# On well quasi orders on languages

Flavio D'Alessandro

Stefano Varricchio

September 18, 2003

Let $S$ be a set. A *quasi order* on $S$ is a binary relation $\leq$ satisfying the reflexive and transitive property.

If $\leq$ is a quasi order on $S$, the equivalence relation associated with $\leq$ is the relation $\sim = \leq \cap \leq^{-1}$, i.e.

$$a \sim b \Leftrightarrow a \leq b \text{ and } b \leq a.$$

**Definition 1** *A quasi-order in $S$ is called a* well quasi-order *(*wqo*) if every nonempty subset $X$ of $S$ has at least one minimal element in $X$ but no more than a finite number of (non-equivalent) minimal elements.*

1. $\leq$ is a well quasi-order.

2. the *ascending chain condition* holds for the closed subsets of $S$.

3. every infinite sequence of elements of $S$ has an infinite ascending subsequence.

4. if $s_1, s_2, \ldots, s_n, \ldots$ is an infinite sequence of elements of $S$, then there exist integers $i, j$ such that $1 \leq i < j$ and $s_i \leq s_j$.

5. there exists neither an infinite strictly descending sequence in $S$ (i.e., $\leq$ is well-founded), nor an infinity of mutually incomparable elements of $S$.

6. $S$ has the *finite basis property*, i.e., every closed subset of $S$ is finitely generated.

**Definition 2** *A quasi-order* $\leq$ *in a semigroup* $S$ *is* monotone *if for all* $x_1, x_2, x_3, x_4 \in S$ *if* $x_1 \leq x_2$ *and* $x_3 \leq x_4$*, then* $x_1 x_3 \leq x_2 x_4$*.*

**Definition 3** *A quasi-order* $\leq$ *in* $S$ *is a* divisibility*, or* division*,* order *if it is monotone and, moreover, for all* $s \in S$ *and* $x, y \in S^1$*,* $s \leq xsy$*.*

**Theorem 1** *(Higman) Let* $S$ *be a semigroup quasi-ordered by a divisibility order* $\leq$*. If there exists a generating set of* $S$ *well quasi-ordered by* $\leq$*, then* $S$ *will also be so.*

**Theorem 2** *Let* $A^*$ *be the free semigroup over a finite alphabet* $A$ *and let* $\leq$ *be the subsequence ordering. Then* $\leq$ *is a wqo on* $A^*$*.*

A semi-Thue system $(A, \pi)$ is called *unitary* when $\pi$ is a finite set of productions of the kind

$$\epsilon \to u,\ u \in I \subseteq A^{+}.$$

Let $\Rightarrow_I^*$ the derivation relation in these systems. If $I = A$, then $\Rightarrow_A^*$ is equal to the subsequence ordering $\leq$ which is a divisibility ordering. Thus, $\Rightarrow_A^*$ is a wqo by the Higman theorem.

**Definition 4** *Let $I \subseteq A^{+}$. We say that $I$ is* subword unavoidable *if there exists a positive integer $k_0$, such that any word $u \in A^*$, $|u| > k_0$, can be written as $u = xwy$, where $w \in I$ and $x, y \in A^*$.*

**Theorem 3** *(Ehrenfeucht, Haussler, Rozenberg, 1983) The relation $\Rightarrow_I^*$ of the unitary semi-Thue system associated with the finite set $I \subseteq A^{+}$ is a wqo if and only if $I$ is subword unavoidable.*

Let $L_I^\epsilon = \{w \in A^* \mid \epsilon \Rightarrow_I^* w\}$.

In general $\Rightarrow_I^*$ is not a wqo on $L_I^\epsilon$. In fact let

$$A = \{a, b, c\}$$

and

$$I = \{ab, c\}$$

Then the sequence

$$acb, aacbb, aaacbbb, \ldots, a^n cb^n \ldots$$

is a bad sequence with respect to $\Rightarrow_I^*$.

Let $I$ be a finite subset of $A^*$. We define the binary relation $\vdash_I$ as follows: we say that $v \vdash_I w$ if

$$v = v_1 v_2 \cdots v_{n+1},$$

$$w = v_1 a_1 v_2 a_2 \cdots v_n a_n v_{n+1},$$

with $a_i \in A$, and $a_1 a_2 \cdots a_n \in I$.

The relation $\vdash_I^*$ is the transitive and reflexive closure of $\vdash_I$.

**Definition 5** *Let $I \subseteq A^*$. We say that $I$ is* subsequence unavoidable *if there exists a positive integer $k_0$, such that any word $u \in A^*$, $|u| > k_0$, can be written as $v_1 a_1 v_2 a_2 \cdots v_n a_n v_{n+1}$ where $a_1 a_2 \cdots a_n \in I$, $a_i \in A$, and $v_1, v_2, \ldots v_{n+1} \in A^*$.*

**Theorem 4** *( Haussler 1983) The relation $\vdash_I^*$ associated with the finite set $I \subseteq A^+$ is a wqo if and only if $I$ is subsequence unavoidable.*

Let $L^{\epsilon}_{\vdash_I} = \{w \in A^* \mid \epsilon \vdash_I^* w\}$.

In general $\vdash_I^*$ is not a wqo on $L^{\epsilon}_{\vdash_I}$. Let

$$A = \{a, \bar{a}, b, \bar{b}, c, \bar{c}, d, \bar{d}\}$$

$$I = \{a\bar{a}, b\bar{b}, c\bar{c}, d\bar{d}\}$$

Let $u_1, u_2, \ldots, u_n, \ldots$ be a sequence of words of $\{b, c\}^*$, with the property that $u_i$ is not a factor of $u_j$ for $i \neq j$.

Let $v_1, v_2, \ldots, v_n, \ldots$ be the sequence of words of $\{b, \bar{b}, c, \bar{c}\}^*$ where $v_i$ is obtained from $u_i$ with the substitution $b \to b\bar{b}$, $c \to c\bar{c}$. For instance if $u_i = cbbc$ then $v_i = c\bar{c}b\bar{b}b\bar{b}c\bar{c}$.

Finally we construct a sequence $w_1, w_2, \ldots, w_n, \ldots$ where each $w_i$ is obtained from $v_i$ as follows.

If, for instance, $v_i = c\bar{c}b\bar{b}b\bar{b}c\bar{c}b\bar{b}c\bar{c}$ then

$$w_i = ac\bar{c}db\bar{b}\bar{a}ab\bar{b}\bar{d}dc\bar{c}\bar{a}ab\bar{b}\bar{d}c\bar{c}\bar{a}$$

**Proposition 1** *The sequence* $w_1, w_2, \ldots, w_n, \ldots$ *is a bad sequence with respect to* $\vdash_I^*$. *Therefore,* $\vdash_I^*$ *is not a wqo on* $L^{\epsilon}_{\vdash_I}$.

The following theorem holds:

**Theorem 5** *For any finite set* $I$ *the relation* $\vdash_I^*$ *is a well quasi order on* $L_I^{\epsilon}$.

Let $T$ be a set of tuples of words. The elements of $T$ are of the kind $(u_1, u_2, \ldots, u_n)$, with $u_i \in A^+$, and $n \geq 1$.

We define the binary relation $\vdash_T$ as follows: we say that $v \vdash_T w$ if $v = v_1 v_2 \cdots v_{n+1}$,

$$w = v_1 u_1 v_2 u_2 \cdots v_n u_n v_{n+1},$$

and $(u_1, u_2 \ldots u_n) \in T$.

The relation $\vdash_T^*$ is the transitive and reflexive closure of $\vdash_T$.

**Definition 6** *Let $T$ be a set of tuples of elements of $A^*$. We say that $T$ is unavoidable if there exists a positive integer $k_0$, such that any word $u \in A^*$, $|u| > k_0$, can be written as $v_1 u_1 v_2 u_2 \cdots v_n u_n v_{n+1}$ where $(u_1, u_2 \ldots u_n) \in T$, and $v_1, v_2, \ldots v_{n+1} \in A^*$.*

**Conjecture 1** *( Haussler 1983) The relation* $\vdash_T^*$ *associated with the finite set* $T$ *of tuples is a wqo if and only if* $T$ *is unavoidable.*

Let $I$ be a finite set of words. We consider the set

$$\bar{I} = \{(u, v) \mid u, v \in A^+ \text{ and } uv \in I\} \cup I$$

The following theorem holds:

**Theorem 6** *For any finite set of words $I$ the relation $\vdash_{\bar{I}}^*$ is a wqo on $L_I^\epsilon$.*

**Definition 7** *We say that $\leq$ is a division order on L, if $\leq$ is monotone and the following conditions hold:*

$$u \leq xuy \text{ for any } u \in L, x, y \in A^* \text{ with } xuy \in L,$$

$$1 \leq u \text{ for any } u \in L.$$

Let $G = (V, A, P, S)$ be a context-free grammar.

Let $V = \{X_1, X_2, \ldots, X_n\}$. Let $L_i$ be the language of the words generated setting $X_i$ as start symbol and let $L = \cup_{i=1}^n L_i$. The following theorem holds:

**Theorem 7** *If $\leq$ is a division order on $L$, then $\leq$ is a well quasi order on $L$.*

Let $I \subseteq A^+$.

1. The language $L_I^\epsilon$ is a context-free language generated by a grammar having only one variable.

2. The language $L_{\vdash_I}^\epsilon$ is not, in general, context-free.

3. The derivation relation $\vdash_I^*$ is a division order on $L_I^\epsilon$ and $L_{\vdash_I}^\epsilon$.

4. The derivation relation $\Rightarrow_I^*$ in general is not a division order neither on $L_I^\epsilon$ nor on $L_{\vdash_I}^\epsilon$.

**Theorem 8** *(Ehrenfeucht, Haussler, Rozenberg, 1983) Let $L \subseteq A^*$. $L$ is regular if and only if it is upwards closed with respect to a monotone well quasi order of $A^*$.*

Let $I \subseteq A^+$. The following conditions are equivalent:

1. The language $L_I^\epsilon$ is a regular language.

2. The derivation relation $\Rightarrow_I^*$ is a well quasi order on $A^*$.

3. $I$ is subword unavoidable.

**Problem:** Characterize the subsets $I$ such that $\Rightarrow_I^*$ is a well quasi order on $L_I^\epsilon$.

**Conjecture 2** *The derivation relation is a well quasi order on* $L_I^\epsilon$ *if and only if it is a well quasi order on* $A^*$.

The above conjecture holds if the following statement is true:

If $I$ is avoidable, then $\Rightarrow_I^*$ is not a well quasi order on $L_I^\epsilon$.