A. Ehrenfeucht

P. Ten Pas

G. Rozenberg

Combinatorial properties of texts

<http://www.numdam.org/item?id=ITA_1993__27_5_433_0>

COMBINATORIAL PROPERTIES OF TEXTS (*)

by A. Ehrenfeucht ([1]), P. Ten Pas ([2]) and G. Rozenberg ([1]), ([2])

Communicated by J. Berstel

Abstract. – *A text is considered as a sorting process. Combinatorial properties of such sorting processes are considered. In particular combinatorial characterizations of* alternating texts *are given. Also a natural surperclass of alternating texts is introduced (the so-called* jump-free texts) *and it is characterized through combinatorial properties and through closure properties.*

Résumé. – *Un texte est considéré comme un procédé de loi. Les propriétés combinatoires de tels procédés de loi sont étudiés. On donne en particulier des caractérisations combinatoires de textes alternants. On introduit également une sur-classe naturelle des textes alternants (appelés les textes sans saut) et on les caractérise tant par des propriétés combinatoires que par des propriétés de clôture.*

1. INTRODUCTION

A *word* can be considered as an ordered pair (λ, ρ) where ρ is a linear order on a (finite) domain D and λ is a function on D [Mostly one assumes that $\rho = (1, 2, \ldots, n)$ for some $n \geq 0$, we refer to such a word as *standard*].

The notion of a text generalizes the notion of a word in that a text is a triple $\tau = (\lambda, \rho_1, \rho_2)$ where ρ_1, ρ_2 are linear orders on the same domain D and λ is a function on D [if (λ, ρ_1) is standard, then τ is called *standard*]. The importance of this generalization is that one can see the text τ as the word (λ, ρ_1) together with the (syntactic) *structure* spanned on it. This structure is determined by the second linear order ρ_2 – it may be a tree as in classical language theory, but it also may be more general than a tree. The

way that ρ_2 determines this structure is established through the theory of 2-*structures* as follows.

Given two linear orders $\{\rho_1, \rho_2\}$ on a common domain D, $P = \{\rho_1 \cap \rho_2,$ $\rho_1 \cap \text{rev}(\rho_2), \text{rev}(\rho_1) \cap \rho_2, \text{rev}(\rho_1) \cap \text{rev}(\rho_2)\} - \{\varnothing\}$ is a partition of the set of 2-edges over D [here $\text{rev}(\rho)$ denotes the reverse of a linear order ρ, and a 2-edge over D is an ordered pair of different elements of D]. Then (D, P) is a 2-structure, as a matter of fact a 2-structure with specific properties which is referred to as a *T-structure* (*see* e. g., [2]).

A 2-*structure* is an ordered pair (D, P), where D is a finite domain and P is a partition of the set $E_2(D)$ of the 2-edges over D; a *labeled* 2-structure has additionally a labeling function which assigns a unique label (syntactic name) to each class of P. The main result of the decomposition theory for labeled 2-structures says that each labeled 2-structure g has a unique hierarchical representation, called the *shape* of g, which is a tree where each node is a labeled 2-structure of a special type. In the case of labeled *T*-structures each node of the shape is either a *linear* labeled 2-structure or a *primitive* labeled 2-structure. Linear labeled 2-structures correspond to linear orders and so if the shape has linear nodes only, then it is essentially a tree; otherwise (*i. e.*, if it contains also primitive nodes) it is more general than a tree.

Hence given a text $\tau = (\lambda, \rho_1, \rho_2)$ one gets a hierarchical representation of τ by considering the shape of the labeled T-structure associated with τ. In this way ρ_2 determines the syntactic structure of the word (λ, ρ_1) which may be a tree or a structure more general than a tree. If the corresponding shape has linear nodes only (*i. e.*, the shape is *locally linear*), then it is a tree; otherwise it is more general than a tree because it can contain also primitive nodes.

The notion of a text was introduced in [3] where also basic properties of texts were investigated; more properties of texts are investigated in [4].

In this paper we investigate combinatorial properties of texts. The point of view taken in this paper is that in a (standard) text $\tau = (\lambda, \rho_1, \rho_2)$, ρ_2 is obtained by permuting ρ_1. We consider combinatorial properties of such permutations.

If we go through elements of the domain of τ according to the linear order ρ_1 and put them one-by-one in the positions where they belong in ρ_2, then in fact we are sorting ρ_1 according to ρ_2; when τ is standard we are simply sorting $\{1, \ldots, n\}$ (the domain of τ) according to ρ_2. The way that ρ_2 is formed when we go through this sorting process is described formally through the notion of a *sorting sequence*.

This paper investigates sorting sequences of texts. They are introduced in Section 2, where also some of their basic properties are proved.

In Section 3 we investigate sorting sequences of alternating texts (hence one may say that we investigate how to form trees through sortings!). We also provide characterizations of alternating texts. The results of this section lead to the introduction of a class of texts more general than alternating texts; it is the class of *jump-free texts*.

The class of jump-free texts is investigated in Section 4 where it is given a combinatorial and an operational characterization.

The basic notions concerning texts and labeled 2-structures are recalled in Section 1.

We would like to point out that in the considerations of this paper (concerning the combinatorial properties of texts) the labeling function fun_τ of a text τ does not play any technical role. Hence in fact we deal in this paper with *naked texts* which are ordered pairs of linear orders (representing finite permutations). However, for the sake of consistency with [3] we have decided to keep here the notion of a text in its original form.

2. PRELIMINARIES

In this section we give some notation and terminology, in particular concerning graphs and trees.

For a set Z, $\#Z$ denotes its cardinality, and $E_2(Z) = \{(x, y): x, y \in Z$ and $x \neq y\}$; each element of $E_2(Z)$ is a 2-*edge over* Z. If $e = (x, y)$ is a 2-edge, then the *reverse of* e, denoted $rev(e)$, is the 2-edge (y, x). For a set of 2-edges $T \subseteq E_2(D)$, the *reverse* of T, denoted $rev(T)$, is the set $\{rev(e) | e \in T\}$. $SING(Z)$ denotes the set of all singletons over Z; $\emptyset$ denotes the empty set. *Unless explicitly clear otherwise, we consider finite sets only.*

Sets X, Y are *overlapping* iff $X - Y \neq \emptyset$, $Y - X \neq \emptyset$, and $X \cap Y \neq \emptyset$. For sets X, Y we write $X \subseteq Y$ if X is included in Y, $X \subset Y$ if X is *strictly* included in Y, and $X \times Y$ denotes the Cartesian product of X, Y. In a partition of a set we assume that each partition class is nonempty.

For a sequence s, $|s|$ denotes its length, and for $1 \leq i \leq |s|$, $s(i)$ denotes the i'th element of s.

By a *function* in this paper we understand a set of ordered pairs φ such that, for all (x, y), $(u, v) \in \varphi$, $x = u$ implies $y = v$. The set $\{x:$ there exists y such that $(x, y) \in \varphi\}$ is the *domain of* φ, denoted by $dom(\varphi)$; we say that φ

is a *function on* dom (φ). If $Z \subseteq \text{dom} (\varphi)$, then $\varphi|_Z$ denotes the restriction of φ to Z.

A (directed) *graph* is an ordered pair $h = (D, T)$, where D is a (finite) nonempty set of *nodes*, denoted by $\text{nd}(h)$, and $T \subseteq D \times D$ is the set of *edges*. h is *antireflexive* iff, for each $x \in D$, $(x, x) \notin T$; h is *transitive* iff, for all x, y, $z \in D$, $(x, y) \in T$ and $(y, z) \in T$ implies $(y, z) \in T$. h is a *linear order iff* h is antireflexive, transitive, and for each $(x, y) \in E_2(D)$, either $(x, y) \in T$ or $(y, x) \in T$.

We carry over to T the terminology and notations concerning h. Hence, T is a *linear order* iff h is a linear order. If $h = (D, T)$ is a linear order such that $D = \{ x_1, \ldots x_n \}$ for some $n \geq 1$, and $(x_j, x_{j+1}) \in T$ for all $j \in \{ 1, \ldots, n-1 \}$, then we write T in the form $(x_1, \ldots, x_n)$. In particular, for a linear order with one element x we write (x). Hence a linear order (D, T) can be specified as a sequence of the elements of D. The terminology and notations concerning sequences carry over to linear orders.

If $h = (D, T)$ is a linear order and $X \subseteq D$, then X is a *segment of* h (or of T) iff for all y, $z \in X$, if $u \in D$ is such that $(y, u) \in T$ and $(u, z) \in T$, then $u \in X$. We use $\text{seg}(h)$ (or $\text{seg}(T)$) to denote all segments of h. For $X \subseteq D$, $h|_x$ denotes the restriction of h to X, i.e., $(X, T \cap E_2(X))$, and $T|_x$ denotes the restriction of T to X, i.e., $T \cap E_2(X)$.

If $h_1 = (D_1, T_1)$, $h_2 = (D_2, T_2)$ are disjoint linear orders (*i.e.*, $D_1 \cap D_2 = \varnothing$), then the *sum of* h_1, h_2, is the linear order $h = (D_1 \cup D_2, T)$, such that $T = T_1 \cup T_2 \cup \{ (x, y) : x \in D_1 \text{ and } y \in D_2 \}$.

Two graphs $h_1 = (D_1, T_1)$, $h_2 = (D_2, T_2)$ are *isomorphic* iff there is a bijection $\varphi : D_1 \to D_2$ such that, for all x, $y \in D_1$, $(x, y) \in T_1$ iff $(\varphi(x), \varphi(y)) \in T_2$; φ is an *isomorphism* between h_1 and h_2.

A graph $t = (D, T)$ is a *tree* iff t is acyclic and there exists a node v of t (the *root* of t, denoted $root(t)$) such that each node of t is reachable from v by a unique path. We use $leaf(t)$ to denote the set of *leaves* of t, and $in(t)$ to denote the set of *inner nodes* of t (*i.e.*, in $(t) = \text{nd}(t) - leaf(t)$). For a node $v \in in(t)$, $ddes_t(v)$ denotes the set of *direct descendants of* v (in t), *i.e.*, nodes x such that $(v, x) \in T$.

We assume that a tree does not have chains, i.e., each inner node of a tree has at least two direct descendants. Also, unless explicitly clear otherwise, we assume that a tree has more than one (hence at least three) nodes; in this way we avoid unnecessary trivial technicalities.

3. TEXTS

In this section we give an overview of basic notions concerning texts. The study of texts has originated in the theory of 2-structures, where a 2-structure is a restricted kind of relational structure.

DEFINITION 3.1: A *labeled 2-structure* (abbreviated *l2s*) is a 3-tuple $g = (D, \Delta, \delta)$, where D is a finite nonempty set, Δ is a finite alphabet, and δ is a function from $E_2(D)$ into Δ. □

Let $g = (D, \Delta, \delta)$ be a *l2s*. D is called the *domain of g*, denoted by dom (g), and δ is called the *labeling function of g*, denoted by lab (g).

The labeling function δ induces an equivalence relation on the set of 2-edges of g as follows: for all e_1, $e_2 \in E_2(D)$, e_1 is *g-equivalent with* e_2 iff $\delta(e_1) = \delta(e_2)$. This equivalence relation on $E_2(D)$, denoted by rel(g), corresponds to a partition of $E_2(D)$ into equivalence classes, which is denoted by part (g).

If we "forget" the labeling of the equivalence classes, then we are in fact considering "(unlabeled) 2-structures": a *2-structure* is a pair (D, α), where D is a finite nonempty set, and α is a partition of $E_2(D)$ [e.g., given by an equivalence relation on $E_2(D)$].

We assume that (labeled) 2-structures are "reversible", *i.e.*, if $g = (D, \Delta, \delta)$ is a *l2s*, and e_1, $e_2 \in E_2(D)$, then $\delta(e_1) = \delta(e_2)$ iff $\delta(\text{rev}(e_1)) = \delta(\text{rev}(e_2))$.

Let $g = (D, \Delta, \delta)$ be a *l2s*. A 2-edge e of g is *symmetric* iff $\delta(e) = \delta(\text{rev}(e))$; otherwise e is called *asymmetric*. For each class $P \in \text{part}(g)$, rev $(P) \in \text{part}(g)$, and either $P = \text{rev}(P)$ (hence P consists of symmetric 2-edges only and is called *symmetric*), or $P \cap \text{rev}(P) = \varnothing$ (hence P consists of asymmetric 2-edges only and is called *antisymmetric*). Consequently, the classes in part(g) can be grouped into sets $\{P, \text{rev}(P)\}$ – such a set is called a *feature of g* (if P is symmetric, then the corresponding feature has one class only; if P is antisymmetric the feature consists of two classes). g is *symmetric* iff all its 2-edges are symmetric and g is *antisymmetric* iff all its 2-edges are asymmetric.

Example 3.1: Consider the *l2s' s* g_1 and g_2 in figure 1 (we use the obvious "graph-theoretic" pictorial representation to specify labeled 2-structures).

g_1 is symmetric and has two features. g_2 has one symmetric class $\{(1, 2), (2, 1), (2, 4), (4, 2), (3, 4), (4, 3)\}$ in part(g_2) and two antisymmetric classes, $\{(1, 3), (1, 4), (2, 3)\}$ and its reverse $\{(3, 1), (4, 1), (3, 2)\}$. Hence g_2 has two features. □

For a *l2s* $g = (D, \Delta, \delta)$ and $X \subseteq D$, the *substructure of g determined by X*, denoted $\text{sub}_g(X)$, is the *l2s* $(X, \Delta, \delta|_X)$.
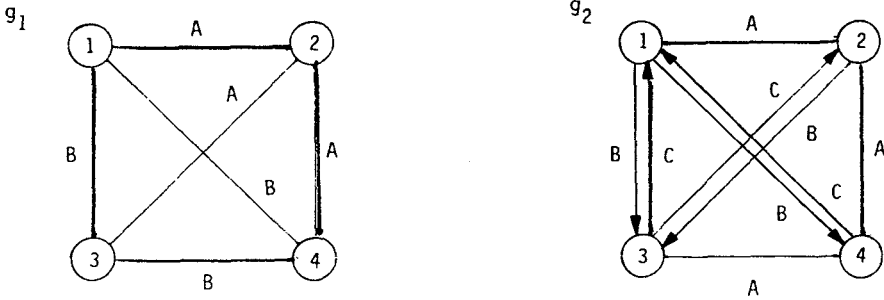
Figure 1.

The basic technical notion concerning labeled 2-structures is the notion of a "clan".

DEFINITION 3.2: Let $g = (D, \Delta, \delta)$ be a $l2s$, and let $X \subseteq D$. X is a *clan* (*of g*) iff for all $x, y \in X$ and all $z \in D - X$, $\delta(z, x) = \delta(z, y)$. □

We use $C(g)$ to denote the set of all clans of g.

Clearly, for each $l2s$ $g = (D, \Delta, \delta)$, $\varnothing \in C(g)$, SING $(D) \subseteq C(g)$, and $D \in C(g)$. These clans are called the *trivial clans of g*, denoted $TC(g)$. A *prime* clan of g is a clan that is not overlapping any other clan of g. The set of all prime clans of g is denoted $PC(g)$. Note that $TC(g) \subseteq PC(g)$.

An important property of clans is the following one: if X and Y are disjoint clans of a $l2s$ g, then for all $x_1, x_2 \in X$ and all $y_1, y_2 \in Y$, (x_1, y_1) is equivalent with (x_2, y_2). This property allows one to form *quotients* as follows. If $g = (D, \Delta, \delta)$ is a $l2s$, and M is a partition of D into clans of g, then g/M is the $l2s$ (M, Δ, δ'), where δ' is such that for all $X, Y \in M$, $\delta'(X, Y) = \delta(x, y)$, where $x \in X$ and $y \in Y$.

The following subclasses of the class of labeled 2-structures are both natural and important.

DEFINITION 3.3: Let $g = (D, \Delta, \delta)$ be a $l2s$.

(1) g is *primitive* iff $C(g) = TC(g)$.

(2) g is *special* iff $PC(g) = TC(g)$.

(3) g is *complete* iff eitther $|D| = 1$ or $|\text{part}(g)| = 1$.

(4) g is *linear* iff either $|D| = 1$ or g is antisymmetric, $\text{part}(g) = \{P, \text{rev}(P)\}$ and there exists a linear order $(x_1, \ldots, x_n)$, $n \geq 2$, of D such that, for all different $i, j \in \{1, \ldots, n\}$, $(x_i, x_j) \in P$ iff $i < j$. □

Example 3.2: Let g and h be as in figure 2.

$C(g) = \emptyset \cup \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2, 3, 4\}\}$, hence g is a primitive $l2s$. h is a linear $l2s$ (consider the linear order $(1, 4, 3, 2)$ of its domain). $\square$
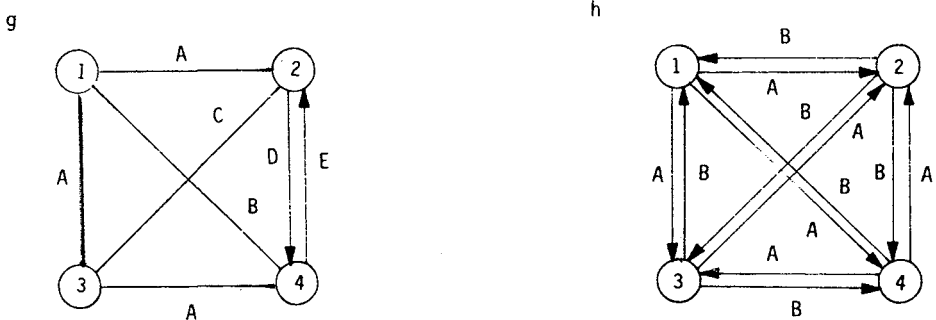

Figure 2.

A basic technical result for labeled 2-structures is that a $l2s$ g is special iff g is complete, linear or primitive. The decomposition theory developed in [1] allows one to define "hierarchical representations" of labeled 2-structures using one of these three types of labeled 2-structures only. These representations are formulated in terms of "$l2s$-labeled tree families".

A tree family is an ordered pair (D, T), where D is a finite nonempty set, and T is a subset of 2^D such that $D \in T$, $\emptyset \notin T$, $\mathrm{SING}(D) \subseteq T$, and for all X, $Y \in T$, X and Y are not overlapping. To each tree family there corresponds a unique tree the nodes of which are the elements of T and the edges are of the form (X, Y) with $X, Y \in T$ where $X \subset Y$ and for no $Z \in T$, $X \subset Z \subset Y$. Due to this correspondence we carry over the terminology and notations for trees to tree families.

DEFINITION 3.4: A $l2s$-*labeled tree family* is a triple $\beta = (D, T, \varphi)$ such that $\alpha = (D, T)$ is a tree family, and φ is a function on $in(\alpha)$ such that, for each $X \in in(\alpha)$, $\varphi(X)$ is a $l2s$ with $\mathrm{dom}(\varphi(X)) = ddes_\alpha(X)$. $\square$

A $l2s$ $g = (D, \Delta, \delta)$ is hierarchically represented by a $l2s$-labeled tree family $\beta = (D, T, \varphi)$ iff $T \subseteq C(g)$, and, for each $X \in in(\beta)$, $\varphi(X) = sub_g(X)/ddes_\beta(X)$. If T is $PC(g) - \{\emptyset\}$, then β is called the *shape* of g, denoted *shape* (g). The main theorem of the decomposition theory for labeled 2-structures says that if $\beta = shape(g)$, then $\varphi(X)$ is special for each $X \in in(\beta)$, and if $shape(g) = shape(g')$ then $g = g'$.

Example 3.3: Let g be the *12s* in figure 3.

Then *shape*(g) is as in figure 4.   □

Given a *12s*-labeled tree framily $\beta=(D, T, \varphi)$ representing g, one can recover g, *e.g.*, in a bottom-up fashion, as follows. For $x, y \in D$, consider $Z \in \text{in}(\beta)$ such that Z has two direct descendents $X, Y \in T$ with $x \in X$ and
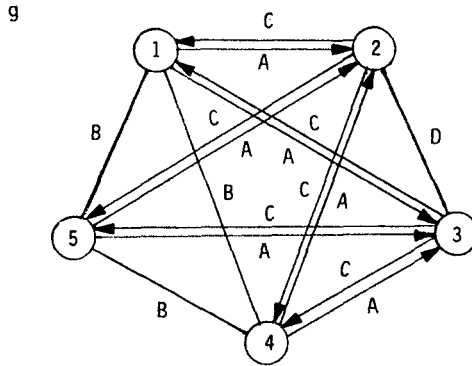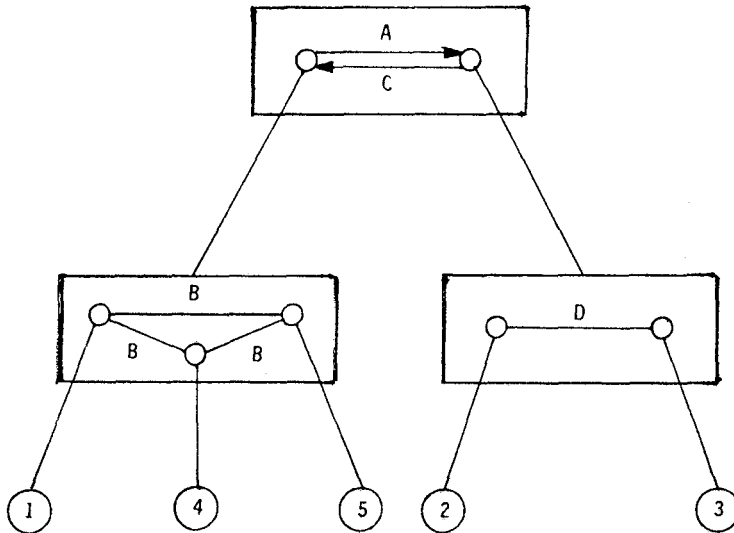
g

Figure 3.

shape(g)

Figure 4.

$y \in Y$. Then the label of the 2-edge (x, y) of g is the label of the 2-edge (X, Y) in the $l2s$ that labels Z in β [hence in $\varphi(Z)$].

Consider two linear orders ρ_1, ρ_2 on a finite nonempty set D. It turns out that $(D, \{\rho_1 \cap \rho_2, \rho_1 \cap \operatorname{rev}(\rho_2), \operatorname{rev}(\rho_1) \cap \rho_2, \operatorname{rev}(\rho_1) \cap \operatorname{rev}(\rho_2)\} - \{\varnothing\})$ is a 2-structure. As a matter of fact it is a $2s$ with very specific properties, referred to as a "T-structure".

DEFINITION 3.5: A T-*structure* (abbreviated Ts) is a $2s$ $g = (D, R)$ such that g is antisymmetric, and for all $X \subseteq D$ with $|X| = 3$, there is a $x \in X$ such that (x, y) is equivalent with (x, z), where $X = \{x, y, z\}$.   □

If we add a labeling function to a Ts, then it becomes a *labeled T-structure* (abbreviated lTs).

It can also be proved that, given a Ts with at most two features, there are two linear orders ρ_1, ρ_2 on its domain such that part$(g) = \{\rho_1 \cap \rho_2, \rho_1 \cap \operatorname{rev}(\rho_2), \operatorname{rev}(\rho_1) \cap \rho_2, \operatorname{rev}(\rho_1) \cap \operatorname{rev}(\rho_2)\} - \{\varnothing\}$. Moreover these two linear orders are unique in the sense that if ρ'_1, ρ'_2 are two other linear orders determining part (g) as above, then

$$\{\rho_1, \rho_2, \operatorname{rev}(\rho_1), \operatorname{rev}(\rho_2)\} = \{\rho'_1, \rho'_2, \operatorname{rev}(\rho'_1\}, \operatorname{rev}(\rho'_2)\}.$$

As a matter of fact two linear orders can represent (as above) an arbitrary T-structure, because it can be proved that an arbitrary Ts is equivalent in a well-defined sense to a Ts with at most two features.

The above results allow one to use a pair of linear orders as a specification of a Ts. On the other hand, through this relationship, hierarchical representations can be assigned to pairs of linear orders, just by taking the hierarchical representations of the corresponding T-structures. Note that, because a Ts is antisymmetric, if a $l2s$-labeled tree family β is a shape of a Ts, then nodes of β are either linear or primitive.

This leads to the notion of a text.

DEFINITION 3.6: A *text* is a 3-tuple $\tau = (\lambda, \rho_1, \rho_2)$, where λ is a finite function, and ρ_1 and ρ_2 are linear orders on dom(λ).   □

We will use fun_τ, dom(τ), $VO(\tau)$, and $HO(\tau)$ to denote λ, dom (λ), ρ_1, and ρ_2, respectively.

Clearly, a word can be considered as a pair (λ, ρ), where ρ is a linear order, and λ is a function on dom(ρ). Hence a text may be considered as a word α, equipped with an additional linear order which assigns, together with the obvious left-to-right order, a syntactic structure (shape) to α.

A *T-function* is an ordered pair (λ, g) such that λ is a function and g is a lTs on the domain of λ; hence a T-function is a lTs together with values assigned (by λ) to elements of its domain. All the notation and terminology concerning labeled T-structures carries over to T-functions.

DEFINITION 3.7: Let $\tau = (\lambda, \rho_1, \rho_2)$ be a text.

(1) The *T-function of* τ, denoted $Tf(\tau)$, is the T-function (λ, g) such that $\text{dom}(g) = \text{dom}(\tau)$, $\text{part}(g) = \{\rho_1 \cap \rho_2, \quad \rho_1 \cap \text{rev}(\rho_2), \quad \text{rev}(\rho_1) \cap \rho_2, \text{rev}(\rho_1) \cap \text{rev}\,\rho_2)\} - \{\varnothing\}$, and $\text{lab}(g) = \delta$, where $\delta(\rho_1 \cap \rho_2) = VH$, $\delta(\rho_1 \cap \text{rev}(\rho_2)) = V\bar{H}$, $\delta(\text{rev}(\rho_1) \cap \rho_2) = \bar{V}H$, and $\delta(\text{rev}(\rho_1) \cap \text{rec}(\rho_2)) = \bar{V}\bar{H}$.

(2) The *shape of* τ, denoted $shape(\tau)$, is the shape of $Tf(\tau)$. □

Note that the labels of classes in $Tf(\tau)$ are very specific, they allow us to recognize which intersection of linear orders is used to define a given class.

Example 3.4: Let $\tau = (\lambda, (1, 6, 3, 4, 2, 5), (2, 4, 6, 3, 5, 1))$, where $\lambda(1, 6, 3, 4, 2, 5) = aabaab$. Then $Tf(\tau)$ is as in figure 5.

shape (τ) is as in figure 6. □

For a text τ, the *length of* τ, denoted by $|\tau|$, is $\#\text{dom}(\tau)$. A *singleton text* is a text τ with $|\tau| = 1$.
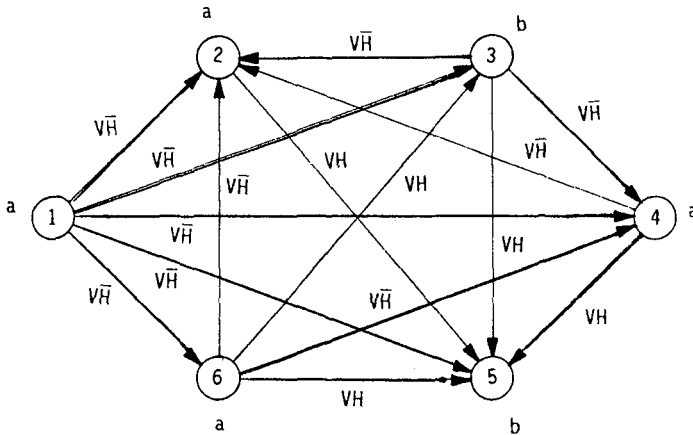
Figure 5.

A text τ is *standard* if $VO(\tau)$ is $(1, 2, \ldots, |\tau|)$. A standard text of length n is determined by a permutation on $\{1, \ldots, n\}$ and a labeling function on $\{1, \ldots, n\}$.

Two texts τ_1 and τ_2 are *isomorphic* if there exists a bijection $\varphi : \text{dom}(\tau_1) \rightarrow \text{dom}(\tau_2)$ such that φ is an isomorphism between
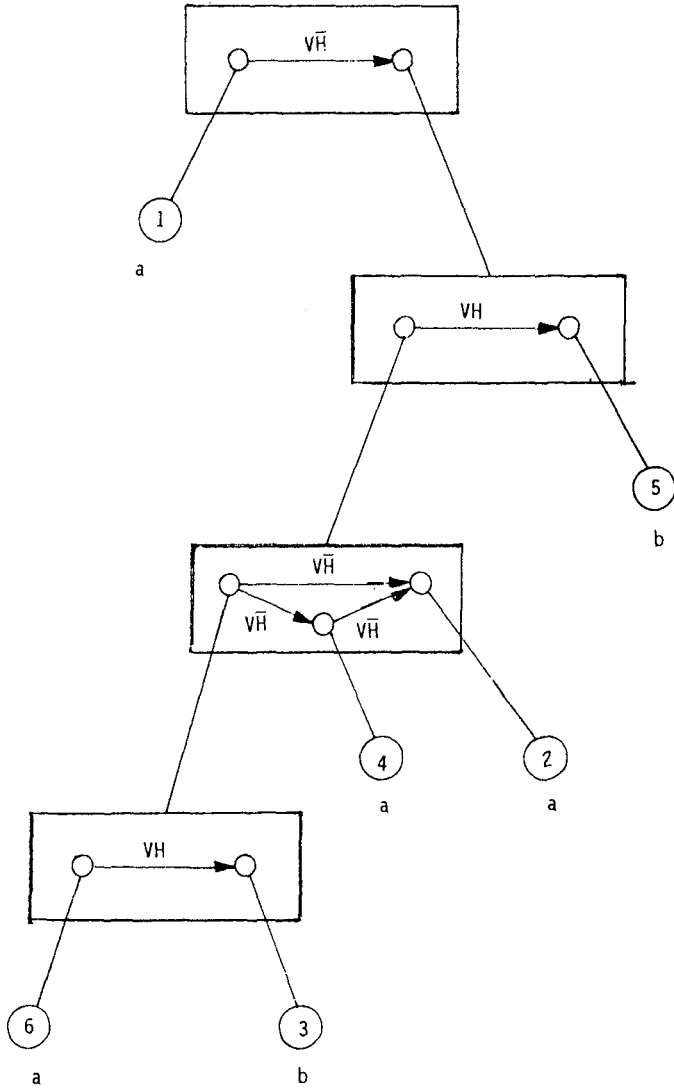
Figure 6.

$(\mathrm{dom}(\tau_1),\ VO(\tau_1))$ and $(\mathrm{dom}(\tau_2),\ VO(\tau_2))$, φ is an isomorphism between $(\mathrm{dom}(\tau_1),\ HO(\tau_1))$ and $(\mathrm{dom}(\tau_2),\ HO(\tau_2))$, and, for each $x \in \mathrm{dom}(\tau_1)$, $\lambda_1(x) = \lambda_2(\varphi(x))$; then φ is an *isomorphism from* τ_1 *onto* τ_2.

It is easy to see that every text is isomorphic with a unique standard text, and hence one often considers standard texts only.

Note that if a shape has linear nodes only, then the features of the nodes are alternately labeled $\{VH, \bar{V}\bar{H}\}$ and $\{V\bar{H}, \bar{V}H\}$. Hence a text τ is called *alternating* iff each inner node of *shape* (τ) is linear, otherwise τ is *non-alternating*. A text τ is *primitive* iff $Tf(\tau)$ is primitive. The classes of alternating, non-alternating, and primitive texts are denoted by **ALT**, **NALT**, and **PRIM** respectively.

In [3] two characterizations of alternating texts are given: a combinatorial one and an operational one. The following notion is used to formulate the combinatorial characterization.

DEFINITION 3.8: Let τ be a text, and let $I = \{i_1, i_2, i_3, i_4\} \subseteq \mathrm{dom}(\tau)$. Then I is a *primitive quartet of* τ iff either

$$\tau|_I = (\lambda|_I, (i_1, i_2, i_3, i_4), (i_2, i_4, i_1, i_3)),$$

or

$$\tau|_I = (\lambda|_I, (i_1, i_2, i_3, i_4), (i_3, i_1, i_4, i_2)).$$

In the former case, I is a *left* primitive quartet of τ, in the latter case I is a *right* primitive quartet of τ.   □

PROPOSITION 3.1: *Let* τ *be a text.* $\tau \in$ **NALT** *iff* τ *has a primitive quartet.*   □

To state the operational characterization, we first recall some operations on texts that were defined in [3]. (Two texts are called *disjoint* iff their domains are disjoint.)

DEFINITION 3.9: Let $\tau = (\lambda, \rho_1, \rho_2)$ and $\tau' = \lambda', \rho_1', \rho_2')$ be disjoint texts.

(1) The V-*reverse of* τ, denoted $V\mathrm{rev}(\tau)$, is the text $(\lambda, \mathrm{rev}(\rho_1), \rho_2)$.

(2) The H-*reverse of* τ, denoted $H\mathrm{rev}(\tau)$, is the text $(\lambda, \rho_1, \mathrm{rev}(\rho_2))$.

(3) The sum *of* τ *and* τ', denoted $\tau \oplus \tau'$, is the text

$(\lambda \cup \lambda', \rho_1 + \rho_1', \rho_2 + \rho_2')$.   □

It is proved in [3] that **ALT** is closed under the above operations. As a matter of fact the following characterization of **ALT** is proved in [3]. (A *singleton text* is a text τ such that $\mathrm{dom}(\tau) = \{x\}$, and $VO(\tau) = HO(\tau) = (x)$.)

PROPOSITION 3.2: (*a*) **ALT** *is the smallest class of texts containing all singleton texts that is closed under* $V\mathrm{rev}$ *and* $\oplus$.

(*b*) **ALT** *is the smallest class of texts containing all singleton texts that is closed under* $H\mathrm{rev}$ *and* $\oplus$.   □

4. SORTING SEQUENCES

As observed in the previous section, a standard text is determined by a permutation, together with a labeling function. In this paper we take this point of view, and in particular we consider permutations corresponding to (standard) texts as sorting processes. In this section we introduce a number of basic notions corresponding to this point of view.

Given a subset X of the domain of a text τ, $HO(\tau)$ determines a partition of X into segments of $HO(\tau)$ as follows.

DEFINITION 4.1: Let τ be a text, and let $X \subseteq \text{dom}(\tau)$, $X \neq \emptyset$. The *H-segmentation of X in* τ, denoted by $H_\tau(X)$, is the m-tuple $(\sigma_1, \ldots, \sigma_m)$, $m \geq 1$, such that

(i) each σ_i, $1 \leq i \leq m$, is a nonempty segment of $HO(\tau)$,

(ii) $\bigcup\limits_{i=1}^{m} \sigma_i = X$,

(iii) for all $1 \leq i \leq j \leq m$, $\sigma_i \cup \sigma_j$ is not a segment of $HO(\tau)$, and

(iv) for all $1 \leq i < j \leq m$, if $x \in \sigma_i$ and $y \in \sigma_j$ then $(x, y) \in HO(\tau)$. $\square$

Similarly, one can define the *V*-segmentation of X in τ; however, since we deal mostly with standard texts, we will not consider *V*-segmentations in this paper.

Example 4.1: Let τ be a standard text such that $HO(\tau) = (3, 6, 1, 4, 5, 2)$. Then $H_\tau(\{1, 3\}) = (\{3\}, \{1\})$, $H_\tau(\{2, 3, 4, 6\}) = (\{3, 6\}, \{4\}, \{2\})$, and $H_\tau(\{2, 4, 5\}) = (\{2, 4, 5\})$. $\square$

A way to investigate the relationship between $HO(\tau)$ and $VO(\tau)$ for a standard text τ is to investigate the sequences of *H*-segmentations corresponding to the sequence $(1), (1, 2), \ldots, (1, \ldots, i), \ldots, (1, \ldots, |\tau|)$ of prefixes of $VO(\tau)$. This sequence is formally defined as follows.

DEFINITION 4.2: Let τ be a standard text of length n. The *sorting sequence of* τ, denoted $sort_\tau$, is the sequence $(H_\tau(X_1), \ldots, H_\tau(X_n))$, where $X_i = \{1, \ldots, i\}$ for $1 \leq i \leq n$. $\square$

We use the following convenient notation. For a text τ, and for i, $k \in \text{dom}(\tau)$ such that either $(k, i) \in VO(\tau)$ or $k = i$, $k(sort_\tau(i))$ denotes the segment of $sort_\tau(i)$ that contains k.

Example 4.2: (Example 4.1 continued): Let τ be the standard text from Example 4.1. Then $sort_\tau$ is as follows.

$sort_\tau(1) = (\{1\})$

$\text{sort}_\tau(2) = (\{1\}, \{2\})$

$\text{sort}_\tau(3) = (\{3\}, \{1\}, \{2\})$

$\text{sort}_\tau(4) = (\{3\}, \{1, 4\}, \{2\})$

$\text{sort}_\tau(5) = (\{3\}, \{1, 2, 4, 5\})$

$\text{sort}_\tau(6) = (\{1, 2, 3, 4, 5, 6\}).$

Hence $1(\text{sort}_\tau(4)) = \{1, 4\}$, $1(\text{sort}_\tau(5)) = \{1, 2, 4, 5\}$, $2(\text{sort}_\tau(4)) = \{2\}$, $3(\text{sort}_\tau(4)) = \{3\}$, and $4(\text{sort}_\tau(4)) = \{1, 4\}$. $\quad\square$

Now we will introduce some parameters which are useful in describing the formation of the sorting sequence of a given text.

Let τ be a standard text of length n and let $i \in \text{dom}(\tau)$, $i > 1$. When i is sorted into $\text{sort}_\tau(i-1)$, it is put in some "gap" between segments of $\text{sort}_\tau(i-1)$. The parameter $\text{jump}_\tau(i)$ will give the number of segments that are passed over when jumping from $i-1$ to i, that is, the number of segments between the segment to the left of this gap and the segment that contains $i-1$. The parameter $\text{type}_\tau(i)$ will represent the way i is placed in the gap.

Let $\text{sort}_\tau(i-1) = (\sigma_1, \ldots, \sigma_m)$.

If σ_k is the segment of $\text{sort}_\tau(i-1)$ that contains $i-1$, $i.e.$,

$\sigma_k = (i-1)(\text{sort}_\tau(i-1))$, then k is denoted by $\text{pos}_\tau(i-1)$.

If σ_l is the segment of $\text{sort}_\tau(i-1)$ that is immediately to the left of i, then l is denoted by $\text{ind}_\tau(i)$. Formally, $\text{ind}_\tau(i)$ is defined as follows:

$\text{ind}_\tau(i) = 0$ if $(i, j) \in HO(\tau)$ for each $j \in \sigma_1$,

$\text{ind}_\tau(i) = m$ if $(j, i) \in HO(\tau)$ for each $j \in \sigma_m$,

$\text{ind}_\tau(i) = l$ if $(j, i) \in HO(\tau)$ for each $j \in \sigma_l$, and $(i, t) \in HO(\tau)$ for each $t \in \sigma_{l+1}$.

For the sake of completeness, we define $\text{pos}_\tau(n) = 1$ (since $n(\text{sort}_\tau(n)) = \text{dom}(\tau)$), and $\text{ind}_\tau(1) = 0$. Hence pos_τ and ind_τ are functions on $\text{dom}(\tau)$.

Now, as described above, $\text{jump}_\tau(i)$ is the number of segments between $\text{pos}_\tau(i-1)$ and $\text{ind}_\tau(i)$, together with a sign for the direction.

Furthermore, $\text{type}_\tau(i)$ says whether i is forming a segment of its own in $\text{sort}_\tau(i)$, or i is joining one segment of $\text{sort}_\tau(i-1)$ (which may be either to the left or to the right of i), or joining two segments of $\text{sort}_\tau(i-1)$ (one to the left and one to the right of i).

The functions jump_τ and type_τ are given formally by the following definition.

DEFINITION 4.3: Let τ be a standard text. Then

(1) $\text{jump}_\tau : \text{dom}(\tau) \to \mathbb{Z}$ is defined by $\text{jump}_\tau(1) = 0$, and for $i = 2, \ldots, n$, $\text{jump}_\tau(i) = \text{ind}_\tau(i) - \text{pos}_\tau(i-1)$;

(2) $\text{type}_\tau : \text{dom}(\tau) \to \{S, R, L, B\}$ is defined by for $i \in \text{dom}(\tau)$ and k such that $HO(\tau)(k) = i$,

$$\text{type}_\tau(i) = \begin{cases} S \text{ if } (HO(\tau)(k-1) > i \text{ or } k=1) \\ \quad \text{and } (HO(\tau)(k+1) > i \text{ or } k=n) \\ L \text{ if } HO(\tau)(k-1) < i \\ \quad \text{and } (HO(\tau)(k+1) > i \text{ or } k=n) \\ R \text{ if } (HO(\tau)(k-1) > i \text{ or } k=1) \\ \quad \text{and } (HO(\tau)(k+1) < i \\ B \text{ if } HO(\tau)(k-1) < i \\ \quad \text{and } HO(\tau)(k+1) < i. \end{cases}$$

Example 4.3: (Examples 4.1 and 4.2 continued): Let τ be the text of Example 4.1. The following table gives pos_τ, type_τ, ind_τ, and jump_τ.

i	$\text{pos}_\tau(i)$	$\text{type}_\tau(i)$	$\text{ind}_\tau(i)$	$\text{jump}_\tau(i)$
1.	1	S	0	0
2.	2	S	1	0
3.	1	S	0	−2
4.	2	L	2	1
5.	2	B	2	0
6.	1	B	1	−1

Note that using type_τ and jump_τ we can reconstruct $HO(\tau)$ step-by-step.

Clearly $HO(\tau)|_{\{1\}} = (1)$. Since $\text{jump}_\tau(2) = 0$, $HO(\tau)|_{\{1,2\}} = (1, 2)$. Since $\text{type}_\tau(2) = S$ and $\text{jump}_\tau(3) = -2$, $HO(\tau)|_{\{1,2,3\}} = (3, 1, 2)$. Since $\text{type}_\tau(3) = S$ and $\text{jump}_\tau(4) = 1$, $HO(\tau)|_{\{1,2,3,4\}} = (3, 1, 4, 2)$. Since $\text{type}_\tau(4) = L$ and $\text{jump}_\tau(5) = 0$, $HO(\tau)|_{\{1,2,3,4,5\}} = (3, 1, 4, 5, 2)$. Finally, since $\text{type}_\tau(5) = B$, $\text{type}_\tau(4) = L$, and $\text{type}_\tau(3) = \text{type}_\tau(2) = S$, $\text{sort}_\tau(5) = (\{3\}, \{1, 4, 5, 2\})$, Hence, since $\text{jump}_\tau(6) = -1$, $HO(\tau)|_{\{1,2,3,4,5,6\}} = (3, 6, 1, 4, 5, 2)$. $\square$

We will prove now that the above example illustrates a general situation: a text τ is uniquely determined by fun_τ, type_τ, and jump_τ.

THEOREM 4.1: *A standard text τ is uniquely determined by the triplet* $\{\text{fun}_\tau, \text{type}_\tau, \text{jump}_\tau\}$.

Proof: We will prove by induction on i that $\text{pos}_\tau(i)$, $\text{sort}_\tau(i)$, and $HO(\tau)|_{\{1, \ldots i\}}$ are uniquely determined by type_τ and jump_τ.

If $i=1$, then, by definition, $\text{pos}_\tau(i)=1$, $\text{sort}_\tau(i)=(\{1\})$, and $HO(\tau)|_{\{1,\ldots i\}}=(1)$.

Let $i>1$, and suppose that $\text{pos}_\tau(i-1)$, $\text{sort}_\tau(i-1)$, and $HO(\tau)|_{\{1,\ldots,i-1\}}$ are uniquely determined by type_τ and jump_τ. Let $k=\text{pos}_\tau(i-1)+\text{jump}_\tau(i)$, let $\text{sort}_\tau(i-1)=(\sigma_1,\ldots,\sigma_m)$ and let $HO(\tau)|_{\{1,\ldots,i-1\}}=(a_1,\ldots,a_{i-1})$. By the definition of jump_τ, $k=\text{ind}_\tau(i)$.

Hence

$$\text{pos}_\tau(i)=\begin{cases} k & \text{if} & \text{type}_\tau(i)\in\{L,B\} \\ k+1 & \text{if} & \text{type}_\tau(i)\in\{S,R\} \end{cases}$$

$$\text{sort}_\tau(i)=\begin{cases} (\sigma_1,\ldots,\sigma_k\cup\{i\},\sigma_{k+1},\ldots,\sigma_m) & \text{if} & \text{type}_\tau(i)=L \\ (\sigma_1,\ldots,\sigma_k,\{i\}\cup\sigma_{k+1},\ldots,\sigma_m) & \text{if} & \text{type}_\tau(i)=R \\ (\sigma_1,\ldots,\sigma_k,\{i\},\sigma_{k+1},\ldots,\sigma_m) & \text{if} & \text{type}_\tau(i)=S \\ (\sigma_1,\ldots,\sigma_k\cup\{i\}\cup\sigma_{k+1},\ldots,\sigma_m) & \text{if} & \text{type}_\tau(i)=B \end{cases}$$

and

$$HO(\tau)|_{\{1,\ldots,i\}}=(a_1,\ldots,a_j,i,a_{j+1},\ldots,a_{i-1}),$$

where j is such that $a_j\in\sigma_k$ and $a_{j+1}\notin\sigma_k$.

This completes the induction.

Hence in particular $HO(\tau)$ is uniquely determined by type_τ and jump_τ, and so τ is uniquely determined by type_τ and jump_τ. $\square$

Example 4.4: Consider τ such that jump_τ and type_τ are as follows:

	1	2	3	4	5	6	7	8
jump_τ . . .	0	0	-1	0	-1	-1	-1	0
type_τ . . .	S	S	R	L	B	R	S	B

The step-by-step construction of $HO(\tau)$ looks as follows (in each step the underlined number is the element of $\text{dom}(\tau)$ that is sorted according to its jump and type).

$HO\,(\tau)|_{\{1\}}$ $\underline{1}$

$HO\,(\tau)|_{\{1,\,2\}}$ 1 $\underline{2}$

$HO\,(\tau)|_{\{1,\,2,\,3\}}$ 1 $\underline{3}$ 2

$HO\,(\tau)|_{\{1,\,2,\,3,\,4\}}$ 1   3 2 $\underline{4}$

$HO\,(\tau)|_{\{1,\,2,\,3,\,4,\,5\}}$ 1 $\underline{5}$ 3 2 4

$HO\,(\tau)|_{\{1,\,2,\,3,\,4,\,5,\,6\}}$ $\underline{6}$ 1 5 3 2 4

$HO\,(\tau)|_{\{1,\,2,\,3,\,4,\,5,\,6,\,7\}}$ $\underline{7}$ 6 1 5 3 2 4

$HO\,(\tau)|_{\{1,\,2,\,3,\,4,\,5,\,6,\,7,\,8\}}$ 7 $\underline{8}$ 6 1 5 3 2 4 $\square$

Remark 4.1: So far, we have only considered standard texts. We could however extend Definitions 4.2 to 4.3 to arbitrary texts, by replacing the order " $<$ " by the order VO.

Hence, if τ is an arbitrary text, then $\mathrm{sort}_\tau = (H_\tau(X_1), \ldots, H_\tau(X_n))$, where $X_i = \{ VO\,(\tau)\,(j) \mid j = 1, \ldots, i \}$ for $1 \leq i \leq n$, and $\mathrm{jump}_\tau\,(VO\,(\tau)\,(j)) = \mathrm{jump}_{\tau'}\,(j)$ for $j = 1, \ldots, |\tau|$, where τ' is the standard text isomorphic with τ. Analogously, one can define pos_τ, type_τ, and ind_τ for an arbitrary text τ.

For reasons of readability, the results we present are essentially in terms of standard texts. However, it should be clear that they can easily be extended to arbitrary texts. For example, the general result corresponding to Theorem 2.1 says that each text τ is uniquely determined by fun_τ, jump_τ, type_τ, and $VO\,(\tau)$. $\square$

The operations of reverses are both natural and important in the theory of texts (*see* [3]). We will now investigate the relationship between sorting sequences of texts and sorting sequences of their reverses.

THEOREM 4.2: *Let τ be a text such that $|\tau| = n$.*

(1)   *If $\tau' = V\,\mathrm{rev}\,(\tau)$, then for all $i \in \mathrm{dom}\,(\tau)$,*

$$-1 \leq |\mathrm{sort}_{\tau'}\,(n-i)| - |\mathrm{sort}_\tau\,(i)| \leq 1.$$

(2)   *If $\tau' = H\,\mathrm{rev}\,(\tau)$, then for all $i \in \mathrm{dom}\,(\tau)$,*

$$|\mathrm{sort}_{\tau'}\,(i)| = |\mathrm{sort}_\tau\,(i)|.$$

Proof: We may assume that τ is a standard text of length n. Let $i \in \mathrm{dom}\,(\tau)$, and let $\mathrm{sort}_\tau\,(i) = (\sigma_1, \ldots, \sigma_m)$, $m \geq 1$. Note that $\bigcup\limits_{j=1}^{m} \sigma_j = \{1, \ldots, i\}$.

Consider $\tau' = V \operatorname{rev}(\tau)$. The element i is sorted in τ' in the $n - i + 1'$ th step, after $n, n-1, \ldots, i+1$ have been sorted. Let

$$\operatorname{sort}_{\tau'}(n-i) = (\sigma'_1, \ldots, \sigma'_l), \quad l \geq 1.$$

Thus, $\displaystyle\bigcup_{j=1}^{m} \sigma'_j = \{ i+1, \ldots, n \}$.

Hence $\displaystyle\left(\bigcup_{j=1}^{m} \sigma_j \right) \cup \left(\bigcup_{j=1}^{l} \sigma'_j \right) = \{ 1, \ldots, n \}$, and the segments of $\operatorname{sort}_\tau(i)$ alternate with the segments of $\operatorname{sort}_{\tau'}(n-i)$. We describe this situation by saying that $\operatorname{sort}_\tau(i)$ and $\operatorname{sort}_{\tau'}(n-i)$ are *complementary*. Consequently, $-1 \leq |\operatorname{sort}_{\tau'}(n-i)| - |\operatorname{sort}_\tau(i)| \leq 1$.

Now consider $\tau' = H \operatorname{rev}(\tau)$. Then $\operatorname{sort}_{\tau'}(i) = (\sigma_m, \ldots, \sigma_1)$. Hence $\operatorname{sort}_{\tau'}(i) = \operatorname{rev}(\operatorname{sort}_\tau(i))$, and $|\operatorname{sort}_{\tau'}(i)| = |\operatorname{sort}_\tau(i)|$. □

Example 4.5: Let τ be a standard text such that $HO(\tau) = (3, 2, 5, 4, 6, 7, 1)$. Then $\operatorname{sort}_\tau(4) = (\{ 2, 3 \}, \{ 4 \}, \{ 1 \})$.

Let $\tau' = V \operatorname{rev}(\tau) = (\operatorname{fun}_\tau, (7, 6, 5, 4, 3, 2, 1), (3, 2, 5, 4, 6, 7, 1)$. Then, since after three sorting steps 7, 6, and 5 are sorted, $\operatorname{sort}_{\tau'}(3) = (\{ 5 \}, \{ 6, 7 \})$. Hence $\operatorname{sort}_\tau(4)$ and $\operatorname{sort}_{\tau'}(3)$ are complementary, and $|\operatorname{sort}_\tau(4)| = 3$, $|\operatorname{sort}_{\tau'}(3)| = 2$.

Now let $\tau' = H \operatorname{rev}(\tau) = (\operatorname{fun}_\tau, (1, 2, 3, 4, 5, 6, 7), (1, 7, 6, 4, 5, 2, 3))$. Then $\operatorname{sort}_{\tau'}(4) = (\{ 1 \}, \{ 4 \}, \{ 2, 3 \}) = \operatorname{rev}(\operatorname{sort}_\tau(4))$, and $|\operatorname{sort}_{\tau'}(4)| = 3$. □

A more detailed description of the relationship between the sorting sequences of a text τ and the sorting sequences of $V \operatorname{rev}(\tau)$ and $H \operatorname{rev}(\tau)$ is given by the following result.

THEOREM 4.3: *Let τ be a standard text of length n, and let τ' be a text such that* $\operatorname{dom}(\tau') = \{ 1, \ldots, n \}$ *and* $\operatorname{fun}_{\tau'} = \operatorname{fun}_\tau$. *Let* $\psi : \{ S, B, L, R \} \to \{ S, B, L, R \}$ *be the bijection defined by* $\psi(S) = B, \psi(B) = S, \psi(L) = R,$ *and* $\psi(R) = L$.

(1) $\tau' = V \operatorname{rev}(\tau)$ *iff* $VO(\tau') = n, n-1, \ldots, 1)$ *and*

(i) $\operatorname{jump}_{\tau'}(n) = 0, \quad$ *and for* $j = 1, \ldots, n-1,$
$\operatorname{jump}_{\tau'}(j) = -\operatorname{jump}_\tau(j+1) - 1,$

(ii) *for* $j = 1, \ldots, n,$
if $j = HO(\tau)(1),$ *then*

$$\operatorname{type}_{\tau'}(j) = S \quad if \quad \operatorname{type}_\tau(j) = R,$$

and

$$\operatorname{type}_{\tau'}(j) = R \quad if \quad \operatorname{type}_\tau(j) = S,$$

if $j = HO(\tau)(n)$, then

$$\text{type}_{\tau'}(j) = S \quad if \quad \text{type}_\tau(j) = L,$$

and

$$\text{type}_{\tau'}(j) = L \quad if \quad \text{type}_\tau(j) = S$$

otherwise

$$\text{type}_{\tau'}(j) = \psi(\text{type}_\tau(j)).$$

(2) $\tau' = H \operatorname{rev}(\tau)$ *iff* $VO(\tau') = (1, 2, \ldots, n)$ *and*

(i) $\quad \begin{aligned} &\text{jump}_{\tau'}(1) = 0, \quad \text{and for} \quad j = 2, \ldots, n, \\ &\text{jump}_{\tau'}(j) = -\text{jump}_\tau(j) - 1, \end{aligned}$

(ii) *for* $j = 1, \ldots, n,$
if $\text{type}_\tau(j) \in \{L, R\}$, *then*

$$\text{type}_{\tau'}(j) = \psi(\text{type}_\tau(j)),$$

otherwise,

$$\text{type}_{\tau'}(j) = \text{type}_\tau(j).$$

Proof: It suffices to prove that if $\tau' = V \operatorname{rev}(\tau)$, resp. $\tau' = H \operatorname{rev}(\tau)$, then conditions [1, (i)], [1, (ii)], resp. [2, (i)], [2, (ii)] hold. The other direction follows then from Theorem 4.1.

(1) Let $\tau' = V \operatorname{rev}(\tau)$. By definition, $\text{jump}_{\tau'}(n) = 0$. Let $j < n$ and consider $\text{sort}_{\tau'}(n - j)$ in order to find $\text{jump}_{\tau'}(j)$. We will compare it to $\text{jump}_\tau(j + 1)$, which depends on $\text{sort}_\tau(j)$. We have already observed that the sorting sequences $\text{sort}_\tau(j)$ and $\text{sort}_{\tau'}(n - j)$ are complementary. We will distinguish two cases: either the first element of $HO(\tau)$ is sorted already in $\text{sort}_\tau(j)$, or it is sorted already in $\text{sort}_{\tau'}(n - j)$. In the first case, *i.e.*, $(HO(\tau)(1), j) \in VO(\tau)$, we have that $\text{ind}_{\tau'}(j) = \text{pos}_\tau(j) - 1$ and $\text{pos}_{\tau'}(j + 1) = \text{ind}_\tau(j + 1)$. In the second case, $(HO(\tau)(1), j) \in VO(\tau')$, and then $\text{ind}_{\tau'}(j) = \text{pos}_\tau(j)$ and $\text{pos}_{\tau'}(j + 1) = \text{ind}_\tau(j + 1) + 1$. In both cases,

$$\begin{aligned} \text{jump}_{\tau'}(j) &= \text{ind}_{\tau'}(j) - \text{pos}_{\tau'}(j + 1) \\ &= \text{pos}_\tau(j) - \text{ind}_\tau(j + 1) - 1 = -\text{jump}_\tau(j + 1) - 1. \end{aligned}$$

To determine the type of j in τ' for $j = 1, \ldots, n$ we first consider the four special cases.

If $j = HO(\tau)(1)$ and $\text{type}_\tau(j) = R$, then $(HO(\tau)(2), j) \in VO(\tau)$. Hence $(j, HO(\tau)(2)) \in VO(\tau')$, which implies that $\text{type}_{\tau'}(j) = S$. Analogously, the other three special cases can be proved.

It is easily seen that if $j \neq HO(\tau)(1)$ and $j \neq HO(\tau)(n)$, then $\text{type}_{\tau'}(j) = \psi(\text{type}_\tau(j))$, because every element other than j that is already sorted in $\text{sort}_\tau(n-j)$ is not yet sorted in $\text{sort}_\tau(j-1)$ and vice versa.

(2) Let $\tau' = H\,\text{rev}(\tau)$. By definition $\text{jump}_{\tau'}(1) = 0$. Let $j > 1$ and let $|\text{sort}_{\tau'}(j)| = m$. Since $\text{sort}_{\tau'}(j) = \text{rev}(\text{sort}_\tau(j))$,

$$\text{jump}_{\tau'}(j) = \text{ind}_{\tau'}(j) - \text{pos}_{\tau'}(j-1)$$

$$= m - \text{ind}_\tau(j) - (m - \text{pos}_\tau(j-1) + 1) = -\text{jump}_\tau(j) - 1.$$

The type of j in τ' for $j \in \{1, \ldots, n\}$ is immediately obtained from the fact that $\text{sort}_{\tau'}(j) = \text{rev}(\text{sort}_\tau(j))$. $\square$

Example 4.6: (Example 4.5 continued.): Let τ be the text from Example 4.5. Let $\tau' = V\,\text{rev}(\tau)$, and consider $\text{sort}_{\tau'}(3)$. Note that $\text{jump}_{\tau'}(4) = 0$. Compare this to inserting 5 into $\text{sort}_\tau(4)$, then indeed $\text{jump}_{\tau'}(4) = -\text{jump}_\tau(5) - 1 = 0$. Furthermore it is clear that $\text{type}_{\tau'}(4) = B = \psi(\text{type}_\tau(4)) = \psi(S)$.

Now let $\tau' = H\,\text{rev}(\tau)$. Then $\text{jump}_{\tau'}(4) = -1 = -\text{jump}_\tau(4) - 1$, and $\text{type}_{\tau'}(4) = \text{type}_\tau(4) = S$. $\square$

Note that $V\,\text{rev}(\tau)$ is not standard, so Theorem 4.3 (1) in this form cannot be applied twice to obtain τ again. Clearly, Theorem 4.3 can easily be extended to the case that τ is not standard (*see* Remark 4.1) — then applying the first part of the theorem twice *does* give the original text.

5. ALTERNATING TEXTS

In this section we investigate the sorting sequences of alternating texts. To start with, we consider the simplest alternating texts.

DEFINITION 5.1: A text τ is *sequential* iff either $VO(\tau) = HO(\tau)$ or $VO(\tau) = \text{rev}(HO(\tau))$. If the former holds, then τ is *forward sequential*, and if the latter holds, then τ is a *backward sequential*. $\square$

Sequential texts are called "T-vectors" in [4]. It is not difficult to see that a text is sequential iff its T-function is linear.

For a text τ we will use $\overline{\text{dom}}(\tau)$ to denote $\text{dom}(\tau) - \{VO(\tau)(1)\}$.

THEOREM 5.1: *Let τ be a text.*

(1) τ *is forward sequential iff for all* $i \in \overline{\text{dom}}(\tau)$, $\text{jump}_\tau(i) = 0$.

(2) τ *is backward sequential iff for all* $i \in \overline{\text{dom}}(\tau)$, $\text{jump}_\tau(i) = -1$.

Proof: (1) Let τ be a forward sequential text. Obviously, $\text{pos}_\tau(i) = 1$ for all $i \in \text{dom}(\tau)$ and $\text{ind}_\tau(i) = 1$ for all $i \in \overline{\text{dom}}(\tau)$. Hence by Definition 4.3 (4), $\text{jump}_\tau(i) = 0$ for all $i \in \overline{\text{dom}}(\tau)$.

Let τ be a text such that $\text{jump}_\tau(i) = 0$ for all $i \in \overline{\text{dom}}(\tau)$. Then $HO(\tau) = VO(\tau)$, and τ is forward sequential.

(2) Let τ be a backward sequential text. Then $\text{pos}_\tau(i) = 1$ for all $i \in \text{dom}(\tau)$ and $\text{ind}_\tau(i) = 0$ for all $i \in \overline{\text{dom}}(\tau)$. Hence, by Definition 4.3 (4), $\text{jump}_\tau(i) = -1$ for all $i \in \overline{\text{dom}}(\tau)$.

Let τ be a text such that $\text{jump}_\tau(i) = -1$ for all $i \in \overline{\text{dom}}(\tau)$. Then $HO(\tau) = \text{rev}(VO(\tau))$, and τ is backward sequential. □

In view of the above result it is natural to consider texts in which the jumps are restricted to $\{0, -1\}$.

DEFINITION 5.2: A text τ is *jump-free* iff for each $i \in \text{dom}(\tau)$, $\text{jump}_\tau(i) \in \{0, -1\}$. □

We will use $J_{\{0, -1\}}$ to denote the class of all jump-free texts.

Our next result supports the naturalness of $J_{\{0, -1\}}$: all alternating texts are in $J_{\{0, -1\}}$.

THEOREM 3.2: *If a text τ is alternating, then τ is jump-free.*

Proof: Let τ be an alternating text. We may assume that τ is standard.

Assume to the contrary that τ is not jump-free, *i.e.*, there exists an $i \in \text{dom}(\tau)$ such that either $\text{jump}_\tau(i) > 0$ or $\text{jump}_\tau(i) < -1$. We will show that this implies that τ has a primitive quartet, which by Proposition 3.1 contradicts the fact that τ is alternating.

Assume that there exists an $i \in \text{dom}(\tau)$ such that $\text{jump}_\tau(i) > 0$. Let $\text{sort}_\tau(i-1) = (\sigma_1, \ldots, \sigma_m)$, $m \geq 1$, and let $t = \text{pos}_\tau(i-1)$. Then $t \neq m$, since $\text{jump}_\tau(i) > 0$. Let $j \in \sigma_{t+1}$, and let $k \in \text{dom}(\tau)$ be such that $(x, k) \in HO(\tau)$ for each $x \in \sigma_t$, and $(k, y) \in HO(\tau)$ for each $y \in \sigma_{t+1}$. Then $j < i-1$ and $k > i$. Consequently, if $I = \{j, i-1, i, k\}$, then $VO(\tau)|_I = (j, i-1, i, k)$ and $HO(\tau)|_I = (i-1, k, j, i)$. Hence I is a left primitive quartet.

Analogously, in the case that there exists an $i \in \text{dom}(\tau)$ such that $\text{jump}_\tau(i) < -1$, we find a right primitive quartet.

Consequently, by Proposition 3.1, τ is not alternating. Hence the assumption that τ is not jump-free yields a contradiction. Consequently τ is jump-free. □

Remark 5.1: Although $\mathbf{ALT} \subseteq J_{\{0, -1\}}$, there exist texts in $J_{\{0, -1\}}$ that are not alternating as seen by the following example.

Let τ be a standard text such that $HO(\tau) = (2, 5, 3, 1, 4)$. Then sort_τ and jump_τ are as follows.

i	$\text{sort}_\tau(i)$	$\text{jump}_\tau(i)$
1.	$(\{1\})$	0
2.	$(\{2\}, \{1\})$	-1
3.	$(\{2\}, \{1, 3\})$	0
4.	$(\{2\}, \{1, 3, 4,\})$	0
5.	$(\{1, 2, 3, 4, 5\})$	-1

Hence $\tau \in J_{\{0, -1\}}$.

On the other hand, $\{1, 2, 4, 5\}$ is a primitive quartet. Consequently, by Proposition 3.1, τ is not alternating.   □

Hence, in order to characterize alternating texts, we have to impose more restrictions on the formation of sorting sequences. First we give a sharper version of Proposition 3.1.

THEOREM 5.3: *Let τ be a text. The following statements are equivalent:*

(1) *τ is not alternating;*

(2) *τ has a primitive quartet;*

(3) *τ has a primitive quartet I such that if $\rho = VO(\tau)|_I$, then $\{\rho(1), \rho(2)\} \in \text{seg}(VO(\tau))$.*

Proof: Let τ be a text. We may assume that τ is standard.

(1) iff (2): This follows directly from Proposition 3.1.

(3) implies (2): Obvious.

(2) implies (3): It is sufficient to consider the case that τ has a right primitive quartet. The case that τ has a left primitive quartet follows by an analogous argument.

Let $I = \{i_1, i_2, i_3, i_4\}$ be a right primitive quartet of τ such that $VO(\tau)|_I = (i_1, i_2, i_3, i_4)$.

Let $J = \{j \in \{1, \ldots, n\} \mid (i_3, j), (j, i_4) \in HO(\tau) \text{ and } i_1 \leq j < i_2\}$. Note that $J \neq \varnothing$, since $i_1 \in J$. Let $j_0 = \max(J)$.

Claim 5.1: Either $(i_4, j_0 + 1) \in HO(\tau)$ or $(j_0 + 1, i_3) \in HO(\tau)$.

Proof of Claim 5.1: If $j_0 + 1 = i_2$, then the claim holds, because $(i_4, i_2) \in HO(\tau)$.

If $j_0 + 1 < i_2$, then $(i_4, j_0 + 1) \notin HO(\tau)$ and $(j_0 + 1, i_3) \notin HO(\tau)$ imply $j_0 + 1 \in J$. But this would contradict the maximality of j_0. Hence either $(i_4, j_0 + 1) \in HO(\tau)$ or $(j_0 + 1, i_3) \in HO(\tau)$. □ *Claim* 3.1.

If $(i_4, j_0 + 1) \in HO(\tau)$, then $HO(\tau)|_{\{j_0, j_0 + 1, i_3, i_4\}} = (i_3, j_0, i_4, j_0 + 1)$. Hence in that case $\{j_0, j_0 + 1, i_3, i_4\}$ is a right primitive quartet.

If $(j_0 + 1, i_3) \in HO(\tau)$, then $j_0 + 1 \neq i_2$ and

$$HO(\tau)|_{\{j_0, j_0 + 1, i_2, i_3\}} = (j_0 + 1, i_3, j_0, i_2),$$

which implies that $\{j_0, j_0 + 1, i_2, i_3\}$ is a left primitive quartet.

Hence (3) is satisfied.

This completes the proof of Theorem 5.3   □

Theorem 5.3 may be viewed as characterizing alternating texts in terms of "forbidden jumps".

Intuitively speaking, forbidding in a standard text the left quartet I with $HO(\tau|_I) = (2, 4, 1, 3)$ means that when a left jump $(2 \leftarrow 1)$ is made, then the jump from 2 to 3, where 3 appears to the right of 1, is not allowed because there is a "gap" to be filled in between 1 and 2. In other words, one could move from 2 to the right of 1 only if the gap between 1 and 2 is "patched" first. Thus, e.g., $(2, 3, 1, 4)$ would be allowed because 3 would patch the gap between 1 and 2 and the one could move to 4 which is to the right of 1.

Forbidding the right quartet I with $HO(\tau|_I) = (3, 1, 4, 2)$ has the same intuition.

The above intuition leads to the following characterization of alternating texts in terms of jumps.

THEOREM 5.4: *A standard text τ is alternating iff for each $i \in \mathrm{dom}(\tau)$*

(i) *if* $\mathrm{jump}_\tau(i+1) < 0$ *and there is a $j > i+1$ such that $(i, j) \in HO(\tau)$, then* $i+1 \in i(\mathrm{sort}_\tau(j))$, *and*

(ii) *if* $\mathrm{jump}_\tau(i+1) \geq 0$ *and there is a $j > i+1$ such that $(j, i) \in HO(\tau)$, then* $i+1 \in i(\mathrm{sort}_\tau(j))$.

Proof: Suppose that τ is a text that is not alternating. We will show that there exists $i \in \mathrm{dom}(\tau)$ such that either (i) or (ii) is not satisfied. By Theorem 5.3, τ has a primitive quartet I such that if $\rho = VO(\tau)|_I$, then $\{\rho(1), \rho(2)\} \in \mathrm{seg}(VO(\tau))$.

It is sufficient to consider only one of the two symmetric cases. Let $I = \{i_1, i_1 + 1, i_3, i_4\}$ be a left primitive quartet of τ such that $VO(\tau)|_I = (i_1, i_1 + 1, i_3, i_4)$. Then $\mathrm{jump}_\tau(i_1 + 1) < 0$, $i_3 > i_1 + 1$, and $(i_1, i_3) \in HO(\tau)$. Since $i_4 > i_3$,

$i_1 + 1 \notin i_1$ (sort$_\tau (i_3)$). Hence in this case, (i) does not hold. If τ has a right primitive quartet, we find that (ii) does not hold.

Suppose now that there exists $i \in \text{dom} (\tau)$ such that (i) does not hold, *i.e.*, jump$_\tau (i+1) < 0$, and for some $j > i+1$, $(i, j) \in HO(\tau)$ and $i+1 \notin i (\text{sort}_\tau (j))$. Then there exists $k > j$ such that $(i+1, k) \in HO(\tau)$ and $(k, i) \in HO(\tau)$. Now $\{ i, i+1, j, k \}$ is a left primitive quartet of τ as in statement (3) of Theorem 5.3. If there is an $i \in \text{dom} (\tau)$ such that (ii) is not satisfied, then by symmetry there is a right primitive quartet as in (3). Hence, by Theorem 5.3, τ is not alternating. □

Example 5.1: Consider the text τ from Remark 5.1, which is not alternating. For this text, jump$_\tau (2) < 0$, $4 > 2$, $(1, 4) \in HO(\tau)$, but $2 \notin 1 (\text{sort}_\tau (4)) = \{ 1, 3, 4 \}$.

The standard text τ such that $HO(\tau) = (1, 3, 4, 2, 5, 6)$ is an alternating text: e.g., jump$_\tau (3) < 0$, and $(1, 5) \in HO(\tau)$, but 2 and 3 are in the same segment in sort$_\tau (5)$, because the gap between 2 and 3 is "patched" by 4. □

Hence a text that is jump-free is alternating if the elements of its domain that have jump -1 satisfy Condition (i) of Theorem 5.4, and those that have jump 0 satisfy Condition (ii) of Theorem 5.4.

6. TWO CHARACTERIZATIONS OF JUMP-FREE TEXTS

In this section, a combinatorial and an operational characterization of $J_{\{0, -1\}}$ are given, similar to the results obtained for alternating texts (Propositions 3.1 and 3.2). As a by-product of the operational characterization we obtain an alternative proof of the fact that each alternating text is jump-free (Theorem 5.2).

First we give a combinatorial characterization (*cf.* Proposition 3.1 and Theorem 5.3).

THEOREM 6.1: *Let τ be a text. The following statements are equivalent.*

(1) $\tau \notin J_{\{0, -1\}}$;

(2) *there is a primitive quartet I of τ such that $\{ \rho(2), \rho(3) \} \in \text{seg} (VO(\tau))$, where $\rho = VO(\tau) |_I$;*

(3) *there is a primitive quartet I of τ such that $\{ \rho(1), \rho(4) \} \in \text{seg} (HO(\tau))$, where $\rho = VO(\tau) |_I$;*

(4) *there is a primitive quartet I of τ such that $\{ \rho(1), \rho(4) \} \in \text{seg} (HO(\tau))$ and $\{ \rho(2), \rho(3) \} \in \text{seg} (VO(\tau))$, where $\rho = VO(\tau) |_I$.*

Proof: We assume that τ is a standard text.

(1) implies (4): This is obtained by reconsidering the proof of Theorem 5.2. Let $i \in \mathrm{dom}(\tau)$ be such that $\mathrm{jump}_\tau(i) \notin \{0, -1\}$. Let $j, k \in \mathrm{dom}(\tau)$ be as in the proof of Theorem 5.2 (for either of the symmetric cases). Note that in both cases j and k can be chosen such that $\{j, k\} \in \mathrm{seg}(HO(\tau))$. Then $\{j, i-1, i, k\}$ is either a left or a right [depending on the sign of $\mathrm{jump}_\tau(i)$] primitive quartet of the demanded form.

(2) implies (1): It is sufficient to consider the case that I is a left primitive quartet (the other case is symmetric).

Let $I = \{i_1, i_2, i_2+1, i_4\}$ where $VO(\tau)|_I = (i_1, i_2, i_2+1, i_4)$. Then $HO(\tau)|_I = (i_2, i_4, i_1, i_2+1)$.

Consider $\mathrm{sort}_\tau(i_2)$. Since $(i_2, i_4), (i_4, i_1) \in HO(\tau)$, $i_1 \notin i_2(\mathrm{sort}_\tau(i_2))$. Hence $\mathrm{jump}_\tau(i_2+1) > 0$. Consequently $\tau \notin J_{\{0, -1\}}$.

(3) implies (1): Again we restrict ourselves to the case that I is a left primitive quartet. The proof resembles the proof of Theorem 5.3.

Let $I = \{i_1, i_2, i_3, i_4\}$ where $VO(\tau)|_I = (i_1, i_2, i_3, i_4)$. Then $HO(\tau)|_I = (i_2, i_4, i_1, i_3)$, and $\{i_1, i_4\} \in \mathrm{seg}(HO(\tau))$.

Let $j_0 = \max\{j \in \mathrm{dom}(\tau) \mid i_2 \leq j < i_3 \text{ and } (j, i_4) \in HO(\tau)\}$. Then $(i_4, j_0+1) \in HO(\tau)$, which includes the case that $j_0+1 = i_3$. Since $\{i_1, i_4\} \in \mathrm{seg}(HO(\tau))$, it follows that $HO(\tau)|_{\{i_1, j_0, j_0+1, i_4\}} = (j_0, i_4, i_1, j_0+1)$, and hence $\mathrm{jump}_\tau(j_0) > 0$. Consequently $\tau \notin J_{\{0, -1\}}$.

(4) implies (2) and (3): Obvious.

This completes the proof of the theorem. □

Example 6.1: (1) Consider the text τ from Remark 5.1. We have already observed that τ is not alternating, since τ has a primitive quartet $(1, 2, 4, 5)$. Note that this is the only primitive quartet of τ. Now, since $(1, 5) \notin \mathrm{seg}(HO(\tau))$, it follows directly from Theorem 6.1 that τ is jump-free.

(2) Let τ be a standard text such that $HO(\tau) = (2, 5, 4, 1, 3)$. Then τ is not jump-free, because $\{1, 2, 3, 4\}$ is a primitive quartet, and $\{1, 4\} \in \mathrm{seg}(HO(\tau))$. □

We establish some closure properties of $J_{\{0, -1\}}$ before giving an operational characterization of this class.

THEOREM 6.2: $J_{\{0, -1\}}$ *is closed under the operations $V\mathrm{rev}$, $H\mathrm{rev}$ and $\oplus$.*

Proof: It follows directly from Theorem 4.3 that $J_{\{0, -1\}}$ is closed under $V\mathrm{rev}$ and $H\mathrm{rev}$. Let τ_1 and τ_2 be disjoint jump-free texts, and let $\tau = \tau_1 \oplus \tau_2$. It is easy to verify that for each $i \in \mathrm{dom}(\tau)$, $\mathrm{jump}_\tau(i) = \mathrm{jump}_{\tau_j}(i)$ (*see* also

Lemma 6.2), where $j \in \{1, 2\}$ is such that $i \in \mathrm{dom}(\tau_j)$. Hence τ is jump-free, and $J_{\{0, -1\}}$ is closed under $\oplus$.   □

Note that Theorem 6.2 and Proposition 1.2 give another proof of Theorem 5.2.

In [3] (and in a slightly different way in [4]) the operation flip was introduced, which exchanges the two linear orders of a text. Hence for a text $\tau = (\lambda, \rho_1, \rho_2)$, $\mathrm{flip}(\tau) = (\lambda, \rho_2, \rho_1)$. Note that $\mathrm{flip} \circ V\mathrm{rev} = H\mathrm{rev} \circ \mathrm{flip}$.

We prove now that $J_{\{0, -1\}}$ is also closed under flip.

THEOREM 6.3: $J_{\{0, -1\}}$ *is closed under flip.*

Proof: Let $\tau \in J_{\{0, -1\}}$ be a text, and let $\tau' = \mathrm{flip}(\tau)$.

Assume to the contrary that $\tau' \notin J_{\{0, -1\}}$. By Theorem 6.1, there exists a primitive quartet $I = \{i_1, i_2, i_3, i_4\}$ such that $\{i_1, i_4\} \in \mathrm{seg}(HO(\tau'))$, $\{i_2, i_3\} \in \mathrm{seg}(VO(\tau'))$, and $VO(\tau')|_{\{i_1, i_2, i_3, i_4\}} = (i_1, i_2, i_3, i_4)$.

Since τ results from τ' by swapping the linear orders, it follows that I is a primitive quartet of τ with $\{\rho(2), \rho(3)\} = \{i_1, i_4\} \in \mathrm{seg}(VO(\tau))$ and $\{\rho(1), \rho(4)\} = \{i_2, i_3\} \in \mathrm{seg}(HO(\tau))$, where $\rho = VO(\tau)|_I$. By Theorem 6.1, this implies that $\tau \notin J_{\{0, 1\}}$; a contradiction.

Hence $\tau' = \mathrm{flip}(\tau) \in J_{\{0, -1\}}$.   □

Example 6.2: Let τ be the text from Example 6.1(2) which is not jump-free. Let $\tau' = \mathrm{flip}(\tau) = (\mathrm{fun}_\tau, (2, 5, 4, 1, 3), (1, 2, 3, 4, 5))$. Then in $\mathrm{sort}_{\tau'}(3)$, 2, 5, and 4 have been sorted, and $\mathrm{sort}_{\tau'}(3) = (\{2\}, \{4, 5\})$. Hence $\mathrm{jump}_{\tau'}(1) = -2$. Consequently, $\tau' = \mathrm{flip}(\tau)$ is not jump-free.   □

The shape of a text gives a decomposition of the text in sequential and primitive texts. In the case of alternating texts, one obtains a decomposition in sequential texts only. As a consequence, Proposition 3.2 describes an alternating text as the composition of singleton texts using the operations $V\mathrm{rev}$ (or $H\mathrm{rev}$) and $\oplus$.

Although $J_{\{0, -1\}}$ is closed under $V\mathrm{rev}$, $H\mathrm{rev}$, $\oplus$, (and flip) we do not have an analogous characterization of jump-free texts in terms of these operations, This is because primitive texts (of length more than 2) can not be obtained from smaller texts using the sum operation; the composition of non-alternating texts requires a separate compositional operation for each primitive text.

To deal with this, we consider an alternative way of composing arbitrary texts. We define a substitution operation on texts as follows.

DEFINITION 6.1: Let τ be a standard text of length n with $HO(\tau) = (b_1, \ldots, b_n)$, and let $(\tau_1, \ldots, \tau_n)$ be a sequence of disjoint texts. Then the *substitution of* $(\tau_1, \ldots, \tau_n)$ *into* τ, denoted $[\tau \leftarrow (\tau_1, \ldots, \tau_n)]$, is the text $(\lambda, \rho_1, \rho_2)$, with $\rho_1 = VO(\tau_1) + \ldots + VO(\tau_n)$, $\rho_2 = HO(\tau_{b_1}) + \ldots + HO(\tau_{b_n})$, and $\lambda = \bigcup\limits_{i=1}^{n} \mathrm{fun}_{\tau_i}$.

Example 4.3: Let τ be a standard text such that $HO(\tau) = (2, 4, 1, 3)$. Let $\tau_1 = (\mathrm{fun}_{\tau_1}, (1, 2, 3), (3, 2, 1))$, $\tau_2 = (\mathrm{fun}_{\tau_2}, (4, 5), (4, 5))$, $\tau_3 = (\mathrm{fun}_{\tau_3}, (6), (6))$, and $\tau_4 = \mathrm{fun}_{\tau_4}, (7), (7))$. Then

$$[\tau \leftarrow (\tau_1, \ldots, \tau_4)] = (\lambda, 1, 2, 3, 4, 5, 6, 7), (4, 5, 7, 3, 2, 1, 6)),$$

where $\lambda = \bigcup\limits_{i=1}^{4} \mathrm{fun}_{\tau_i}$. □

Now indeed, the substitution operation can be used to describe a text as the composition of singleton texts according to its shape (*see* Theorem 6.4). The reason for this is that substitution of texts amounts to inverting the quotient construction on the corresponding T-functions. This connection is given by the following fact, which is obtained by immediate verification of the definitions.

FACT 6.1: Let $\tau, \tau_1, \ldots, \tau_n$ be texts, and let τ' be a standard text of length n. Then

$$\tau = [\tau' \leftarrow (\tau_1, \ldots, \tau_n)] \quad \text{iff} \quad Tf(\tau') = Tf(\tau)/\{\mathrm{dom}(\tau_1), \ldots, \mathrm{dom}(\tau_n)\}$$

(modulo isomorphisms). □

Note that the sum operation is a special case of substitution: $\oplus$ is the substitution of alternating texts into a forward sequential text. From [3] (*cf.* Proposition 3.2) we know that **ALT** is closed under $\oplus$; we now show that **ALT** is closed under general substitution.

LEMMA 6.1: **ALT** *is closed under substitution.*

Proof: Let τ be a standard alternating text of length n, let $\tau_1, \ldots, \tau_n$ be disjoint alternating texts, and let $\tau' = [\tau \leftarrow (\tau_1, \ldots, \tau_n)]$. Assume to the contrary that τ' is not alternating. Then, by Proposition 3.1, τ' has a primitive quartet $I = \{i_1, i_2, i_3, i_4\}$. Since for each $j \in \{1, \ldots, n\}$ $\mathrm{dom}(\tau_j)$ is a segment of $HO(\tau')$, it follows that either there exists $j \in \{1, \ldots, n\}$ such that $\{i_1, \ldots, i_4\} \subseteq \mathrm{dom}(\tau_j)$ or there exist $j_1, j_2, j_3, j_4 \in \{1, \ldots, n\}$ mutually different such that $i_k \in \mathrm{dom}(\tau_{j_k})$ for $k = 1, \ldots, 4$. Consequently, by Definition

6.1, either $\{i_2, \ldots, i_4\}$ is a primitive quartet of τ_j, or $\{j_1, j_2, j_3, j_4\}$ is a primitive quartet of τ, which contradicts the fact that both τ_j and τ are alternating.

Hence τ' is alternating, and **ALT** is closed under substitution. $\square$

A *primitive pair* is a text τ with $|\tau| = 2$. Note that each primitive pair is a primitive and sequential text, and that each sequential text can be obtained from primitive pairs by substitution.

For a class of texts K, K^S denotes the smallest class of texts containing K that is closed under substitution.

THEOREM 6.4: *The class of all texts* **TXT** *equals* **PRIM**^{S}.

Proof: Clearly, **TXT** $\supseteq$ **PRIM**^{S}.

In order to show that **TXT** $\subseteq$ **PRIM**^{S}, let τ be a text with $|\tau| = n$. We will prove by induction on n that $\tau \in$ **PRIM**^{S}.

If $n = 1$, then τ is a singleton text, which is primitive. Now suppose that $n > 1$, and that all texts of length less than n are in **PRIM**^{S}.

Consider *shape* (τ). Let g be the lTs that labels the root of the shape. Then g is linear or primitive. Hence the text τ' for which $g = Tf(\tau')$ is sequential or primitive. In the former case τ' can be obtained from primitive pairs by substitution, and hence in both cases $\tau' \in$ **PRIM**^{S}. Let $X_1, \ldots, X_S$ be such that $\mathrm{dom}(g) = \{X_1, \ldots, X_S\}$ and $VO(\tau') = (X_1, \ldots, X_S)$. Let $\tau_i = \tau|_{X_i}$ for $i = 1, \ldots, s$. Since $|\tau_i| < n$ for $i = 1, \ldots, s$, it follows from the inductive assumption that $\tau_1, \ldots, \tau_S$ are all in **PRIM**^{S}. By the definition of the shape, $Tf(\tau') = g = Tf(\tau)/\{X_1, \ldots, X_s\}$. Hence, by Fact 6.1, $[\tau' \leftarrow (\tau_1, \ldots, \tau_S)] = \tau$. Hence $\tau \in$ **PRIM**^{S}. This completes the induction step in this proof.

Consequently, **TXT** $\subseteq$ **PRIM**^{S}.

Thus, **TXT** = **PRIM**^{S}. $\square$

In particular, alternating texts can be characterized by substitution of alternating primitive texts. It should be clear from the proof of Theorem 6.4 that **ALT** $\subseteq$ (**PRIM** $\cap$ **ALT**)^S; then it follows by Lemma 6.1 that (**PRIM** $\cap$ **ALT**)^S = **ALT**. Note that **PRIM** $\cap$ **ALT** consists of all singleton texts and all primitive pairs, and that the substitution of texts τ_1 and τ_2 into a primitive pair corresponds to either $\tau_1 \oplus \tau_2$ or

$$H \operatorname{rev} (H \operatorname{rev}(\tau_1) \oplus H \operatorname{rev}(\tau_2)) = V \operatorname{rev}(V \operatorname{rev}(\tau_2) \oplus V \operatorname{rev}(\tau_1)).$$

Hence in fact we have restated Proposition 3.2 in terms of substitution.

For jump-free texts we now have an analogous characterization in terms of substitution. First we prove that $J_{(0, -1)}$ is closed under substitution.

It is possible, although tedious, to characterize the substitution operation in terms of jump and type, as in Theorem 4.3. For our purpose here it suffices to construct the jump function of a text that is the result of substitution.

Recall that for a text τ, $\overline{\mathrm{dom}}\,(\tau) = \mathrm{dom}\,(\tau) - \{\,VO\,(\tau)\,(1)\,\}$.

LEMMA 6.2: *Let τ be a standard text of length n and let $\tau_1, \ldots, \tau_n$ be disjoint texts. Let $\tau' = [\tau \leftarrow (\tau_1, \ldots, \tau_n)]$. Then for each $i \in \mathrm{dom}\,(\tau')$, if $k \in \{\,1, \ldots, n\,\}$ is such that $i \in \mathrm{dom}\,(\tau_k)$, then*

$$\mathrm{jump}_{\tau'}(i) = \begin{cases} \mathrm{jump}_{\tau_k}(i) & \text{if} \quad i \in \overline{\mathrm{dom}}\,(\tau_k) \\ \mathrm{jump}_{\tau}(k) & \text{otherwise.} \end{cases}$$

Proof: We may assume that τ' is standard. Let $i \in \mathrm{dom}\,(\tau')$, and let k be such that $i \in \mathrm{dom}\,(\tau_k)$. Note that $\mathrm{dom}\,(\tau_k)$ is a segment of $HO\,(\tau')$. Hence there is an $l \geq 0$ such that for each $j \in \mathrm{dom}\,(\tau_k)$, $\mathrm{pos}_{\tau'}(j) = l + \mathrm{pos}_{\tau_k}(j)$, and

$\mathrm{ind}_{\tau'}(j) = l + \mathrm{ind}_{\tau_k}(j)$. Thus, for each $i \in \overline{\mathrm{dom}}\,(\tau_k)$,

$$\mathrm{jump}_{\tau'}(i) = \mathrm{ind}_{\tau'}(i) - \mathrm{pos}_{\tau'}(i-1)$$
$$= l + \mathrm{ind}_{\tau_k}(i) - (l + \mathrm{pos}_{\tau_k}(i-1)) = \mathrm{jump}_{\tau_k}(i).$$

If $i = VO\,(\tau_1)\,(1) = 1$, then $\mathrm{jump}_{\tau'}(i) = 0 = \mathrm{jump}_{\tau}(1)$. If $i = VO\,(\tau_k)\,(1)$, $k \geq 2$, then $i - 1 \in \tau_{k-1}$. Note that in the latter case, $\mathrm{ind}_{\tau'}(i) = \mathrm{ind}_{\tau}(k)$ and $\mathrm{pos}_{\tau'}(i-1) = \mathrm{pos}_{\tau}(k-1)$. Hence

$$\mathrm{jump}_{\tau'}(i) = \mathrm{ind}_{\tau'}(i) - \mathrm{pos}_{\tau'}(i-1)$$
$$= \mathrm{ind}_{\tau}(k) - \mathrm{pos}_{\tau}(k-1) = \mathrm{jump}_{\tau}(k). \quad \square$$

Example 6.4: Let $\tau, \tau_1, \tau_2, \tau_3, \tau_4$ be the texts from Example 6.3, and let $\tau' = [\tau \leftarrow (\tau_1, \ldots, \tau_4)]$. Then $\mathrm{jump}_{\tau'}$ is as follows.

	1	2	3	4	5	6	7
$\mathrm{jump}_{\tau'}$ ..	$\underline{0}$	-1	-1	$-\underline{1}$	0	$\underline{1}$	$-\underline{1}$

The underlined numbers are the jumps of τ, the other ones are jumps of τ_1 and τ_2.

COROLLARY 6.1: : $J_{\{0, -1\}}$ *is closed under substitution.*　□

Now we are able to give an operational characterization of $J_{\{0, -1\}}$.

THEOREM 6.5: *The class* $J_{\{0, -1\}}$ *equals* $(\text{PRIM} \cap J_{\{0, -1\}})^S$.

Proof: By Corollary 6.1, $J_{\{0, -1\}}$ is closed under substitution. Hence $(\text{PRIM} \cap J_{\{0, -1\}})^S \subseteqq J_{\{0, -1\}}$. Since subtexts (induced by clans) and quotients of jump-free texts are jumpfree, the proof of Theorem 6.4 applies, and hence $J_{\{0, -1\}} \subseteqq (\text{PRIM} \cap J_{\{0, -1\}})^S$.　□

Example 6.5: Let τ be the standard text such that $HO(\tau) = (2, 3, 4, 9, 5, 1, 7, 8, 6)$ and $\text{fun}_\tau(VO(\tau)) = (a, a, b, c, b, a, a, b, b)$. Then shape$(\tau)$ is as in figure 7.

Hence $\tau = [\tau' \leftarrow (\tau_1, \ldots, \tau_5)]$, with $\tau', \tau_1, \ldots, \tau_5$ as follows. τ' is a primitive jump-free standard text such that $HO(\tau') = (2, 5, 3, 1, 4)$; τ_1, τ_3, τ_5 are the singleton texts such that $\text{dom}(\tau_1) = \{1\}$, $\text{fun}_{\tau_1}(1) = a$, $\text{dom}(\tau_3) = \{5\}$, $\text{fun}_{\tau_3}(5) = b$, and $\text{dom}(\tau_5) = \{9\}$, $\text{fun}_{\tau_5}(9) = b$;

$\tau_2 = [\tau'' \leftarrow (\tau_6, \tau_7)]$, where τ'' is a forward primitive pair, τ_6 is the singleton text with $\text{dom}(\tau_6) = \{2\}$, $\text{fun}_{\tau_6}(2) = a$, and τ_7 is a forward primitive pair with $\text{dom}(\tau_7) = \{3, 4\}$, $\text{fun}_{\tau_7}(VO(\tau_7)) = (b, c)$;

$\tau_4 = [\tau''' \leftarrow (\tau_8, \tau_9)]$, where τ''' is a backward primitive pair, τ_8 is the singleton text with $\text{dom}(\tau_8) = \{6\}$, $\text{fun}_{\tau_8}(6) = a$, and τ_7 is a forward primitive pair with $\text{dom}(\tau_9) = \{7, 8\}$, $\text{fun}_{\tau_9}(VO(\tau_9)) = (a, b)$.

Hence τ is obtained from a primitive jump-free text by substituting primitive pairs and singleton texts.　□

DISCUSSION

In this paper we have investigated sorting sequences of texts. A sorting sequence describes step-by-step how the first linear order of a text is sorted according to the second linear order. We have defined some parameters describing the situation at each step, such as the positions of the last sorted element and the next element, the distance between these positions (given by the parameter jump), and the way the next element is filled in (given by the parameter type). It is proved that each (standard) text τ is uniquely determined by its labeling function fun_τ and the parameters jump_τ and type_τ.
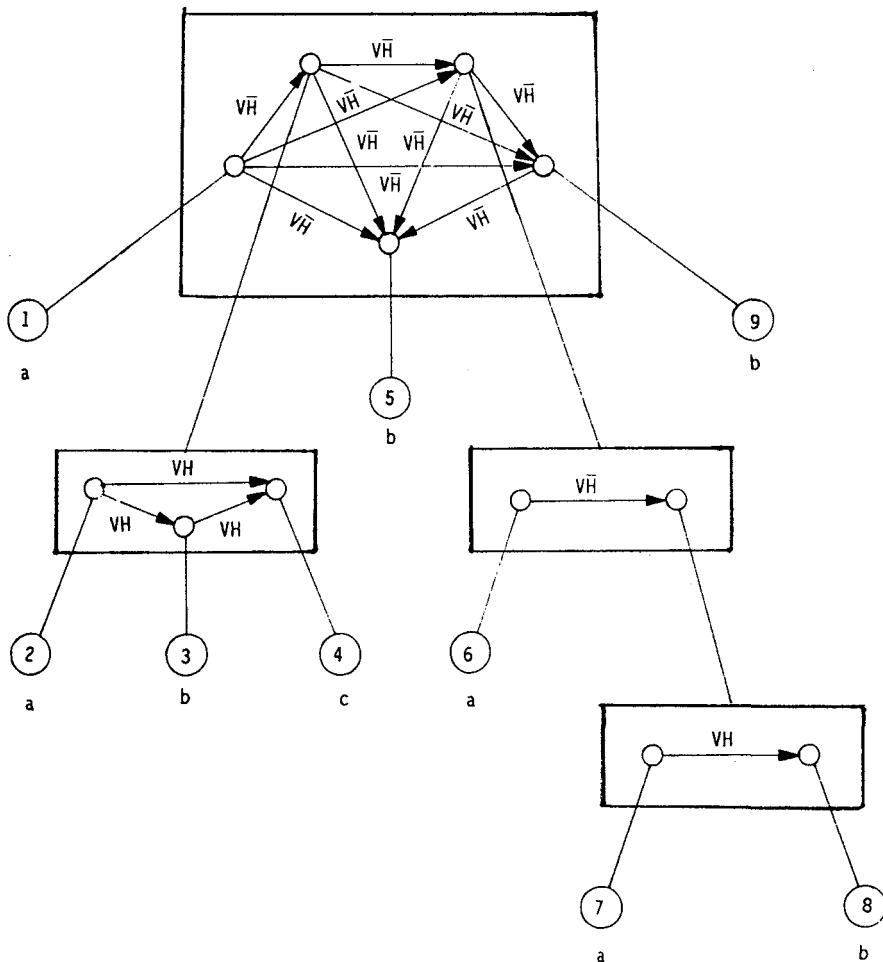
Figure 7.

This yields an alternative way to represent standard jump-free texts: a text $(\lambda, (1, \ldots, n), \rho_2)$ can be represented by the ordered pair (λ, δ), where δ is a function from $\{1, \ldots, n\}$ to an alphabet Δ with $|\Delta| = 8$ such that the letters of Δ correspond with the possible values of jump (0 or -1) and type (S, L, R or B).

In [3] it is shown that alternating texts can be characterized combinatorially using the notion of "primitive quartet", or operationally in terms of the sum and reverse operations. Here we show that the class of jump-free texts can be characterized by forbidding special types of primitive quartets. This characterization illustrates the connection between jump-free and alternating

texts. It could be interesting to consider texts where other types of primitive quartets are forbidden, and see how these relate to jump-free texts and alternating texts.

It is also proved that each jump-free text is obtained from primitive jump-free texts by the substitution operation. However, we still do not know how primitive jump-free texts look like. Clearly, the jump-freeness is a serious restriction, e.g., there is no jump-free primitive text of length 6. There exist jump-free texts on 7 and 8 elements, e.g.,

$$(\lambda, (1, 2, 3, 4, 5, 6, 7), (2, 4, 7, 5, 3, 1, 6))$$
$$(\lambda, (1, 2, 3, 4, 5, 6, 7, 8), (7, 1, 5, 2, 4, 6, 8, 3)).$$

In constructing primitive jump-free texts, one has two constraints: because of the primitivity it is not allowed to construct segments of subsequent elements, and because of the jump-freeness subsequent elements cannot be "too far apart". It is our opinion, that these somewhat contradictory constraints make this problem quite challenging.

Finally, we would like to point out that the connection between the representation of texts by sorting sequences considered here, and the representation of texts by matrices considered in [4] should be investigated.

REFERENCES

1. A. EHRENFEUCHT and G. ROZENBERG, Theory of 2-structures, Part II: representation through labeled tree families, *Theoret. Comput. Sci.*, 1990, *70*, pp. 305-342.
2. A. EHRENFEUCHT and G. ROZENBERG, Angular 2-structures, *Theoret. Comput. Sci.*, 1992, *92*, pp. 227-248.
3. A. EHRENFEUCHT and G. ROZENBERG, *T*-structures, *T*-functions, and texts, *Theoret. Comput. Sci.* (to appear).
4. A. EHRENFEUCHT, T. HARJU and G. ROZENBERG, Matrix oriented texts, Leiden University Techn. Rep. 91-24.