}a}\ u)\tau$, where $\underline{y_{02}a}$ denotes $y_{02}a$ when $e \neq a$ and $\underline{y_{02}a} = 1$, when $e = a$. Note that $e \notin (\underline{y_{02}a}\ u)\tau$ and also that e must be a letter of $x_{02}a$; let $x_{02}a = x'_{02}e\underline{x''_{02}a}$, where where $e \notin x''_{02}\tau$. Then

$$x = x_{01}cx'_{02}e\underline{x''_{02}}ax_1bx_2$$
$$y = y_{01}\underline{ey_{02}}ay_{10}cy_{11}by_2 = y_{01}w \; .$$

Consider the loop $h = e\underline{y_{02}}ay_{10}cx'_{02}$. We claim that this loop can be inserted after y_{01} in y by using Rule (1a). For we have $(e\underline{y_{02}}ay_{10}c)\tau \subset w\tau$ by the definition of w above. Also $x'_{02}\tau \subset (x_{02}a)\tau \subset w\tau$. Thus $h\tau \subset w\tau$.

Next we must verify that $h\tau \subset y_{01}\tau$. By construction e is rightmost in y . Thus $f \in cx'_{02}\tau$ implies $(f, e) \in x\tau_2 = y\tau_2$ and $f \in y_{01}\tau$. Hence $cx'_{02}\tau \subset y_{01}\tau$. In fact we have $(x_{01})\tau \subset y_{01}\tau$ by the same argument. Now $f \in (\underline{ey_{02}}ay_{10})\tau$ implies $(f, c) \in y\tau_2 = x\tau_2$ and $f \in x_{01}\tau$ because c is rightmost in x as shown. Thus $f \in y_{01}\tau$. Altogether, $h\tau \subset y_{01}\tau$. Inserting two copies of the loop h we have

$$y = y_{01}\underline{ey_{02}}ay_{10}cy_{11}by_2$$
$$\sim y_{01}(\underline{ey_{02}}ay_{10}cx'_{02})^2\underline{ey_{02}}ay_{10}cy_{11}by_2$$
$$= y_{01}\underline{ey_{02}}a(y_{10}cx'_{02}\underline{ey_{02}}a)^2y_{10}cy_{11}by_2$$

Let $z_1 = y_{01}\underline{ey_{02}a}$, $p = y_{10}c$, $q = x'_{02}\underline{ey_{02}a}$, $z = y_{11}$, and $z_2 = by_2$. Then

$$y \sim z_1(pq)^2pzz_2 \; .$$

We now show that $z\tau \subset z_1\tau \cap z_2\tau$. In fact, $f \in y_{11}\tau$ implies $(c, f) \in y\tau_2$ and so (f, c) in $y\tau_2 = x\tau_2$ because

$(c, b) = (c, d)$ is proper in y. Thus $f \in x_{01}\tau \subset y_{01}\tau$, and we have $f \in z_1\tau$. Therefore $z\tau \in z_1\tau$. Similarly $f \in y_{11}\tau$ implies $(f, b) \in y\tau_2$ and $(b, f) \in y\tau_2$. Hence $f \in y_2\tau$ and $z\tau \subset z_2\tau$.

Let $z' = x_1$. Then $x_1\tau \subset z_1\tau \cap z_2\tau$ by similar arguments. We are now in a position to apply Rule (1b):

$$\begin{aligned} y &\sim z_1(pq)^2pzz_2 \\ &\sim z_1(pq)^2z'z_2 \\ &= y_{01}\underline{ey_{02}}a(y_{10}cx'_{02}\underline{ey_{02}a})^2x_1by_2 \\ &= [y_{01}(\underline{ey_{02}a}y_{10}cx'_{02})^2ey_{02}]ax_1by_2 \\ &= y'_1ax_1by'_2 = y' \end{aligned}$$

which has the desired form. We can also write

$$y' = y_{01}ey_{02}g^2ax_1by_2 = y_0g^2ax_1by_2$$

where $g = ay_{10}cx'_{02}ey_{02}$. Recall that proper singularities do not overlap. In $y = y_0ay_{10}cy_{11}by_2$ we have the proper singularities in y_0ay_{10} and in y_2 and the pair (c, b). By Proposition 3 the segment $ay_{10}cy_{11}b$ has only one proper singularity; hence there are none in ay_{10}. Now in y' we have the proper singularities of y_0ay_{10} and y_2 and the pair (a, b) which replaced (c, b). The segment g^2 is free of singularities, since each pair $(f, f') \in g\tau \times g\tau$ appears at least twice in g^2 if $f \neq f'$, and g^2 can't have any singular letters. This leaves the possibility that there is a proper singularity in y_0g of the type $f \in y_0$, $f' \in g$. But $g\tau \subset y_{01}\tau \subset y_0\tau$. Hence either $(f', f) \in y_0\tau_2$ and (f, f') is not singular, or $(f, f') \in y_0\tau_2$ and the signularity in y_0g was not proper. Thus y' has only the proper singularities of y with (c, d) replaced by (a, b).

Case 2: $a = c$, $b \neq d$

This follows by left-right symmetry from Case 1. This time a loop is inserted on the right side and Rule (1c) is applied.

Case 3: $a \neq c$, $b \neq d$

Proceed as in Case 1 inserting first the left loop, then the right loop, and apply Rule (1).

In all cases of (iii) we can transform y into y' in such a way that the proper singularities of y' are the same as those of y except that (c, d) has been replaced by (a, b).

Now consider two words $x, y \in A^*$ such that $x \;{}_2\!\sim y$. Clearly each singular letter of x must also be a singular letter of y and vice versa. Also, if (a, b) has a proper occurrence in x then either (a, b) is also proper in y , or (a, b) occurs in y with another proper pair (c, d) , as in Propositions 3 and 4. As shown above, we can find y' such that $y' \sim y$ and the singularities of y' are those of y , with the exception that (c, d) has been replaced by (a, b). By repeating this process we find $y' \sim y$ such that y' has exactly the same singularities as x. It is easily verified that these singularities must appear in y' in the same order as in x. Thus we may assume at this point that x and y have the same singularities and that they have the form:

$$x = x_0 s_1 x_1 s_2 \cdots s_m x_m$$
$$y = y_0 s_1 y_1 s_2 \cdots s_m y_m$$

where $m \geq 0$, x_i , $i = 0, \ldots, m$, do not have any singularities of x and either $s_i = e$, $e \in A$, or $s_i = aw_ib$ is a proper singular pair of x.

5. Segments Between Singularities

Refer to the factorizations of x and y above that show all the proper singularities. In this section we will show that the segments y_i between proper singularities can be replaced by the segments x_i by using only Rule (1). The main result here is Lemma 2, but we need several preliminary results first.

Proposition 5. Let

$$x = \bar{x}_1\bar{x}_2\bar{x}_3 = (x_0 s_1 \ldots x_i s_i) x_{i+1} (s_{i+1} x_{i+2} \cdots s_m x_m) ,$$

$i \geq 0$, $m \geq 0$, where $\bar{x}_1 = x_0 s_1 \cdots x_i s_i$, $\bar{x}_2 = x_{i+1}$, and $\bar{x}_3 = (s_{i+1} x_{i+2} \cdots s_m x_m)$, and let

$$y = \bar{y}_1\bar{y}_2\bar{y}_3 = (y_0 s_1 \cdots y_i s_i) y_{i+1} (s_{i+1} y_{i+2} \cdots s_m y_m)$$

be similarly defined, where $x \underset{2}{\sim} y$, x and y are coterminal, and x and y have the same proper singularities. Then $\bar{x}_2$ and $\bar{y}_2$ are coterminal and

$$\bar{x}_1\tau = \bar{y}_1\tau \qquad \bar{x}_3\tau = \bar{y}_3\tau$$
$$(\bar{x}_1\bar{x}_2)\tau = (\bar{y}_1\bar{y}_2)\tau \qquad (\bar{x}_2\bar{x}_3)\tau = (\bar{y}_2\bar{y}_3)\tau$$

<u>Proof</u>: If x has no proper singularities then $\bar{x}_2 = x$ and $\bar{y}_2 = y$ and the claims easily follow. If x has exactly one singularity then either $\bar{x}_1 = 1$, $\bar{x}_2 = x_0$, $\bar{x}_3 = s_1x_1$ or $\bar{x}_1 = x_0s_1$, $\bar{x}_2 = x_1$, and $\bar{x}_3 = 1$. In the first case $\bar{y}_1 = 1$, $\bar{y}_2 = y_0$ and $\bar{y}_3 = s_1y_1$. Again the claim is easily verified here, and the second case is symmetric. The general case follows easily with the aid of Proposition 1(c).

<u>Proposition 6</u>. Let $x \in A^*$ have the factorization

$$x = x_1x_2x_3 = x_1x_{21}ax_{22}x_3 ,$$

where $x_2 = x_{21}ax_{22}$, $a \in A$, and $a \notin (x_1x_{21})\tau$. If x_2 has no singularities of x , then

$$(x_{21}a)\tau \subset (x_{22}x_3)\tau .$$

<u>Proof</u>: Since a appears in x_2 and x_2 has no singularities of x , we have $(a, a) \in x\tau_2$. Because $a \notin (x_1x_{21})\tau$, we must have $a \in (x_{22}x_3)\tau$. Also $e \in x_{21}\tau$ implies $(e, a) \in x_2\tau_2$. Since x_2 has no singularities of x , we have $(a, e) \in x\tau_2$ and $e \in (x_{22}x_3)\tau$. Thus $(x_{21}a)\tau \in (x_{22}x_3)\tau$.

<u>Proposition 7</u>. Let $x,y \in A^*$ have the factorizations

$$x = x_1x_2x_3 = x_1x_{21}ax_{22}x_3$$
$$y = y_1y_2y_3 = y_1y_{21}ay_{22}y_3$$

where x_2 and y_2 have no singularities of x, and $x_2 = x_{21}ax_{22}$, $y_2 = y_{21}ay_{22}$, $a \in A$, $a \notin (x_1x_{21})\tau \cup (y_1y_{21})\tau$. Then $(x_2x_3)\tau = (y_2y_3)\tau$ implies $(x_{22}x_3)\tau = (y_{22}y_3)\tau$.

<u>Proof</u>: $(x_{22}x_3)\tau = (x_{21}ax_{22}x_3)\tau = (x_2x_3)\tau$ by Proposition 6. Similarly $(y_{22}y_3)\tau = (y_2y_3)\tau$ and the claim follows. □

Let $x,y \in A^*$ be such that $x\tau = y\tau$ and let B be a given subset of $x\tau$. Let $\bar{x}$ and $\bar{y}$ be prefixes of x and y respectively. The pair $(\bar{x}, \bar{y})$ is called a <u>B-pair</u> iff

$$\overline{x}\tau = \overline{y}\tau \supset B .$$

Let $P_B(x, y)$ be the set of all B-pairs of x and y. This set is nonempty since $(x, y) \in P_B(x, y)$. Define the binary relation $\leq$ on $P_B(x, y)$ by

$$(x_1, y_1) \leq (x_2, y_2) \text{ iff } |x_1| \leq |x_2| \text{ and } |y_1| \leq |y_2| .$$

One verifies that $\leq$ is a partial order on $P_B(x, y)$.

<u>Proposition 8</u>. $P_B(x, y)$ has a unique minimal element with respect to $\leq$.

Proof: Because P is finite it suffices to show that for all $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ in $P_B(x, y)$ there exists $\bar{p} = (\bar{x}, \bar{y}) \in P_B(x, y)$ such that $\bar{p} \leq p_1$ and $\bar{p} \leq p_2$. If $p_1 \leq p_2$, let $\bar{p} = p_1$. If $p_2 \leq p_1$, let $\bar{p} = p_2$. Now suppose neither $p_1 \leq p_2$ nor $p_2 \leq p_1$. Suppose also that $|x_1| > |x_2|$. Then, since $p_1 \nleq p_2$, we must have $|y_1| < |y_2|$. Now

$$x_2\tau \subset x_1\tau = y_1\tau \subset y_2\tau = x_2\tau.$$

Let $\bar{p} = (x_2, y_1)$. Then $\bar{p}$ is a B-pair and $\bar{p} \leq p_1$, $\bar{p} \leq p_2$. Similarly, if $|x_1| < |x_2|$, then $|y_1| > |y_2|$. Let $\bar{p} = (x_1, y_2)$; then $\bar{p}$ is the required B-pair. Finally the case $|x_1| = |x_2|$ cannot occur, for then either $p_1 \leq p_2$ or $p_2 \leq p_1$. □

Lemma 2. Let x and y be coterminal paths such that $x \; {}_2\!\sim y$ and suppose that x and y have the factorizations:

$$x = x_1x_2x_3 \qquad\qquad y = y_1y_2y_3$$

where x_2 and y_2 are coterminal and do not contain any singularities of x and

$$x_1\tau = y_1\tau \qquad\qquad x_3\tau = y_3\tau$$
$$(x_1x_2)\tau = (y_1y_2)\tau \qquad (x_2x_3)\tau = (y_2y_3)\tau$$

Then $y \sim y_1x_2y_3$.

Proof: The proof proceeds by induction on $|x_2| + |y_2|$.

Basis: $|x_2| + |y_2| = 0$

Here $x_2 = y_2 = 1$ and $y = y_1 1 y_3 \sim y_1x_2y_3$.

Induction Step: $|x_2| + |y_2| > 0$

We assume that the lemma holds for all cases where $|x_2| + |y_2| \le k$. Suppose now that $|x_2| + |y_2| = k+1$. The proof will be decomposed into several cases.

Case 1: $x_2\tau \subset x_1\tau$ and $x_2\tau \subset x_3\tau$

Here $y_2\tau \subset (y_1y_2)\tau = (x_1x_2)\tau = x_1\tau = y_1\tau$. Similarly $y_2\tau \subset y_3\tau$. Also $x_2\tau \subset y_1\tau \cap y_3\tau$. By Rule 1(a)

$$y = y_1y_2y_3 \sim y_1x_2y_3.$$

Case 2: $x_2\tau \not\subset x_1\tau$

Note that $y_2\tau \not\subset y_1\tau$; otherwise $x_2\tau \subset (x_1x_2)\tau = (y_1y_2)\tau = y_1\tau = x_1\tau$, which is a contradiction. Let a be the first letter of x_2 from the left that does not appear in x_1. Similarly let b be the first letter of y_2 from the left that is not in y_1. Then $x_2 = x_{21}ax_{22}$, $y_2 = y_{21}by_{22}$ and

$$x = x_1x_{21}ax_{22}x_3\ , \quad \text{where} \quad a \notin (x_1x_{21})\tau = x_1\tau\ , \tag{5}$$

$$y = y_1y_{21}by_{22}y_3\ , \quad \text{where} \quad b \notin (y_1y_{21})\tau = y_1\tau\ . \tag{6}$$

We consider next two subcases.

Case 2.1: $a = b$

Here we have

$$y = y_1y_{21}ay_{22}y_3\ , \quad \text{where} \quad a \notin (y_1y_{21})\tau = y_1\tau\ , \tag{7}$$

and x is as in (5). Now x_{21} and y_{21} are coterminal and $x_{21}\tau, y_{21}\tau \subset y_1\tau$. By Proposition 6, $y_{21}\tau \subset (y_{22}y_3)\tau$. By Propositions 6 and 7, $x_{21}\tau \subset (x_{22}x_3)\tau = (y_{22}y_3)\tau$. By Rule (1a)

$$y = (y_1)(y_{21})(ay_{22}y_3) \sim (y_1)(x_{21})(ay_{22}y_3) = y'. \tag{8}$$

Now let $x_1' = x_1x_{21}a$, $x_2' = x_{22}$, and $x_3' = x_3$. Then

$$x = x_1'x_2'x_3' = (x_1x_{21}a)(x_{22})(x_3). \tag{9}$$

Similarly, let $y_1' = y_1x_{21}a$, $y_2' = y_{22}$, and $y_3' = y_3$. Then

$$y = y_1'y_2'y_3' = (y_1x_{21}a)(y_{22})(y_3). \tag{10}$$

We verify the 4 conditions of the lemma:

(i) $x_1'\tau = (x_1x_{21}a)\tau = (y_1x_{21}a)\tau = y_1'\tau$.

(ii) $(x_1'x_2')\tau = (x_1x_2)\tau = (y_1y_2)\tau = (y_1'y_2')\tau$.

(iii) $x_3'\tau = x_3\tau = y_3\tau = y_3'\tau$.

(iv) $(x_2'x_3')\tau = (x_{22}x_3)\tau = (y_{22}y_3)\tau = (y_2'y_3')\tau$ by Proposition 7.

Note that x_2' is a proper factor of x_2 and y_2' is a proper factor of y_2. Hence x_2' and y_2' do not contain any singularities of x. Evidently $|x_2'| + |y_2'| < |x_2| + |y_2|$ and we can apply the induction hypothesis:

$$y' = y_1'y_2'y_3' \sim y_1'x_2'y_3' = y_1x_{21}ax_{22}y_3 = y_1x_2y_3.$$

Altogether $y \sim y' \sim y_1x_2y_3$ and the induction step goes through in this case.

Case 2.2: $a \neq b$

Refer to (5) and (6). Since $b \in (y_1y_2)\tau - y_1\tau = (x_1x_2)\tau - x_1\tau$ we must have $b \in x_{22}\tau$. Similarly $a \in y_{22}\tau$ and

$$x = x_1x_2x_3 = x_1(x_{21}ax_{22})x_3 = x_1x_{21}a(s_1bs_2)x_3, \tag{11}$$

where $x_{22} = s_1bs_2$ and $b \notin (x_1x_{21}as_1)\tau$, and

$$y = y_1y_2y_3 = y_1(y_{21}by_{22})y_3 = y_1y_{21}b(t_1at_2)y_3, \tag{12}$$

where $y_{22} = t_1at_2$ and $a \notin (y_1y_{21}bt_1)\tau$. In other words the leftmost appearances of b in x and a in y are shown.

Let $(as_1)\tau \cup (bt_1)\tau = B$. The prefixes x_1x_2 of x_1x_2 and y_1y_2 of y_1y_2 satisfy

$$(x_1x_2)\tau = (y_1y_2)\tau \supset B.$$

Thus (x_1x_2, y_1y_2) is a B-pair. By Proposition 8, there exists a minimal B-pair $(\bar{x}, \bar{y})$. Since $b \in B$ and $b \notin (x_1x_{21}as_1)\tau$ we have

$$|x_1x_{21}as_1b| \le |\bar{x}| \le |x_1x_2|. \tag{13}$$

Similarly

$$|y_1y_{21}bt_1a| \le |\bar{y}| \le |y_1y_2| \tag{14}$$

Let c be the last letter of $\bar{x}$ and d the last letter of $\bar{y}$, and let $\bar{x} = pc$ and $\bar{y} = qd$. We claim first that $c \neq d$. Note that $c \notin p\tau$, for otherwise the pair $(p, \bar{y})$ would be a shorter B-pair. Similarly $d \notin q\tau$. Assume now that $c = d$. If $c \notin B$, then (p, q) is a B-pair, contradicting the assumption that (pc, qc) is minimal. Thus $c \in B = (as_1bt_1)\tau$. Since $|x_1x_{21}as_1b| \le |pc|$ and $c \notin p\tau$, the condition $c \in (as_1b)\tau$ implies $c = b$. But then $c \in (y_1y_{21}bt_1)\tau$ and $y_1y_{21}bt_1$ is a proper prefix of $\bar{y}$. This implies $c \in q\tau$ which is a contradiction. Hence we cannot have $c \in (as_1b)\tau$ and we must have $c \in t_1\tau$. This is again a contradiction of the fact that $c \notin q\tau$. Therefore $c \neq d$.

From (13) and (11) it is clear that either $c = b$ or $c \neq b$ and $c \in s_2$. Both cases can be handled by the notation

$$pc = x_1 x_{21} a s_1 b s_{21}. \tag{15}$$

For if $c = b$, let $s_{21} = 1$. Otherwise let s_{21} be the shortest prefix of s_2 that ends in c. In either case let $s_2 = s_{21}s_{22}$. Similarly

$$qd = y_1 y_{21} b t_1 a t_{21} \tag{16}$$

where $t_2 = t_{21}t_{22}$ and $t_{21} = 1$ if $d = a$, and t_{21} is the shortest prefix of t_2 that ends in d, otherwise. Now let

$$f = a s_1 b s_{21},$$
$$g = b t_1 a t_{21}.$$

We now arrive at the decompositions of x and y :

$$\begin{aligned} x &= x_1x_2x_3 = x_1x_{21}ax_{22}x_3 = x_1x_{21}as_1bs_2x_3 \\ &= x_1x_{21}as_1bs_{21}s_{22}x_3 = x_1x_{21}fs_{22}x_3 = pcs_{22}x_3, \end{aligned} \tag{17}$$

$$\begin{aligned} y &= y_1y_2y_3 = y_1y_{21}by_{22}y_3 = y_1y_{21}bt_1at_2y_3 \\ &= y_1y_{21}bt_1at_{21}t_{22}y_3 = y_1y_{21}gt_{22}y_3 = qdt_{22}y_3. \end{aligned} \tag{18}$$

Consider next where c can appear in y. Since $c \in (pc)\tau = (qd)\tau$, we must have $c \in (y_1y_{21}bt_1at_{21})\tau$. If $c \in (y_1y_{21})\tau$ then $c \in x_1\tau$ and $c \in p\tau$ which is a contradiction. Hence $c \in (bt_1at_{21})\tau = g\tau$. Similarly $d \in (as_1bs_{21})\tau = f\tau$. Let

$$f = as_1bs_{21} = u_1du_2c, \quad \text{where} \quad d \notin u_2\tau, \tag{19}$$

$$g = bt_1at_{21} = v_1cv_2d, \quad \text{where} \quad c \notin v_2\tau. \tag{20}$$

In other words we take the rightmost appearances of d in f and c in g. We now have the factorizations illustrated in Figure 2. Of necessity, the figure shows a particular case and should only be used as a visual aid.

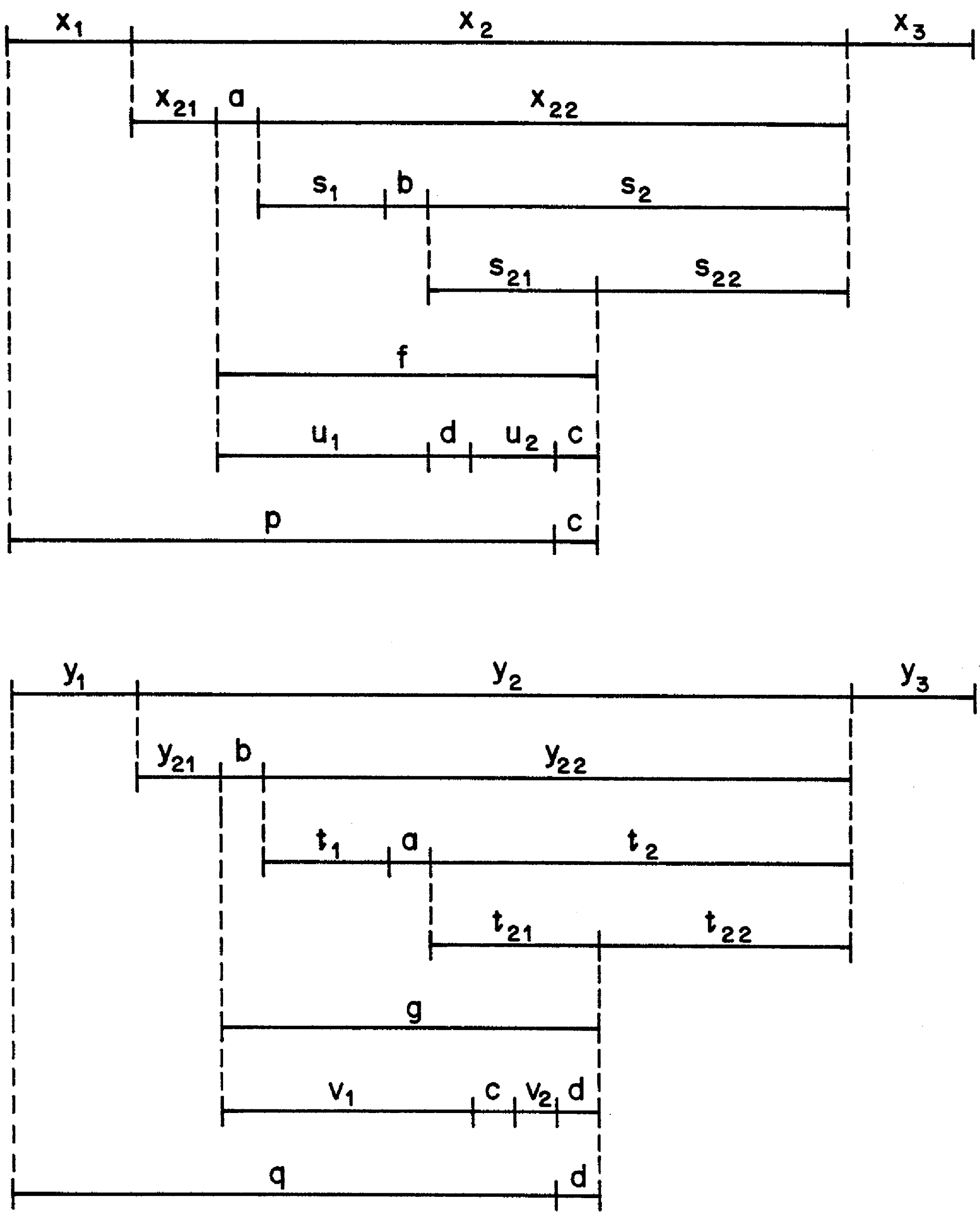

Figure 2 Illustrating Factorizations of x and y

We will deal with the factorization

$$x = x_1'x_2'x_3' = (x_1x_{21}f)(s_{22})(x_3), \qquad (21)$$

where $x_1' = x_1x_{21}f$, $x_2' = s_{22}$, $x_3' = x_3$. We begin with

$$y = y_1y_{21}gt_{22}y_3$$

and we will show that $y \sim y'$ where

$$y' = y_1'y_2'y_3' = (y_1x_{21}f)(v_2dt_{22})(y_3) \qquad (22)$$

where $y_1' = y_1x_{21}f$, $y_2' = v_2dt_{22}$, and $y_3' = y_3$. The proof is given in Lemma 3 below. Assuming this result we next show that all the conditions of Lemma 2 apply to (21) and (22).

First, $x_2' = s_{22}$ is a proper factor of x_2 and $y_2' = v_2dt_{22}$ is a proper factor of y_2. Hence x_2' and y_2' contain no singularities of x. Second, x_2' and y_2' are coterminal. Third, $y \sim y'$ (Lemma 3) implies $y \;{}_2\!\sim y'$ and so $x \;{}_2\!\sim y'$. Finally, we verify the four conditions on the alphabets of the factors:

(i) $x_1'\tau = (x_1x_{21}f)\tau = (y_1x_{21}f)\tau = y_1'\tau$.

(ii) $(x_1'x_2')\tau = (x_1x_2)\tau = (y_1y_2)\tau = (qd)\tau \cup t_{22}\tau$
$= (pc)\tau \cup t_{22}\tau = (x_1x_{21}f)\tau \cup t_{22}\tau$
$= (y_1x_{21}f)\tau \cup t_{22}\tau = (y_1x_{21}f)\tau \cup (v_2d)\tau \cup t_{22}\tau$

because $(v_2d)\tau \subset (y_1y_2)\tau$. Therefore
$(x_1'x_2')\tau = (y_1x_{21}fv_2dt_{22})\tau = (y_1'y_2')\tau$.

(iii) $x_3'\tau = x_3\tau = y_3\tau = y_3'\tau$.

(iv) Since y_1' ends in f which ends in c, $e \in (y_2'y_3')\tau$ implies $(c, e) \in y'\tau_2 = x\tau_2$. Hence $e \in (s_{22}x_3)\tau$, because $c \notin p\tau$. Therefore $(y_2'y_3')\tau \subset (x_2'x_3')\tau$.

Conversely, $(x_2'x_3')\tau = (s_{22}x_3)\tau \subset (x_2x_3)\tau = (y_2y_3)\tau$ $= (y_{21}gt_{22}y_3)\tau$. By Proposition 6 applied to the letter d in g, $(y_{21}g)\tau \subset (t_{22}y_3)\tau$. Hence $(x_2'x_3')\tau \subset (t_{22}y_3)\tau \subset (y_2'y_3')\tau$. Thus $(x_2'x_3')\tau = (y_2'y_3')\tau$.

Now all the conditions of Lemma 2 are satisfied. Since $|x_2'| + |y_2'| < |x_2| + |y_2|$, the induction hypothesis applies and

$$y' = y_1'y_2'y_3' \sim y_1'x_2'y_3' = y_1x_{21}fs_{22}y_3 = y_1x_2y_3.$$

Therefore $y \sim y' \sim y_1x_2y_3$ as claimed, and the induction step goes through.

<u>Case 3: $x_2\tau \not\subset x_3\tau$</u>

This follows from Case 2 by left-right duality.

Since the induction step goes through in all cases, the lemma holds.

<u>Lemma 3</u>. Let x, y, and y' be defined as in the proof of Lemma 2. Then $y \sim y'$.

<u>Proof</u>:

(a) We first show that the graph consisting of the edges in $C = f\tau \cup g\tau$. is strongly connected. Since the node $b\omega$ is connected to $a\alpha = f\alpha$ by the path t_1, all the nodes in the path as_1b are connected to $f\alpha$. Let $s_{21} = s_{21}'s_{21}''$ where s_{21}' is the longest prefix of s_{21} that is connected to $f\alpha$. Similarly, $a\omega$ is connected to $b\alpha = g\alpha$ by s_1. Let $t_{21} = t_{21}'t_{21}''$ where t_{21}' is the longest prefix of t_{21} connected to $g\alpha$. See Figure 3.

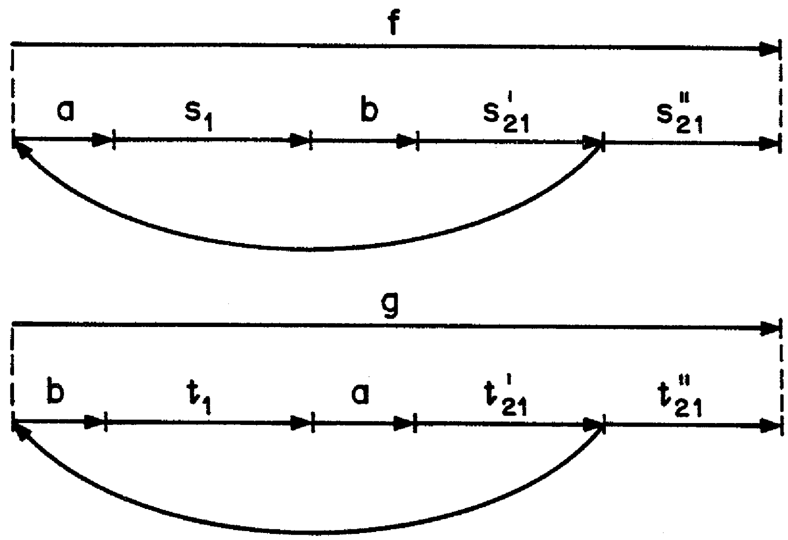

Figure 3

Now s''_{21} cannot have any edges in common with $as_1bs'_{21}$ or $bt_1at'_{21}$. Otherwise the ω end of the common edge could be connected to $f\alpha$. Hence

$$s''_{21}\tau \cap (bt_1at'_{21})\tau = \phi.$$

Also, $(pc)\tau \supset (qd)\tau$, i.e.

$$(x_1x_{21}as_1bs'_{21})\tau \cup s''_{21}\tau \supset (y_1y_{21}bt_1at'_{21})\tau = (y_1y_{21})\tau \cup (bt_1at'_{21})\tau.$$

Consequently we have:

$$(x_1x_{21}as_1bs'_{21})\tau \supset (y_1y_{21}bt_1at'_{21})\tau.$$

Similarly the reverse inclusion holds and

$$(x_1x_{21}as_1bs'_{21})\tau = (y_1y_{21}bt_1at'_{21})\tau \supset B = (as_1)\tau \cup (bt_1)\tau.$$

Therefore $(x_1x_{21}as_1bs'_{21}, y_1y_{21}bt_1at'_{21})$ is a B-pair. However (pc, qd) is a minimal B-pair. Hence we must have $s'_{21} = s_{21}$, $t'_{21} = t_{21}$, $f\omega$ is connected to $f\alpha$ and $g\omega$ is connected to $g\alpha$. Hence the graph is strongly connected since f and g have a common edge.

(b) In view of (a) there exist paths h and k such that

$$h\alpha = f\omega, \quad h\omega = f\alpha, \quad h\tau \subset C$$
$$k\alpha = g\omega, \quad k\omega = g\alpha, \quad k\tau \subset C.$$

Let $f' = u_2ch$ and $g' = v_2dk$. Then $f'fg'g$ is a loop about the vertex $d\omega = g\omega$ and $f'fg'g \subset C$. Now

$$\begin{aligned} y &= y_1y_{21}gt_{22}y_3 \\ &\sim (y_1y_{21}g)(f'fg'g)^3t_{22}y_3 \end{aligned}$$

by (1a), because $(y_1y_{21}g)\tau = (qd)\tau = (tc)\tau \supset C$, and $C \subset (t_{22}y_3)\tau$ by Proposition 6. Thus

$$\begin{aligned} y &\sim y_1y_{21}g(f'fg'g)^3t_{22}y_3 \\ &= y_1y_{21}(gf')((fg')(gf'))^2fg'gt_{22}y_3 \\ &= [y_1][y_{21}(gf')][(fg')(gf')]^2[fg'gt_{22}y_3] \end{aligned}$$

Now Rule (1c) can be applied, yielding

$$y \sim y_1x_{21}(fg'gf')^2fg'gt_{22}y_3$$

where we have replaced $y_{21}gf'$ by x_{21}. The alphabet conditions on x_{21} and y_{21} we easily verified. Thus

$$\begin{aligned} y &\sim y_1x_{21}(fg'gf')^2fg'gt_{22}y_3 \\ &= y_1x_{21}fg'(gf'fg')^2gt_{22}y_3 \\ &\sim y_1x_{21}fg'gt_{22}y_3\,, \quad \text{by Rule (1a)} \\ &= y_1x_{21}fv_2d(kv_1cv_2d)t_{22}y_3 \\ &\sim y_1x_{21}fv_2dt_{22}y_3\,, \quad \text{by Rule (1a)} \\ &= y'. \end{aligned}$$

Hence the lemma holds.

This concludes the proof of Lemmas 2 and 3. By combining Lemmas 1 and 2 we have the theorem.

References

[1] Brzozowski, J.A., Hierarchies of Aperiodic Languages, *Revue Francaise d'Automatique, Informatique et Recherche Operationnelle*, Vol. 10, No. 8, pp. 33-49, August 1976.

[2] Brzozowski, J.A., and Simon, I., Characterizations of Locally Testable Events, *Discrete Mathematics*, Vol. 4, No. 3, pp. 243-271, March 1973.

[3] Eilenberg, S., *Automata, Languages and Machines*, Vol. B, Academic Press, New York, 1976.

[4] Knast, R., Bilocally Testable Languages, Technical Report, Institute of mathematics, Polish Academy of Sciences, Poznan, Poland, 1979.