

# Well-quasi-orderings on word languages

main

81b14ea82bdf369ba5dbe8de29fa418756900125

2025-10-15 17:43:32 +0200

Anonymized for review

Anonymized for review

**Abstract.** The set of finite words over a well-quasi-ordered set is itself well-quasi-ordered. This seminal result by Higman is a cornerstone of the theory of well-quasi-orderings and has found numerous applications in computer science. However, this result is based on a specific choice of ordering on words, the (scattered) subword ordering. In this paper, we describe to what extent other natural orderings (prefix, suffix, and infix) on words can be used to derive Higman-like theorems. More specifically, we are interested in characterizing *languages* of words that are well-quasi-ordered under these orderings, and explore their properties and connections with other language theoretic notions. We furthermore give decision procedures when the languages are given by various computational models such as automata, context-free grammars, and automatic structures.

## 1 Introduction

A *well-quasi-ordered* set is a set  $X$  equipped with a quasi-order  $\preceq$  such that every infinite sequence  $(x_n)_{n \in \mathbb{N}}$  of elements taken in  $X$  contains an increasing pair  $x_i \preceq x_j$  with  $i < j$ . Well-quasi-orderings serve as a core combinatorial tool powering many termination arguments, and was successfully applied to the verification of infinite state transition systems [?,?]. One of the appealing properties of well-quasi-orderings is that they are closed under many operations, such as taking products, finite unions, and finite powerset constructions [?]. Perhaps more surprisingly, the class of well-quasi-ordered sets is also stable under the operation of taking finite words and finite trees labelled by elements of a well-quasi-ordered set [?,?].

Note that in the case of finite words and finite trees, the precise choice of ordering is crucial to ensure that the resulting structure is well-quasi-ordered. The celebrated result of Higman states that the set of finite words over an ordered alphabet  $(X, \preceq)$  is well-quasi-ordered by the so-called subword embedding relation [?]. Let us recall that the subword relation for words over  $(X, \preceq)$  is defined as follows: a word  $u$  is a *subword* of a word  $v$ , written  $u \leq^* v$ , if there exists an increasing function  $f: \{1, \dots, |u|\} \rightarrow \{1, \dots, |v|\}$  such that  $u_i \preceq v_{f(i)}$  for all  $i \in \{1, \dots, |u|\}$ .

However, there are many other natural orderings on words that could be considered in the context of well-quasi-orderings, even in the simplified setting of a finite alphabet  $\Sigma$  equipped with the equality relation. In this setting, the three alternatives we consider are the *prefix relation* ( $u \sqsubseteq_{\text{pref}} v$  if there exists  $w$  with  $uw = v$ ), the *suffix relation* ( $u \sqsubseteq_{\text{suff}} v$  if there exists  $w$  such that  $wu = v$ ), and the *infix relation* ( $u \sqsubseteq_{\text{infix}} v$  if there exists  $w_1, w_2$  such that  $w_1uw_2 = v$ ). Note that these three relations straightforwardly generalize to infinite quasi-ordered alphabets. Unfortunately, it is easy to see that none of these constructions are well-quasi-ordered as soon as the alphabet contains two distinct letters: for instance, the infinite sequence  $ab^na$  is well-quasi-ordered by the subword relation but by neither the prefix relation, nor the suffix relation, nor the infix relation.

While this dooms well-quasi-orderedness of these relations in the general case, there may be *subsets* of  $\Sigma^*$  which are well-quasi-ordered by these relations. As a simple example, take the case of finite sets of (finite) words which are all well-quasi-ordered regardless of the ordering considered. This raises the question of characterizing exactly which subsets  $L \subseteq \Sigma^*$  are well-quasi-ordered with respect to the prefix relation (respectively, the suffix relation or the infix relation), and designing suitable decision procedures.

Let us argue that these decision procedures fit a larger picture in the research area of well-quasi-orderings. Indeed, there has been recent breakthroughs in deciding whether a given order is a well-quasi-order, for instance in the context of the verification of infinite state transition systems [?] or in the context of logic [?]. In the graph theory community, recent works have studied classes of graphs that are well-quasi-ordered by the induced subgraph relation using similar language theoretic techniques [?, ?, ?]. Furthermore, a previous work by Kuske shows that any *reasonable*<sup>1</sup> partially ordered set  $(X, \leq)$  can be embedded into  $\{a, b\}^*$  with the infix relation [?, Lemma 5.1]. Phrased differently, one can encode a large class of partially ordered sets as subsets of  $\{a, b\}^*$ . As a consequence, the following decision problem provides a reasonable abstract framework for deciding whether a given partially ordered set is well-quasi-ordered: given a language  $L \subseteq \Sigma^*$ , decide whether  $L$  is well-quasi-ordered by the infix relation.

The runtime of an algorithm based on well-quasi-orderings is deeply related to the “complexity” of the underlying quasi-order [?]. One way to measure this complexity is to consider its so-called ordinal invariants: for instance, the maximal order type (or m.o.t.), originally defined by De Jongh and Parikh [?], is the order type of the maximal linearization of a well-quasi-ordered set. In the case of a finite set, the m.o.t. is precisely the size of the set. Better runtime bounds were obtained by considering two other parameters [?]: the ordinal height introduced by Schmidt [?], and the ordinal width of Kříž and Thomas [?]. Therefore, when characterizing well-quasi-ordered languages, we will also be interested in deriving upper bounds on their ordinal invariants. This analysis also allows us to better compare the well-quasi-orderings. We refer to Section 2 for a more detailed introduction to these parameters and ordinal computations in general.

---

<sup>1</sup> This will be made precise in Lemma 41.

*Contributions* We focus on languages over a finite alphabet  $\Sigma$ . In this setting, we first characterize languages that are well-quasi-ordered by the prefix relation (and symmetrically, by the suffix relation), and derive tight bounds on their ordinal invariants. These generic results are then used to devise a decision procedure for checking whether a language is well-quasi-ordered by the prefix relation, provided the language is given as input as a finite automaton (Corollary 34). A summary of these results can be found in Figure 1.

| $L$       | Characterisation                              | $\mathfrak{w}(L)$ | $\mathfrak{o}(L)$ |
|-----------|---|-------------------|-------------------|
| arbitrary | Theorem 35: finite unions of chains           | $< \omega$        | $< \omega^2$      |
| regular   | Corollary 34: finite unions of regular chains | $< \omega$        | $< \omega^2$      |

Fig. 1: Summary of results for the prefix relation (and symmetrically, for the suffix relation).

We then turn our attention to the infix relation. In this case, we notice that Lemma 5.1 from [?] imply that there are well-quasi-ordered languages for the infix relation that have arbitrarily large ordinal invariants (except for the ordinal height, which is always at most  $\omega$ ). Therefore, we focus on two natural semantic restrictions on languages: on the one hand, we consider bounded languages, that is, languages included in some  $w_1^* \cdots w_k^*$  for some finite choice of words  $w_1, \dots, w_k$ ; on the other hand, we consider downwards closed languages, that is, languages closed under taking infixes. In both cases, we provide a very precise characterization of well-quasi-ordered languages by the infix relation, and derive tight bounds on their ordinal invariants. These results are summarized in Figure 2. We furthermore notice that for downwards closed languages that are well-quasi-ordered by the infix relation, being bounded is the same as being regular (Lemma 65), and that a bounded language is well-quasi-ordered by the infix relation if and only if its downwards closure is well-quasi-ordered by the infix relation (Corollary 51). This shows that, for bounded languages, being well-quasi-ordered implies that their downwards closure is a regular language, which is a weakening of the usual result that the downwards closure of *any language* for the scattered subword relation is always a regular language.

Turning our attention to decision procedures, we consider two computational models respectively tailored to downwards closed languages and to bounded languages. For downwards closed languages, we consider a model based on representations of infinite words (Section 5.2), for which we provide a decision procedure (Theorem 513). The model used to represent these infinite words is based on automatic sequences and morphic sequences [?], which are well-studied in the context of symbolic dynamics. For bounded languages, we consider the model of amalgamation systems [?], which is an abstract computational model that encompasses many classical ones, such as finite automata, context-free grammars, and Petri nets [?]. We show that if a language recognized by an amalgamation

| $L$              | Characterisation   | $\mathfrak{w}(L)$ | $\mathfrak{o}(L)$ |
|------------------|--|-------------------|-------------------|
| arbitrary        | Lemma 41: countable well-quasi orders with finite initial segments                 | $< \omega_1$      | $< \omega_1$      |
| bounded          | Theorem 42: finite union of products of chains for the prefix and suffix relations | $< \omega^2$      | $< \omega^3$      |
| downwards closed | Theorem 56: finite union of infixes of ultimately uniformly recurrent words        | $< \omega^2$      | $< \omega^3$      |

Fig. 2: Summary of results for the infix relation, the bounds on  $\mathfrak{w}(L)$  and  $\mathfrak{o}(L)$  are tight, and respectively proven in Corollary 48 and Corollary 57.

117 system is well-quasi-ordered by the infix relation, then it is a bounded language  
 118 (Theorem 61), and is therefore regular. Furthermore, we show that we can de-  
 119 cide whether a given language recognized by an amalgamation system is well-  
 120 quasi-ordered by the infix relation (Theorem 62). We defer the introduction of  
 121 amalgamation systems to Section 6.1.

122 *Related work* The study of alternative well-quasi-ordered relations over finite  
 123 words is far from new. For instance, orders obtained by so-called *derivation*  
 124 *relations* were already analysed by Bucher, Ehrenfeucht, and Haussler [?], and  
 125 were later extended by D'Alessandro and Varricchio [?,?]. However, in all those  
 126 cases the orderings are *multiplicative*, that is, if  $u_1 \preceq v_1$  and  $u_2 \preceq v_2$  then  
 127  $u_1 u_2 \preceq v_1 v_2$ . This assumption does not hold for the prefix, suffix, and infix  
 128 relations.

129 A similar question was studied by Atminas, Lozin, and Moshkov [?], in the  
 130 hope of finding characterizations of classes of *finite graphs* that are well-quasi-  
 131 ordered by the *induced subgraph relation* [?, Section 7]. In this setting, it is  
 132 common to refer to classes of graphs via a list of *forbidden patterns*, which are  
 133 finite graphs that cannot be found as induced subgraphs in the class. Applying  
 134 this reasoning to finite words with the infix relation, they provide an efficient  
 135 decision procedure for checking whether a language  $L \subseteq \Sigma^*$  is well-quasi-ordered  
 136 by the infix relation whenever said language is given as input via a list of *forbid-*  
 137 *den factors* [?, Theorem 1, Theorem 2]. The key construction of their paper is to  
 138 study languages  $L$  that are *regular* (recognized by some finite deterministic au-  
 139 tomata), for which they can decide whether  $L$  is well-quasi-ordered by the infix  
 140 relation [?, Theorem 1]. Because it is easy to transform a list of forbidden factors  
 141 into a regular language [?, Theorem 1], this yields the desired decision procedure.  
 142 Our work extends this result in several ways: first, we also consider the prefix  
 143 relation and the suffix relation, then we consider non-regular languages, and fi-  
 144 nally, we provide very precise descriptions of the well-quasi-ordered languages,  
 145 as well as tight bounds on their ordinal invariants.

146 *Outline* We introduce in Section 2 the necessary background on well-quasi-  
 147 orders and ordinal invariants. In Section 3, which is relatively self-contained,

we study the prefix relation and prove in Theorem 35 the characterization of well-quasi-ordered languages by the prefix relation. In Section 4, we obtain the infix analogue of Theorem 35 specifically for bounded languages (Theorem 42). In Section 5, we study the downwards closed languages, characterize them using a notion of ultimately uniformly recurrent words borrowed from symbolic dynamics (Theorem 56), and compute bounds on their ordinal invariants in Corollary 57. Finally, we generalize these results to all amalgamation systems in Section 6 in (Theorem 61), and provide a decision procedure for checking whether a language is well-quasi-ordered by the infix relation (resp. prefix and suffix) in this context (Theorem 62).

*Acknowledgements* We would like to thank participants of the 2024 edition of Autobóz for their helpful comments and discussions. We would also like to thank Vincent Jugé for his pointers on word combinatorics.

## 2 Preliminaries

*Finite words.* In this paper, we use upper Greek letters  $\Sigma, \Gamma$  to denote finite alphabets,  $\Sigma^*$  to denote the set of finite words over  $\Sigma$ , and  $\varepsilon$  for the empty word in  $\Sigma^*$ . In order to give some intuition on the decision problems, we will sometimes use the notion of *finite automata*, *regular languages*, and Monadic Second Order logic (**MSO**) over finite words, and assume the reader to be familiar with them. We refer to the textbook of [?] for a detailed introduction. However, we will require no prior knowledge on word combinatorics.

*Orderings and Well-Quasi-Orderings.* A *quasi-order* is a reflexive and transitive binary relation, it is a *partial order* if it is furthermore antisymmetric. A *total order* is a partial order where any two elements are comparable. Let now us introduce some notations for well-quasi-orders. A sequence  $(x_i)_{i \in \mathbb{N}}$  in a set  $X$  is *good* if there exist  $i < j$  such that  $x_i \leq x_j$ . It is *bad* otherwise. Therefore, a well-quasi-ordered set is a set where every infinite sequence is good. A *decreasing sequence* is a sequence  $(x_i)_{i \in \mathbb{N}}$  such that  $x_{i+1} < x_i$  for all  $i$ , a *chain* is a sequence such that  $x_i \leq x_{i+1}$  for all  $i$ , and an *antichain* is a set of pairwise incomparable elements. An equivalent definition of a well-quasi-ordered set is that it contains no infinite decreasing sequences, nor infinite antichains. We refer to [?] for a detailed survey on well-quasi-orders.

The prefix relation (resp. the suffix relation and the infix relation) on  $\Sigma^*$  are always *well-founded*, i.e., there are no infinite decreasing sequences for this ordering. In particular, for a language  $L \subseteq \Sigma^*$  to be well-quasi-ordered, it suffices to prove that it contains no infinite antichain.

A useful operation on quasi-ordered sets is to compute the *upwards closure* of a set  $S$  for a relation  $\preceq$ , which is defined as  $\uparrow_{\preceq} S \triangleq \{y \in \Sigma^* \mid \exists x \in S. x \preceq y\}$ . In this paper, we will also use the symmetric notion of *downwards closure*:  $\downarrow_{\preceq} S \triangleq \{y \in \Sigma^* \mid \exists x \in S. y \preceq x\}$ . Abusing notations, we will write  $\uparrow w$  and  $\downarrow w$  for the upwards and downwards closure of a single element  $w$ , omitting the

ordering relation when it is clear from the context. A set  $S$  is called *downwards closed* if  $\downarrow S = S$ .

*Ordinal Invariants.* An *ordinal* is a well-founded totally ordered set. We use  $\alpha, \beta, \gamma$  to denote ordinals, and use  $\omega$  to denote the first infinite ordinal, i.e., the set of natural numbers with the usual ordering. We also use  $\omega_1$  to denote the first *uncountable* ordinal. We only assume superficial familiarity with ordinal arithmetic, and refer to the books of Kunen [?] and Krivine [?, Chapter II] for a detailed introduction to this domain. Given a tree  $T$  whose branches are all finite we can define an ordinal  $\alpha_T$  inductively as follows: if  $T$  is a leaf then  $\alpha_T = 0$ , if  $T$  has children  $(T_i)_{i \in \mathbb{N}}$  then  $\alpha_T = \sup\{\alpha_{T_i} + 1 \mid i \in \mathbb{N}\}$ . We say that  $\alpha_T$  is the *rank* of  $T$ .

Let  $(X, \leq)$  be a well-quasi-ordered set. One can define three well-founded trees from  $X$ : the tree of bad sequences, the tree of decreasing sequences, and the tree of antichains. The rank of these trees are called respectively the *maximal order type* of  $X$  written  $\mathfrak{o}(X)$  [?], the *ordinal height* of  $X$  written  $\mathfrak{h}(X)$  [?], and the *ordinal width* of  $X$  written  $\mathfrak{w}(X)$  [?]. These three parameters are called the *ordinal invariants* of a well-quasi-ordered set  $X$ . As an example, for  $(\mathbb{N}, \leq)$ , all bad sequences are descending and antichains have size at most 1. In fact,  $(\mathbb{N}, \leq)$  is itself an ordinal, namely  $\omega$ . Hence it is its own maximal order type and ordinal height, and its ordinal width is 1. We refer to the survey of [?] for a detail discussion on these concepts and their computation on specific classes of well-quasi-ordered sets.

We will use the following inequality between ordinal invariants, due to [?], and that was recalled in [?, Theorem 3.8]:  $\mathfrak{o}(X) \leq \mathfrak{h}(X) \otimes \mathfrak{w}(X)$ , where  $\otimes$  is the *commutative ordinal product*, also known as the *Hessenberg product*. We will not recall the definition of this product here, and refer to [?, Section 3.5] for a detailed introduction to this concept. The only equalities we will use are  $\omega \otimes \omega = \omega^2$  and  $\omega^2 \otimes \omega = \omega^3$ .

### 3 Prefixes and Suffixes

In this section, we study the well-quasi-ordering of languages under the prefix relation. Let us immediately remark that the map  $u \mapsto u^R$  that reverses a word is an order-bijection between  $(X^*, \sqsubseteq_{\text{pref}})$  and  $(X^*, \sqsubseteq_{\text{suff}})$ , that is,  $u \sqsubseteq_{\text{pref}} v$  if and only if  $u^R \sqsubseteq_{\text{suff}} v^R$ . Therefore, we will focus on the prefix relation in the rest of this section, as  $(L, \sqsubseteq_{\text{pref}})$  is well-quasi-ordered if and only if  $(L^R, \sqsubseteq_{\text{suff}})$  is.

The next remark we make is that  $\Sigma^*$  is not well-quasi-ordered by the prefix relation as soon as  $\Sigma$  contains two distinct letters  $a$  and  $b$ . As an example of infinite antichain, we can consider the set of words  $a^n b$  for  $n \in \mathbb{N}$ . As mentioned in the introduction, there are however some languages that are well-quasi-ordered by the prefix relation. A simple example being the (regular) language  $a^* \subseteq \{a, b\}^*$ , which is order-isomorphic to natural numbers with their usual orderings  $(\mathbb{N}, \leq)$ .

In order to characterize the existence of infinite antichains for the prefix relation, we will introduce the following tree.

**Definition 31** The **tree of prefixes** over a finite alphabet  $\Sigma$  is the infinite tree  $T$  whose nodes are the words of  $\Sigma^*$ , and such that the children of a word  $w$  are the words  $wa$  for all  $a \in \Sigma$ .

We will use this tree of prefixes to find simple witnesses of the existence of infinite antichains in the prefix relation for a given language  $L$ , namely by introducing antichain branches.

**Definition 32** An **antichain branch** for a language  $L$  is an infinite branch  $B$  of the tree of prefixes such that from every point of the branch, one can reach a word in  $L \setminus B$ . Formally:  $\forall u \in B, \exists v \in \Sigma^*, uv \in L \setminus B$ .

Let us illustrate the notion of antichain branch over the alphabet  $\Sigma = \{a, b\}$ , and the language  $L = a^*b$ . In this case, the set  $a^*$  (which is a branch of the tree of prefixes) is an antichain branch for  $L$ . This holds because for any  $a^k$ , the word  $a^k \sqsubseteq_{\text{pref}} a^kb$  belongs to  $L \setminus a^*$ . In general, the existence of an antichain branch for a language  $L$  implies that  $L$  contains an infinite antichain, and because the alphabet  $\Sigma$  is assumed to be finite, one can leverage the fact that the tree of prefixes is finitely branching to prove that the converse holds as well.

**Lemma 33** Let  $L \subseteq \Sigma^*$  be a language. Then,  $L$  contains an infinite antichain if and only if there exists an antichain branch for  $L$ . ▷ Proven p.21

One immediate application of Lemma 33 is that antichain branches can be described inside the tree of prefixes by a monadic second order formula (MSO-formula), allowing us to leverage the decidability of MSO over infinite binary trees [?, Theorem 1.1]. This result will follow from our general decidability result (Theorem 62) but is worth stating on its own for its simplicity.

**Corollary 34** If  $L$  is regular, then the existence of an infinite antichain is decidable. ▷ Proven p.22

Let us now go further and fully characterize languages  $L$  such that the prefix relation is well-quasi-ordered, without any restriction on the decidability of  $L$  itself.

**Theorem 35.** A language  $L \subseteq \Sigma^*$  is well-quasi-ordered by the prefix relation if and only if  $L$  is a union of chains. ▷ Proven p.21

As an immediate consequence, we have a very fine-grained understanding of the ordinal invariants of such well-quasi-ordered languages, which can be leveraged in bounding the complexity of algorithms working on such languages.

**Corollary 36** Let  $L \subseteq \Sigma^*$  be a language that is well-quasi-ordered by the prefix relation. Then, maximal order type of  $L$  strictly smaller than  $\omega^2$ , the ordinal height of  $L$  is at most  $\omega$ , and its ordinal width is finite. Furthermore, these bounds are tight.

*Proof.* The upper bounds follow from the fact that  $L$  is a finite union of chains. The tightness can be obtained by considering the languages  $L_k \triangleq \bigcup_{i=0}^{k-1} a^i b^*$  for  $k \in \mathbb{N}$ , which are well-quasi-ordered by the prefix relation (as they are finite unions of chains), and satisfy that  $\mathfrak{w}(L_k) = k$ ,  $\mathfrak{h}(L_k) = \omega$ , and therefore  $\mathfrak{o}(L_k) = k \cdot \omega$ .

## 4 Infixes and Bounded Languages

In this section, we study languages equipped with the infix relation. As opposed to the prefix and suffix relations, the infix relation can lead to very complicated well-quasi-ordered languages. Formally, the upcoming Lemma 41 due to Kuske shows that *any* countable partial-ordering with finite initial segments can be embedded into the infix relation of a language. To make the former statement precise, let us recall that an *order embedding* from a quasi-ordered set  $(X, \preceq)$  into a quasi-ordered set  $(Y, \preceq')$  is a function  $f: X \rightarrow Y$  such that for all  $x, y \in X$ ,  $x \preceq y$  if and only if  $f(x) \preceq' f(y)$ . When such an embedding exists, we say that  $X$  *embeds into*  $Y$ . Recall that a quasi-ordered set  $(X, \preceq)$  is a partial ordering whenever the relation  $\preceq$  is antisymmetric, that is  $x \preceq y$  and  $y \preceq x$  implies  $x = y$ . A simplified version of the embedding defined in Lemma 41 is illustrated for the subword relation in Figure 5 page 23.

**Lemma 41** [?, Lemma 5.1] *Let  $(X, \preceq)$  be a partially ordered set, and  $\Sigma$  be an alphabet with at least two letters. Then the following are equivalent:*

1.  $X$  embeds into  $(\Sigma^*, \sqsubseteq_{\text{infix}})$ ,
2.  $X$  is countable, and for every  $x \in X$ , its downwards closure  $\downarrow_{\preceq} x$  is finite (that is,  $(X, \preceq)$  has *finite initial segments*).

As a consequence of Lemma 41, we cannot replay proofs of Section 3, and will actually need to leverage some regularity of the languages to obtain a characterization of well-quasi-ordered languages under the infix relation. This regularity will be imposed through the notion of *bounded languages*, i.e., languages  $L \subseteq \Sigma^*$  such that there exists words  $w_1, \dots, w_n$  satisfying  $L \subseteq w_1^* \cdots w_n^*$ . Let us now state the main theorem of this section.

**Theorem 42.** *Let  $L$  be a bounded language of  $\Sigma^*$ . Then,  $L$  is a well-quasi-order when endowed with the infix relation if and only if it is included in a finite union of products  $S_i \cdot P_i$  where  $S_i$  is a chain for the suffix relation, and  $P_i$  is a chain for the prefix relation, for all  $1 \leq i \leq n$ .*

Let us first remark that if  $S$  is a chain for the suffix relation and  $P$  is a chain for the prefix relation, then  $SP$  is well-quasi-ordered for the infix relation. This proves the (easy) right-to-left implication of Theorem 42.

In order to prove the (difficult) left-to-right implication of Theorem 42, we will rely heavily on the combinatorics of periodic words. Let us use a slightly non-standard notation by saying that a non-empty word  $w \in \Sigma^+$  is *periodic* with



period  $x \in \Sigma^*$  if there exists a  $p \in \mathbb{N}$  such that  $w \sqsubseteq_{\text{infix}} x^p$ . The *periodic length* of a word  $u$  is the minimal length of a period  $x$  of  $u$ .

The reason why periodic words built using a given period  $x \in \Sigma^+$  are interesting for the infix relation is that they naturally create chains for the prefix and suffix relations. Indeed, if  $x \in \Sigma^+$  is a finite word, then  $\{x^p \mid p \in \mathbb{N}\}$  is a chain for the infix relation. Note that in general, the downwards closure of a chain is *not* a chain (see Remark 43). However, for the chains generated using periodic words, the downwards closure  $\downarrow_{\sqsubseteq_{\text{infix}}} \{x^p \mid p \in \mathbb{N}\}$  is a *finite union* of chains. Because this set will appear in bigger equations, we introduce the shorter notation  $\text{P}\downarrow(x)$  for the set of infixes of words of the form  $x^p$ , where  $p \in \mathbb{N}$ .

**Remark 43** Let  $(X, \preceq)$  be a quasi-ordered set, and  $L \subseteq X$  be such that  $(L, \preceq)$  is well-quasi-ordered. It is not true in general that  $(\downarrow L, \preceq)$  is well-quasi-ordered. In the case of  $(\Sigma^*, \sqsubseteq_{\text{infix}})$  a typical example is to start from an infinite antichain  $A$ , together with an enumeration  $(w_i)_{i \in \mathbb{N}}$  of  $A$ , and build the language  $L \triangleq \{\prod_{i=0}^n w_i \mid i \in \mathbb{N}\}$ . By definition,  $L$  is a chain for the infix ordering, hence well-quasi-ordered. However,  $\downarrow_{\sqsubseteq_{\text{infix}}} L$  contains  $A$ , and is therefore not well-quasi-ordered.

**Lemma 44** Let  $x \in \Sigma^+$  be a word. Then  $\text{P}\downarrow(x)$  is a finite union of chains for the infix, prefix and suffix relations simultaneously. ▷ Proven p.23

The following combinatorial Lemma 46 connects the property of being well-quasi-ordered to a property of the periodic lengths of words in a language, based on the assumption that some factors can be iterated. It is the core result that powers the analysis done in the upcoming Theorems 42 and 61. It is fundamentally based on a classical result of combinatorics on words (Lemma 45) that we recall here for the sake of completeness.

**Lemma 45 ([?, Theorem 1])** Let  $u, v \in \Sigma^+$  be two words and  $n = \gcd(|u|, |v|)$ . If there exists  $p, q \in \mathbb{N}$  such that  $u^p$  and  $v^q$  have a common prefix of length at least  $|uv| - n$ , then there exists  $z \in \Sigma^+$  such that  $u$  and  $v$  are powers of  $z$ , and in particular  $z$  has length at most  $\min\{|u|, |v|\}$ .

**Lemma 46** Let  $L \subseteq \Sigma^*$  be a language that is well-quasi-ordered by the infix relation. Let  $k \in \mathbb{N}$ ,  $u_1, \dots, u_{k+1} \in \Sigma^*$ , and  $v_1, \dots, v_k \in \Sigma^+$  be such that  $w[\mathbf{n}] \triangleq (\prod_{i=1}^k u_i v_i^{n_i}) u_{k+1}$  belongs to  $L$  for arbitrarily large values of  $\mathbf{n} \in \mathbb{N}^k$ . Then, there exists  $x, y \in \Sigma^+$  of size at most  $\max\{|v_i| \mid 1 \leq i \leq k\}$  such that for all  $\mathbf{n} \in \mathbb{N}^k$  one of the following holds: ▷ Proven p.23

1.  $w[\mathbf{n}] \in u_1 \text{P}\downarrow(x)$ ,
2.  $w[\mathbf{n}] \in \text{P}\downarrow(x) u_{k+1}$ ,
3.  $w[\mathbf{n}] \in \text{P}\downarrow(x) u_i \text{P}\downarrow(y)$  for some  $1 \leq i \leq k+1$ .

**Lemma 47** Let  $L \subseteq \Sigma^*$  be a bounded language that is well-quasi-ordered by the infix relation. Then, there exists a finite subset  $E \subseteq (\Sigma^*)^3$ , such that: ▷ Proven p.24

$$L \subseteq \bigcup_{(x,u,y) \in E} \text{P}\downarrow(x) u \text{P}\downarrow(y) \quad .$$

▷ Back to p.8

347 *Proof (Proof of Theorem 42 as stated on page 8).* We apply Lemma 47, and  
 348 conclude because  $P\downarrow(x)$  is a finite union of chains for the prefix, suffix and infix  
 349 relations (Lemma 44).

350 **Corollary 48** *Let  $L$  be a bounded language of  $\Sigma^*$  that is well-quasi-ordered by*  
 351 *the infix relation. Then, the ordinal width of  $L$  less than  $\omega^2$ , its ordinal height*  
 352 *is at most  $\omega$ , and its maximal order type less than  $\omega^3$ . Furthermore, those three*  
 353 *bounds are tight.*

354 *Proof.* Upper bounds are a direct consequence of Theorem 42, and the tight-  
 355 ness is witnessed by the languages:  $L_k \triangleq \bigcup_{i=2}^{k+1} (ab^i a)^* (ba^i b)^*$ , that are bounded  
 356 languages of  $\{a, b\}^*$ , well-quasi-ordered by the infix relation, and have ordinal  
 357 width, ordinal height and maximal order type respectively equal to  $\omega \cdot k$ ,  $\omega$  and  
 358  $\omega^2 \cdot k$ .

## 359 5 Infixes and Downwards Closed Languages

360 Let us now discuss another classical restriction that can be imposed on languages  
 361 when studying well-quasi-orders, that of being downwards closed. Indeed, the  
 362 Lemma 41 crucially relies on constructing languages that are *not* downwards  
 363 closed, and we have shown in Remark 43 that the downwards closure of a well-  
 364 quasi-ordered language is not necessarily well-quasi-ordered.

### 365 5.1 Characterization of Well-Quasi-Ordered Downwards Closed 366 Languages

367 An immediate consequence of Theorem 42 is that if  $L$  is a bounded language,  
 368 then considering  $L$  or its downwards closure  $\downarrow_{\sqsubseteq_{\text{infix}}} L$  is equivalent with respect  
 369 to being well-quasi-ordered by the infix relation, as opposed to the general case  
 370 illustrated in Remark 43.

▷ Proven p.25

371 **Corollary 51** *Let  $L$  be a bounded language of  $\Sigma^*$ . Then,  $L$  is a well-quasi-order*  
 372 *when endowed with the infix relation if and only if  $\downarrow_{\sqsubseteq_{\text{infix}}} L$  is.*

373 The Corollary 51 is reminiscent of a similar result for the subword embed-  
 374 ding, stipulating that for any language  $L \subseteq \Sigma^*$ , the downwards closure  $\downarrow_{\leq^*} L$   
 375 is described using finitely many excluded subwords, hence is regular. However,  
 376 this is not the case for the infix relation, even with bounded languages, as we  
 377 will now illustrate with the following example.

378 **Example 52** *Let  $L \triangleq a^* b^* \cup b^* a^*$ . This language is bounded, is downwards*  
 379 *closed for the infix relation, is well-quasi-ordered for the infix relation, but is*  
 380 *characterized by an infinite number of excluded infixes, respectively of the form*  
 381  *$ab^k a$  and  $ba^k b$  where  $k \geq 1$ .*

To strengthen Example 52, we will leverage the *Thue-Morse sequence*  $\mathbf{t} \in \{0,1\}^{\mathbb{N}}$ , which we will use as a black-box for its two main characteristics: it is cube-free and uniformly recurrent. Being *cube-free* means that no (finite) word of the form  $uuu$  is an infix of  $\mathbf{t}$ , and being *uniformly recurrent* means that for every word  $u$  that is an infix of  $\mathbf{t}$ , there exists  $k \geq 1$  such that  $u$  occurs as an infix of every  $k$ -sized infix  $v \sqsubseteq_{\text{infix}} \mathbf{t}$ . We refer the reader to a nice survey of Allouche and Shallit for more information on this sequence and its properties [?].

**Theorem 53.** *Let  $w \in \Sigma^{\mathbb{N}}$  be a uniformly recurrent word. Then, the set of finite infixes of  $w$  is well-quasi-ordered for the infix relation.*

*Proof.* Let  $L$  be the set of finite infixes of  $w$ . Consider a sequence  $(u_i)_{i \in \mathbb{N}}$  of words in  $L$ . Without loss of generality, we may consider a subsequence such that  $|u_i| < |u_{i+1}|$  for all  $i \in \mathbb{N}$ . Because  $\mathbf{t}$  is uniformly recurrent, there exists  $k \geq 1$  such that  $u_1$  is an infix of every word  $v$  of size at least  $k$ . In particular,  $u_1$  is an infix of  $u_k$ , hence the sequence  $(u_i)_{i \in \mathbb{N}}$  is good.

**Lemma 54** *The language  $I_{\mathbf{t}}$  of infixes of the Thue-Morse sequence is downwards closed for the infix relation, well-quasi-ordered for the infix relation, but is not bounded.*

*Proof.* By construction  $I_{\mathbf{t}}$  is downwards closed for the infix relation, and by Theorem 53, it is well-quasi-ordered.

Assume by contradiction that  $I_{\mathbf{t}}$  is bounded. In this case, there exist words  $w_1, \dots, w_k \in \Sigma^*$  such that  $I_{\mathbf{t}} \subseteq w_1^* \cdots w_k^*$ . Since  $I_{\mathbf{t}}$  is infinite and downwards closed, there exists a word  $u \in I_{\mathbf{t}}$  such that  $u = w_i^3$  for some  $1 \leq i \leq k$ . This is a contradiction, because  $u \sqsubseteq_{\text{infix}} \mathbf{t}$ , which is cube-free.

One may refine our analysis of the Thue-Morse sequence to obtain precise bounds on the ordinal invariants of its language of infixes.

**Lemma 55** *Under  $\sqsubseteq_{\text{infix}}$ , the maximal order type of  $I_{\mathbf{t}}$  is  $\omega$ , the ordinal height of  $I_{\mathbf{t}}$  is  $\omega$ , the ordinal width of  $I_{\mathbf{t}}$  is  $\omega$ .*

*Proof.* We first show that  $\omega$  is an upper bound for each of these measure, before showing that the bounds are tight.

Let us prove that these are upper bounds for the ordinal invariants of  $I_{\mathbf{t}}$ . The bound of the ordinal height holds for any language  $L$ , as the length of a decreasing sequence of words is bounded by the length of its first element. For the maximal order type, we remark that the uniform recurrence of  $\mathbf{t}$  means that the maximal length of a bad sequence is determined by its first element, hence that it is at most  $\omega$ . Finally, because the ordinal width is at most the maximal order type (as per Section 2, using for instance the results of [?] or [?, Theorem 3.8] stating  $\text{o}(X) \leq \text{h}(X) \otimes \text{w}(X)$ ): we conclude that the ordinal width is also at most  $\omega$ .

Now, let us prove that these bounds are tight. It is clear that  $\text{h}(I_{\mathbf{t}}) = \omega$ : given any number  $n \in \mathbb{N}$ , one can construct a decreasing sequence of words in

422  $I_t$  of length  $n$ , for instance by considering the first  $n$  prefixes of the Thue-Morse  
 423 sequence by decreasing size. Let us now prove that  $\mathfrak{w}(I_t) = \omega$ . To that end,  
 424 we can leverage the fact that the number of infixes of size  $n$  in  $I_t$  is bounded  
 425 below by a non-constant affine function in  $n$  [?], and that two words of length  
 426  $n$  are comparable for the infix relation if and only if they are equal. Hence,  
 427 there cannot be a bound on the size of an antichain in  $I_t$ , and we conclude that  
 428  $\mathfrak{w}(I_t) = \omega$ . Finally, because the ordinal width is at most the maximal order type,  
 429 we conclude that the maximal order type of  $I_t$  is also  $\omega$ .

430 We prove in the upcoming Theorem 56 that the status of the Thue-Morse  
 431 sequence is actually representative of downwards closed languages for the infix  
 432 relation. To that end, let us introduce the notation  $\text{Infixes}(w)$  for the set of finite  
 433 infixes of a (possibly infinite or bi-infinite) word  $w \in \Sigma^* \cup \Sigma^{\mathbb{N}} \cup \Sigma^{\mathbb{Z}}$ . We say  
 434 that an infinite word  $w \in \Sigma^{\mathbb{N}}$  is *ultimately uniformly recurrent* if there exists a  
 435 bound  $N_0 \in \mathbb{N}$  such that  $w_{\geq N_0}$  is uniformly recurrent. We extend this notion to  
 436 finite words by considering that they all are ultimately uniformly recurrent, and  
 437 to bi-infinite words by considering that they are ultimately uniformly recurrent  
 438 if and only if both their left-infinite and right-infinite parts are.

▷ Proven p.13

439 **Theorem 56.** *Let  $L$  be a well-quasi-ordered language for the infix relation that*  
 440 *is downwards closed. Then, there exist finitely many ultimately uniformly recur-*  
 441 *rent words  $w_1, \dots, w_n \in \Sigma^* \cup \Sigma^{\mathbb{N}} \cup \Sigma^{\mathbb{Z}}$  such that  $L = \bigcup_{i=1}^n \text{Infixes}(w_i)$ .*

442 Thanks to Theorem 56, and by analysing the ordinal invariants of infixes  
 443 of an ultimately uniformly recurrent infinite word  $w$  (Lemma 59), we conclude  
 444 that the ordinal invariants of a well-quasi-ordered downwards closed language  
 445 are relatively small.

▷ Proven p.28

446 **Corollary 57** *Then, the maximal order type of  $L$  is strictly less than  $\omega^3$ , its*  
 447 *ordinal height is at most  $\omega$ , and its ordinal width is at most  $\omega^2$ .*  
 448 *Furthermore, those bounds are tight.*

449 To connect infixes of a (bi)-infinite word to downwards closed languages, a  
 450 useful notion is that of directed sets. A subset  $I \subseteq X$  is *directed* if, for every  
 451  $x, y \in I$ , there exists  $z \in I$  such that  $x \leq z$  and  $y \leq z$ . Given a well-quasi-order  
 452  $(X, \leq)$ , one can always decompose  $X$  into a finite union of *order ideals*, that is,  
 453 non-empty sets  $I \subseteq X$  that are downwards closed and directed for the relation  
 454  $\leq$ . In our case, a well-quasi-ordered order ideal for the infix relation is the set of  
 455 finite infixes of a finite, infinite, or bi-infinite word  $w \in \Sigma^* \cup \Sigma^{\mathbb{N}} \cup \Sigma^{\mathbb{Z}}$  (Lemma 58).

▷ Proven p.25

456 **Lemma 58** *Let  $L \subseteq \Sigma^*$  be an order ideal for a well-quasi-ordered infix relation.*  
 457 *Then  $L$  is the set of finite infixes of a finite, infinite or bi-infinite word  $w$ .*

▷ Proven p.25

458 **Lemma 59** *Let  $w \in \Sigma^{\mathbb{N}}$  be an infinite word. Then, the set of finite infixes of  $w$*   
 459 *is well-quasi-ordered for the infix relation if and only if  $w$  is ultimately uniformly*  
 460 *recurrent.*

▷ Proven p.26

**Lemma 510** *Let  $w \in \Sigma^{\mathbb{Z}}$  be a bi-infinite word. Then, the set of finite infixes of  $w$  is well-quasi-ordered for the infix relation if and only if  $w$  is ultimately uniformly recurrent as a bi-infinite word.*

We are now ready to conclude the proof of Theorem 56.

*Proof (Proof of Theorem 56 as stated on page 12).* It is clear that the set of finite infixes of a finite, infinite or bi-infinite ultimately uniformly recurrent word is well-quasi-ordered for the infix relation thanks to Lemma 59.

Conversely, let us consider a well-quasi-ordered language  $L$  that is downwards closed for the infix relation. Because it is a well-quasi-ordered set, it can be written as a finite union of order ideals  $L = \bigcup_{i=1}^n L_i$ .

For every such ideal  $L_i$ , we can apply Lemma 58, and conclude that  $L_i$  is the set of finite infixes of a finite, infinite or bi-infinite word  $w_i$ . Because the languages  $L_i$  are well-quasi-ordered, we can apply Lemma 59, and conclude that  $w_i$  is ultimately uniformly recurrent.

▷ Back to p.12

Finally, we comment on the ordinal invariants of the set of finite infixes of an ultimately uniformly recurrent infinite word, from which the bounds of Corollary 57 naturally follow (the proof is in Appendix D page 28).

**Lemma 511** *Let  $w \in \Sigma^{\mathbb{N}}$  be an ultimately uniformly recurrent word. Then, the set of finite infixes of  $w$  has ordinal width less than  $\omega \cdot 2$ . Furthermore, this bound is tight.*

▷ Proven p.27

**Lemma 512** *Let  $w \in \Sigma^{\mathbb{Z}}$  be a bi-infinite word. Then, the set of finite infixes of  $w$  is well-quasi-ordered for the infix relation if and only if  $w_+$  and  $w_-$  are two ultimately uniformly recurrent words. In this case, the ordinal width of the set of finite infixes of  $w$  is less than  $\omega \cdot 3$ , and this bound is tight.*

▷ Proven p.27

## 5.2 Decision Procedures

As we have demonstrated, infinite (or bi-infinite words) can be used to represent languages that are well-quasi-ordered for the infix relation by considering their set of finite infixes. Let us formalise the representation of languages by sets of bi-infinite words that we will use in this section, following the characterization of Lemma 58. A *sequence representation* of a language  $L \subseteq \Sigma^*$  is a finite set of triples  $(w_i^-, a_i, w_i^+)_{1 \leq i \leq n}$  where  $w_i^-, w_i^+ \in \Sigma^{\mathbb{N}} \cup \Sigma^*$  are two potentially infinite words, and  $a_i \in \Sigma$  is a letter, such that

$$L = \bigcup_{i=1}^n \text{Infixes}(\text{reversed}(w_i^-) a_i w_i^+) \quad .$$

Given an effective representation of sequences, one obtains an effective representation of languages via sequence representations. In this section, we will rely on definitions originating from the area of symbolic dynamics, that precisely

study infinite words whose generation follows from a finitely described process. However, we will not assume that the reader is familiar with this domain, and we will use as black-boxes key results from this area.

A first model that one can use to represent infinite words is the model of *automatic sequences*. In this case, the infinite word  $w$  is described by a finite state automaton, that can compute the  $i$ -th letter of the word  $w$  given as input the number  $i$  written in some base  $b \in \mathbb{N}$ . An example of such a sequence is the Thue-Morse sequence that can be described by a finite automaton using a binary representation of the indices. The good algorithmic properties of automatic sequences come from the fact that a Presburger definable property that uses letters of the sequence can be (trivially) translated into a finite automaton that reads the base  $b$  representation of the free variables (that are indices of the sequence). In particular, it follows that one can decide if an automatic sequence is ultimately uniformly recurrent, a proof of this folklore result can be found in the appendix at Lemma D1. Based on this, we now prove:

**Theorem 513.** *Given a sequence representation of a language  $L \subseteq \Sigma^*$  where all infinite words are automatic sequences, one can decide whether  $L$  is well-quasi-ordered for the infix relation.*

*Proof.* It is easy to see that  $L$  is well-quasi-ordered for the infix relation if and only if for every triple  $(w_i^-, a_i, w_i^+)$  in the sequence representation of  $L$ , the (potentially bi-infinite) word  $\text{reversed}(w_i^-)a_iw_i^+$  defines a well-quasi-ordered language. By Lemma 512, this is the case if and only if both  $w_i^-$  and  $w_i^+$  are ultimately uniformly recurrent. Since one can decide whether an automatic sequence is ultimately uniformly recurrent using Lemma D1, we conclude the proof.

In fact, automatic sequences are part of a larger family of sequences studied in symbolic dynamics, called morphic sequences. Let us first recall that a *morphism* is a function  $f: \Sigma^* \rightarrow \Gamma^*$  such that for every  $u, v \in \Sigma^*$ ,  $f(uv) = f(u)f(v)$ . A *morphic sequence*  $w$  is an infinite word obtained by iterating a morphism  $f: \Sigma^* \rightarrow \Sigma^*$  on a letter  $a \in \Sigma$  such that  $f(a)$  starts with  $a$ , and then applying a homomorphism  $h: \Sigma^* \rightarrow \Gamma^*$ . The infinite word  $f^\omega(a)$  is the limit of the sequence  $(f^n(a))_{n \in \mathbb{N}}$ , which is well-defined because  $f(a)$  starts with  $a$ , and the morphic sequence is  $w \triangleq h(f^\omega(a))$ .

Every automatic sequence is a morphic sequence, but not the other way around. We refer the reader to a short survey of [?] for more details on the possible variations on the definition of morphic sequences and their relationships. It was relatively recently proven that one can decide whether a morphic sequence is uniformly recurrent [?, Theorem 1]. We were not able to find in the literature whether one can decide ultimate uniform recurrence, but conjecture that it is the case, which would allow us to decide whether a language represented by morphic sequences is well-quasi-ordered for the infix relation.

**Conjecture 13.** *Given a morphic sequence  $w \in \Sigma^\mathbb{N}$ , one can decide whether it is ultimately uniformly recurrent.*

## 538 6 Infixes and Amalgamation Systems

539 In the previous section, we have represented languages that are downwards closed  
 540 by the infix relation as infixes of infinite words. However, there are many other  
 541 natural ways to represent languages, such as finite automata or context-free  
 542 grammars. In this section, we are going to show that our results on bounded lan-  
 543 guages can be applied to a large class of systems, called amalgamation systems,  
 544 that includes as particular examples finite automata and context-free grammars.

545 Our first result, of theoretical nature, is that amalgamation systems cannot  
 546 define well-quasi-ordered languages that are not bounded. This implies that all  
 547 the results of Section 4, and in particular Theorem 42, can safely be applied to  
 548 amalgamation systems.

549 **Theorem 61.** *Let  $L \subseteq \Sigma^*$  be a language recognized by an amalgamation system.* ▷ Proven p.31  
 550 *If  $L$  is well-quasi-ordered by the infix relation then  $L$  is bounded.*

551 Our second focus is of practical nature: we want to give a decision procedure  
 552 for being well-quasi-ordered. This will require us to introduce *effectiveness as-*  
 553 *sumptions* on the amalgamation systems. While most of them will be innocuous,  
 554 an important consequence is that we have to consider *classes of languages* rather  
 555 than individual ones, for instance: the class of all regular language, or the class  
 556 of all context-free languages. Such classes will be called effective amalgamative  
 557 classes (Section 6.1). In the following theorem, we prove that under such assump-  
 558 tions, testing well-quasi-ordering is inter-reducible to testing whether a language  
 559 of the class is empty, which is usually the simplest problem for a computational  
 560 model.

561 **Theorem 62.** *Let  $\mathcal{C}$  be an effective amalgamative class of languages. Then the* ▷ Proven p.31  
 562 *following are equivalent:*

- 563 1. *Well-quasi-orderedness of the infix relation is decidable for languages in  $\mathcal{C}$ .*
- 564 2. *Well-quasi-orderedness of the prefix relation is decidable for languages in  $\mathcal{C}$ .*
- 565 3. *Emptiness is decidable for languages in  $\mathcal{C}$ .*

### 566 6.1 Amalgamation Systems

567 Let us now formally introduce the notion of amalgamation systems, and recall  
 568 some results from [?] that will be useful for the proof of Theorem 61. The notion  
 569 of amalgamation system is tailored to produce *pumping arguments*, which is ex-  
 570 actly what our Lemma 46 talks about. At the core of a pumping argument, there  
 571 is a notion of a *run*, which could for instance be a sequence of transitions taken  
 572 in a finite state automaton. Continuing on the analogy with finite automata,  
 573 there is a natural ordering between runs, i.e., a run is smaller than another one  
 574 if one can “delete” loops of the larger run to obtain the other. Typical pumping  
 575 arguments then rely on the fact that *minimal* runs are of finite size, and that  
 576 all other runs are obtained by “gluing” loops to minimal runs. Generalizing this  
 577 notion yields the notion of amalgamation systems.

Let us recall that over an alphabet  $(\Sigma, =)$  a subword embedding between two words  $u \in \Sigma^*$  and  $v \in \Sigma^*$  is a function  $\rho: [1, |u|] \rightarrow [1, |v|]$  such that  $u_i = v_{\rho(i)}$  for all  $i \in [1, |u|]$ . We write  $\text{Hom}^*(u, v)$  the set of all subword embeddings between  $u$  and  $v$ . It may be useful to notice that the set of finite words over  $\Sigma$  forms a category when we consider subword embeddings as morphisms, which is a fancy way to state that  $\text{id} \in \text{Hom}^*(u, u)$  and that  $f \circ g \in \text{Hom}^*(u, w)$  whenever  $g \in \text{Hom}^*(u, v)$  and  $f \in \text{Hom}^*(v, w)$ , for any choice of words  $u, v, w \in \Sigma^*$ .

Given a subword embedding  $f: u \rightarrow v$  between two words  $u$  and  $v$ , there exists a unique decomposition  $v = G_0^f u_1 G_1^f \cdots G_{k-1}^f u_k G_k^f$  where  $G_i^f = v_{f(i)+1} \cdots v_{f(i+1)-1}$  for all  $1 \leq i \leq k-1$ ,  $G_k^f = v_{f(k)+1} \cdots v_{|v|}$ , and  $G_0^f = v_1 \cdots v_{f(1)-1}$ . We say that  $G_i^f$  is the  $i$ -th *gap word* of  $f$ . We encourage the reader to look at Figure 6 to see an example of the gap words resulting from a subword embedding between two words. These gap words will be useful to describe how and where runs of a system (described by words) can be combined.

**Definition 63** An *amalgamation system* is a tuple  $(\Sigma, R, \text{can}, E)$  where  $\Sigma$  is a finite alphabet,  $R$  is a set of so-called runs,  $\text{can}: R \rightarrow (\Sigma \uplus \{\#\})^*$  is a function computing a *canonical decomposition* of a run, and  $E$  describes the so-called *admissible embeddings* between runs: If  $\rho$  and  $\sigma$  are runs from  $R$ , then  $E(\rho, \sigma)$  is a subset of the subword embeddings between  $\text{can}(\rho)$  and  $\text{can}(\sigma)$ . We write  $\rho \trianglelefteq \sigma$  if  $E(\rho, \sigma)$  is non-empty. If we want to refer to a specific embedding  $f \in E(\rho, \sigma)$ , we also write  $\rho \trianglelefteq_f \sigma$ . Given a run  $r \in R$ , and  $i \in [0, |\text{can}(r)|]$ , the *gap language* of  $r$  at position  $i$  is  $L_i^r \triangleq \{G_i^f \mid \exists s \in R. \exists f \in E(r, s)\}$ . An amalgamation system furthermore satisfies the following properties:

1.  $(R, E)$  Forms a Category. For all  $\rho, \sigma, \tau \in R$ ,  $\text{id} \in E(\rho, \rho)$ , and whenever  $f \in E(\rho, \sigma)$  and  $g \in E(\sigma, \tau)$ , then  $g \circ f \in E(\rho, \tau)$ .
2. Well-Quasi-Ordered System.  $(R, \trianglelefteq)$  is a well-quasi-ordered set.
3. Concatenative Amalgamation. Let  $\rho_0, \rho_1, \rho_2$  be runs with  $\rho_0 \trianglelefteq_f \rho_1$  and  $\rho_0 \trianglelefteq_g \rho_2$ . Then for all  $0 \leq i \leq |\text{can}(\rho_0)|$ , there exists a run  $\rho_3 \in R$  and embeddings  $\rho_1 \trianglelefteq_{g'} \rho_3$  and  $\rho_2 \trianglelefteq_{f'} \rho_3$  satisfying two conditions: (a)  $g' \circ f = f' \circ g$  (we write  $h$  for this composition) and (b) for every  $0 \leq j \leq |\rho_0|$ , the gap word  $G_j^h$  is either  $G_j^f G_j^g$  or  $G_j^h = G_j^g G_j^f$ . Specifically, for  $i$  we may fix  $G_i^h = G_i^f G_i^g$ . We refer to Figure 7 for an illustration of this property.

The yield of a run is obtained by projecting away the separator symbol  $\#$  from the canonical decomposition, i.e.  $\text{yield}(\rho) = \pi_\Sigma(\rho)$ . The language recognized by an amalgamation system is  $\text{yield}(R)$ .

We say a language  $L$  is an *amalgamation language* if there exists an amalgamation system recognizing it.

Intuitively, the definition of an amalgamation system allows the comparison of runs, and the proper “gluing” of runs together to obtain new runs. A number of well-known language classes can be seen to be recognized by amalgamation systems, e.g., regular languages [?, Theorem 5.3], reachability and coverability languages of VASS [?, Theorem 5.5], and context-free languages [?, Theorem 5.10].



We can now show a simple lemma that illuminates much of the structure of amalgamation systems whose language is well-quasi-ordered by  $\sqsubseteq_{\text{infix}}$ . Note that Lemma 64 uses Lemma 46 in its proof, and our Theorem 61 follows from it.

**Lemma 64** *Let  $L$  be an amalgamation language recognized by  $(\Sigma, R, E, \text{can})$  that is well-quasi-ordered by  $\sqsubseteq_{\text{infix}}$ . Let  $\rho$  be a run with  $\rho = a_1 \cdots a_n$ , and let  $\sigma, \tau$  be runs with  $\rho \sqsubseteq_f \sigma$  and  $\rho \sqsubseteq_g \sigma$ .  
For any  $0 \leq \ell \leq n$ , we have  $G_\ell^f \sqsubseteq_{\text{infix}} G_\ell^g$  or vice versa.* ▷ Proven p.30

If we additionally assume that such a language is closed under taking infixes, we obtain an even stronger structure: All such languages are regular!

**Lemma 65** *Let  $L \subseteq \Sigma^*$  be a downwards closed language for the infix relation that is well-quasi-ordered. Then, the following are equivalent:* ▷ Proven p.31

- (i)  $L$  is a regular language,
- (ii)  $L$  is recognized by some amalgamation system,
- (iii)  $L$  is a bounded language,
- (iv) There exists a finite set  $E \subseteq (\Sigma^*)^3$  such that  $L = \bigcup_{(x,u,y) \in E} \text{P}\downarrow(x)u\text{P}\downarrow(y)$ .

Combining Lemmas 54 and 65, we can conclude that the collection of infixes of the Thue-Morse sequence cannot be recognized by *any* amalgamation system.

To construct a decision procedure for well-quasi-orderedness under  $\sqsubseteq_{\text{infix}}$ , we need our amalgamation systems to satisfy certain *effectiveness assumptions*. We require that for an amalgamation system  $(\Sigma, R, E, \text{can})$ ,  $R$  is recursively enumerable, the function  $\text{can}(\cdot)$  is computable, and for any two runs  $\rho, \sigma \in R$ , the set  $E(\rho, \sigma)$  is computable. Additionally, we require the class to be effectively closed under rational transductions [?, Chapter 5, page 64].

Under these assumptions, one can transform the inclusion test of Equation (1) of Theorem 42 into an effective procedure, using pumping arguments from [?, Section 4.2], which, in turn, allows us to prove Theorem 62. Since the class  $\mathcal{C}_{\text{aut}}$  of regular languages and the class  $\mathcal{C}_{\text{cfg}}$  of context-free languages are examples of effective amalgamative classes, the following corollary is immediate.

**Corollary 66** *Let  $\mathcal{C} \in \{\mathcal{C}_{\text{aut}}, \mathcal{C}_{\text{cfg}}\}$ . It is decidable whether a language in  $\mathcal{C}$  is well-quasi-ordered by the infix relation. Furthermore, whenever it is well-quasi-ordered by the infix relation, it is a bounded language.*

## 7 Conclusion

We have described the landscapes of well-quasi-ordered languages for the natural orderings on finite words: prefix, suffix, and infix relations. While the prefix and suffix relation exhibit very simple behaviours, the infix relation can encode many complex quasi-orders (and even simulate the subword ordering). In the case of languages that are described by simple computational models, or languages that are “structurally simple” (bounded languages, downwards closed languages), we

showed that only very simple well-quasi-orders can be obtained: they are essentially isomorphic to disjoint unions of copies of finite sets,  $(\mathbb{N}, \leq)$ , and  $(\mathbb{N}^2, \leq)$ . Finally, under effectiveness assumptions on the language (such as being recognized by an amalgamation system, or being the set of infixes of an automatic sequence), we proved the decidability of being well-quasi-ordered for the infix relation. We believe that these very encouraging results pave the way for further research on deciding which sets are well-quasi-ordered for other orderings. Let us now discuss some possible research directions and remarks.

*Towards infinite alphabets* In this paper, we restricted our attention to *finite* alphabets, having in mind the application to regular languages. However, the conclusions of Theorem 42, Corollary 57, and Theorem 35 could be conjectured to hold in the case of infinite alphabets (themselves equipped with a well-quasi-ordering). This would require new techniques, as the finiteness of the alphabet is crucial to all of our positive results.

*Monoid equations* It could be interesting to understand which monoids  $M$  recognize languages that are well-quasi-ordered by the infix, prefix or suffix relations. This research direction is connected to finding which classes of graphs of *bounded clique-width* are well-quasi-ordered with respect to the *induced subgraph relation*, as shown in [?], and recently revisited in [?].

*Lexicographic orderings* There is another natural ordering on words, the *lexicographic ordering*, which does not fit well in our current framework because it is always of ordinal width 1. However, the order-type of the lexicographic ordering over regular languages has already been investigated in the context of infinite words [?], and it would be interesting to see if one can extend these results to decide whether such an ordering is well-founded for languages recognized by amalgamation systems.

*Factor Complexity* Let us conclude this section with a few remarks on the notion of factor complexity of languages. Recall that the *factor complexity* of a language  $L \subseteq \Sigma^*$  is the function  $f_L : \mathbb{N} \rightarrow \mathbb{N}$  such that  $f_L(n)$  is the number of distinct words of size  $n$  in  $L$ . We extend the notion of factor complexity to finite, infinite, and bi-infinite words as the factor complexity of their set of finite infixes. For the prefix relation and the suffix relation, all well-quasi-ordered languages have a bounded factor complexity, since they are finite unions of chains.

While there clearly are languages with low factor complexity that are not well-quasi-ordered for the infix relation, such as the language  $L \triangleq \downarrow ab^*a$ ; one would expect that languages that are well-quasi-ordered for the infix relation would have a low factor complexity.

In some sense, our results confirm this intuition in the case of languages described by a simple computational model. For languages recognized by amalgamation systems, being well-quasi-ordered implies being a bounded language, and therefore being included in some finite union of languages of the form  $w_1^*w_2w_3^*$ . Hence, these languages have at most a quadratic factor complexity. This is also

701 the case for languages described as the infixes of a finite set of pairs of morphic  
702 sequences. Indeed, the factor complexity of a morphic sequence that is uniformly  
703 recurrent is linear [?, Theorem 24], therefore the factor complexity of a language  
704 given by sequence representation using morphic sequences is at most quadratic.

705 However, there are downwards closed languages that are well-quasi-ordered  
706 for the infix relation but have an exponential factor complexity: the  $(5, 3)$ -  
707 Toeplitz word is uniformly recurrent [?, p. 499], and has exponential factor  
708 complexity [?, Theorem 5]. This shows that our computational models somehow  
709 fail to capture vast classes of well-quasi-ordered languages with a high factor  
710 complexity. It would be interesting to understand which new proof techniques  
711 would be required to obtain decidability for these languages.

## 712 A Proofs for Section 1

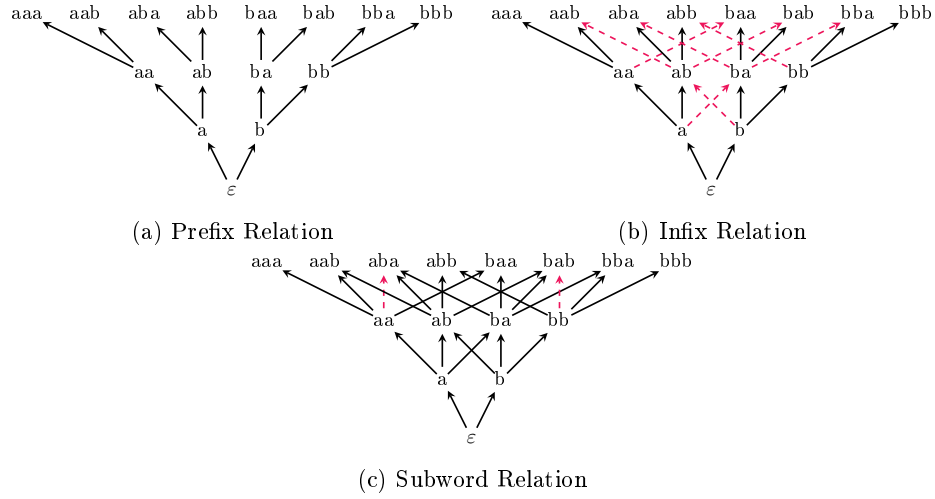


Fig. 3: A simple representation of the subword relation, prefix relation, and infix relation, on the alphabet  $\{a, b\}$  for words of length at most 3. The figures are Hasse Diagrams, representing the successor relation of the order. Furthermore, we highlight in dashed red relations that are added when moving from the prefix relation to the infix one, and to the infix relation to the subword one.

713 **B Proofs for Section 3**

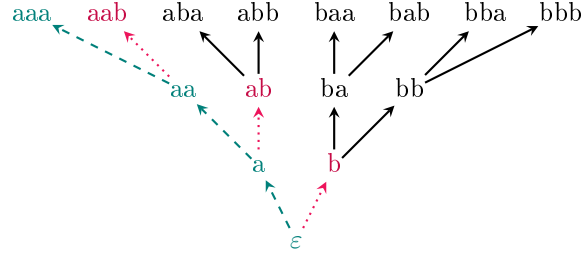


Fig. 4: An antichain branch for the language  $a^*b$ , represented in the tree of prefixes over the alphabet  $\{a, b\}$ . The branch is represented with dashed lines in turquoise, and the antichain is represented in dotted lines in blood-red.

714 *Proof (Proof of Lemma 33 as stated on page 7).* Assume that  $L$  contains an  
 715 antichain branch. Let us construct an infinite antichain as follows. We start with  
 716 a set  $A_0 \triangleq \emptyset$  and a node  $v_0$  at the root of the tree. At step  $i$ , we consider a word  
 717  $w_i$  such that  $v_i$  is a prefix of  $w_i$ , and  $w_i \in L \setminus B$ , which exists by definition of  
 718 antichain branches. We then set  $A_{i+1} \triangleq A_i \cup \{w_i\}$ . To compute  $v_{i+1}$ , we consider  
 719 the largest prefix of  $w_i$  that belongs to  $B$ , and set  $v_{i+1}$  to be the successor of  
 720 this prefix in  $B$ . By an immediate induction, we conclude that for all  $i \in \mathbb{N}$ ,  $A_i$   
 721 is an antichain, and that  $v_i$  is a node in the antichain branch  $B$  such that  $v_i$  is  
 722 not a prefix of any word in  $A_i$ .

723 Conversely, assume that  $L$  contains an infinite antichain  $A$ . Let us construct  
 724 an antichain branch. Let us consider the subtree of the tree of prefixes that  
 725 consists in words that are prefixes of words in  $A$ . This subtree is infinite, and by  
 726 König's lemma, it contains an infinite branch. By definition this is an antichain  
 727 branch.

▷ Back to p.7

728 *Proof (Proof of Corollary 34 as stated on page 7).* If  $L$  is regular, then it is  
 729 MSO-definable, and there exists a formula  $\varphi(x)$  in MSO that selects nodes of  
 730 the tree of prefixes that belong to  $L$ . Now, to decide whether there exists an  
 731 antichain branch for  $L$ , we can simply check whether the following formula is  
 732 satisfied:

$$\exists B. B \text{ is a branch} \wedge \forall x \in B, \exists y. y \text{ is a child of } x \wedge \varphi(y) \wedge y \notin B.$$

733 Because the above formula is an MSO-formula over the infinite  $\Sigma$ -branching tree,  
 734 whether it is satisfied is decidable as an easy consequence of the decidability of  
 735 MSO over infinite binary trees [?, Theorem 1.1].

▷ Back to p.7

736 *Proof (Proof of Theorem 35 as stated on page 7).* Assume that  $L$  is a finite  
 737 union of chains. Because the prefix relation is well-founded, and that finite unions  
 738 of chains have finite antichains, we conclude that  $L$  is well-quasi-ordered.

Conversely, assume that  $L$  is well-quasi-ordered by the prefix relation. Let us define  $S_{\text{split}}$  the set of words  $w \in \Sigma^*$  such that there exists two words  $wu$  and  $wv$  both in  $L$  that are not comparable for the prefix relation. Let  $S = S_{\text{split}} \cup \min_{\subseteq_{\text{pref}}} L$ . Assume by contradiction that  $S$  is infinite. Then,  $S$  equipped with the prefix relation is an infinite tree with finite branching, and therefore contains an infinite branch, which is by definition an antichain branch for  $L$ . This contradicts the assumption that  $L$  is well-quasi-ordered.

Now, let  $w$  be a maximal element for the prefix ordering in  $S$ . The upward closure of  $w$  in  $L$ ,  $(\uparrow_{\subseteq_{\text{pref}}} w) \cap L$ , must be a finite union of chains. Otherwise at least two of the chains would share a common prefix in  $w\Sigma$ , contradicting the maximality of  $w$ .

In particular, letting  $S_{\text{max}}$  be the set of all maximal elements of  $S$ , we conclude that

$$L \subseteq S \cup \bigcup_{w \in S_{\text{max}}} (\uparrow_{\subseteq_{\text{pref}}} w) \cap L \quad .$$

Hence, that  $L$  is finite union of chains.

*Proof (Proof of Corollary 34 as stated on page 7).* If  $L$  is regular, then it is MSO-definable, and there exists a formula  $\varphi(x)$  in MSO that selects nodes of the tree of prefixes that belong to  $L$ . Now, to decide whether there exists an antichain branch for  $L$ , we can simply check whether the following formula is satisfied:

$$\exists B. B \text{ is a branch} \wedge \forall x \in B, \exists y. y \text{ is a child of } x \wedge \varphi(y) \wedge y \notin B \quad .$$

Because the above formula is an MSO-formula over the infinite  $\Sigma$ -branching tree, whether it is satisfied is decidable as an easy consequence of the decidability of MSO over infinite binary trees [?, Theorem 1.1].

761 **C Proofs for Section 4**

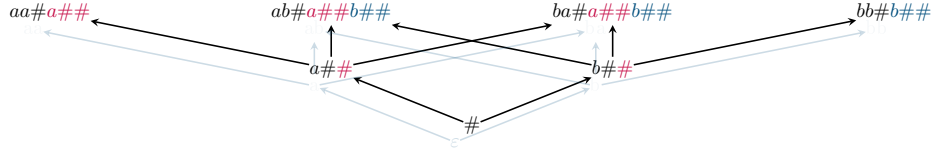


Fig. 5: Representation of the subword relation for  $\{a, b\}^*$  inside the infix relation for  $\{a, b, \#\}^*$  using a simplified version of Lemma 41, restricted to words of length at most 3.

762 *Proof (Proof of Lemma 44 as stated on page 9).* Let  $x \in \Sigma^+$  be a word, and  
 763 let  $P_x$  be the (finite) set of all prefixes of  $x$ , and  $S_x$  be the (finite) set of all  
 764 suffixes of  $x$ . Assume that  $w \in \text{Pl}(x)$ , then  $w = ux^pv$  for some  $u \in S_x$ ,  $v \in P_x$ ,  
 765 and  $p \in \mathbb{N}$ . We have proven that

$$\text{Pl}(x) \subseteq \bigcup_{u \in P_x} \bigcup_{v \in S_x} ux^*v \quad .$$

766 Let us now demonstrate that for all  $(u, v) \in S_x \times P_x$ , the language  $ux^*v$  is a  
 767 chain for the infix, suffix and prefix relations. To that end, let  $(u, v) \in S_x \times P_x$  and  
 768  $\ell, k \in \mathbb{N}$  be such that  $\ell < k$ , let us prove that  $ux^\ell v \sqsubseteq_{\text{infix}} ux^k v$ . Because  $v \sqsubseteq_{\text{pref}} x$ ,  
 769 we know that there exists  $w$  such that  $vw = x$ . In particular,  $ux^\ell vw = ux^{\ell+1}$ ,  
 770 and because  $\ell < k$ , we conclude that  $ux^{\ell+1} \sqsubseteq_{\text{pref}} ux^k v$ . By transitivity,  $ux^\ell v \sqsubseteq_{\text{pref}}$   
 771  $ux^k v$ , and *a fortiori*,  $ux^\ell v \sqsubseteq_{\text{infix}} ux^k v$ . Similarly, because  $u \sqsubseteq_{\text{suff}} x$ , there exists  
 772  $w$  such that  $wu = x$ , and we conclude that  $ux^\ell v \sqsubseteq_{\text{suff}} wux^\ell v = x^{\ell+1}v \sqsubseteq_{\text{suff}} ux^k v$ . ▷ Back to p.9

773

774 *Proof (Proof of Lemma 46 as stated on page 9).* Note that the result is obvious  
 775 if  $k = 0$ , and therefore we assume  $k \geq 1$  in the following proof.

776 Let us construct a sequence of words  $(w_i)_{i \in \mathbb{N}}$ , where  $w_i \triangleq w[\mathbf{n}_i]$  for some  
 777 well-chosen indices  $\mathbf{n}_i \in \mathbb{N}^k$ . The goal being that if  $w[\mathbf{n}_i]$  is an infix of  $w[\mathbf{n}_j]$ ,  
 778 then it can intersect at most *two* iterated words, with an intersection that is  
 779 long enough to successfully apply Lemma 45. In order to achieve this, let us first  
 780 define  $s$  as the maximal size of a word  $v_i$  ( $1 \leq i \leq k$ ) and  $u_j$  ( $1 \leq j \leq k+1$ ).  
 781 Then, we consider  $\mathbf{n}_0 \in \mathbb{N}^k$  such that  $\mathbf{n}_0$  has all its components greater than  
 782  $s!$  and such that  $w[\mathbf{n}_0]$  belongs to  $L$ . Then, we inductively define  $\mathbf{n}_{i+1}$  as the  
 783 smallest vector of numbers greater than  $\mathbf{n}_i$ , such that  $w[\mathbf{n}_{i+1}]$  belongs to  $L$ , and  
 784 with  $\mathbf{n}_i$  having all components greater than  $2|w[\mathbf{n}_i]|$ .

785 Let us assume that  $k \geq 2$  in the following proof for symmetry purposes, and  
 786 argue later on that when  $k = 1$  the same argument goes through. Because  $L$  is  
 787 well-quasi-ordered by the infix relation, there exists  $i < j$  such that  $w[\mathbf{n}_i]$  is an  
 788 infix of  $w[\mathbf{n}_j]$ . Now, because of the chosen values for  $\mathbf{n}_j$ , there exists  $1 \leq \ell \leq k-1$   
 789 such that one of the three following equations holds:

$$\begin{aligned}
790 \quad & - w[\mathbf{n}_i] \sqsubseteq_{\text{infix}} v_\ell^{n_{j,\ell}} u_{\ell+1} v_{\ell+1}^{n_{j,\ell+1}}, \\
791 \quad & - w[\mathbf{n}_i] \sqsubseteq_{\text{infix}} u_\ell v_\ell^{n_{j,\ell}}, \\
792 \quad & - w[\mathbf{n}_i] \sqsubseteq_{\text{infix}} v_\ell^{n_{j,\ell}} u_{\ell+1}.
\end{aligned}$$

793 In the sake of simplicity, we will only consider one of the three cases, namely  
794  $w[\mathbf{n}_i] \sqsubseteq_{\text{infix}} v_\ell^{n_{j,\ell}} u_{\ell+1}$ , the other two being similar. Because the lengths used in  
795  $\mathbf{n}_i$  are all sufficiently large, we know that for every  $k$ ,  $v_k^{n_{i,k}}$  is an infix of a  $v_\ell^p$   
796 for some non-zero  $p$ . Therefore, we can apply Lemma 45 to conclude that there  
797 exists a word  $x$  such that every  $v_k$  is a power of a conjugate of  $x$  (a cyclic shift of  
798  $x$ ), and  $v_\ell$  is a power of  $x$ . We can therefore rewrite  $w[\mathbf{n}_i]$  as  $u_1(\sigma_1(x))^{n_{i,1}} u_2 \dots$ ,  
799 where  $\sigma_k$  is some conjugacy operation (cyclic shift). Now, in order for  $w[\mathbf{n}_i]$  to  
800 be an infix of  $x^{p \times n_{j,\ell}} u_{\ell+1}$ , we must conclude that all the  $u_k$ 's are suffixes or  
801 prefixes of  $x$ , and that they align properly with the  $\sigma_k(x)$ 's to form an infix  
802 of some power of  $x$ , except for the last one. In particular,  $w[\mathbf{n}_i] \in \text{Pl}(x) u_{\ell+1}$ ,  
803 but also, every other choice of  $\mathbf{n}$  will lead to a word in  $\text{Pl}(x) u_{\ell+1}$ , because the  
804 alignment constraints are stable under pumping.

805 In the case of two iterated words, the reasoning is similar, distinguishing  
806 between the  $v_i$ 's that are occurring before and after the junction of the two  
807 iterated words.

808 When  $k = 1$ , the situation is a bit more specific since we only have two  
809 cases: either  $w_i \sqsubseteq_{\text{infix}} u_1 v_1^{n_j}$  or  $w_i \sqsubseteq_{\text{infix}} v_1^{n_j} u_2$ , and we conclude with an identical  
810 reasoning.

811 *Proof (Proof of Lemma 47 as stated on page 9).* Let  $w_1, \dots, w_n$  be such that  
812  $L \subseteq w_1^* \dots w_n^*$ . Let us define  $m \triangleq \max\{|w_i| \mid 1 \leq i \leq n\}$

813 Let  $w[\mathbf{k}] \triangleq w_1^{k_1} \dots w_n^{k_n}$  be a map from  $\mathbb{N}^k$  to  $\Sigma^*$ . We are interested in the  
814 intersection of the image of  $w$  with  $L$ . Let us assume for instance that for all  
815  $\mathbf{k} \in \mathbb{N}^n$ , there exists  $\ell \geq \mathbf{k}$  such that  $w[\ell] \in L$ . Then, leveraging Lemma 46, we  
816 conclude that there exists  $x, y$  of size at most  $\max\{|w_i| \mid 1 \leq i \leq n\}$  such that  
817  $w[\mathbf{k}] \in \text{Pl}(x) \cup \text{Pl}(x) \text{Pl}(y)$ , and we conclude that  $L \subseteq \text{Pl}(x) \cup \text{Pl}(x) \text{Pl}(y)$ .

818 Now, it may be the case that one cannot simultaneously assume that two  
819 component of the vector  $\mathbf{k}$  are unbounded. In general, given a set  $S \subseteq \{1, \dots, n\}$   
820 of indices, we say that  $S$  is admissible if there exists a bound  $N_0$  such that for  
821 all  $\mathbf{b} \in \mathbb{N}^S$ , there exists a vector  $\mathbf{k} \in \mathbb{N}^n$ , such that  $\mathbf{k}$  is greater than  $\mathbf{b}$  on the  $S$   
822 components, and the other components are below the bound  $N_0$ . The language  
823 of an admissible set  $S$  is the set of words obtained by repeating  $w_i$  at most  $N_0$   
824 times if it is not in  $S$  ( $w_i^{\leq N_0}$ ) and arbitrarily many times otherwise ( $w_i^*$ ). Note  
825 that  $L \subseteq \bigcup_{S \text{ admissible}} L(S)$ .

826 Now, admissible languages are ready to be pumped according to Lemma 46.  
827 For every admissible language, the size of a word that is not iterated is at most  
828  $N_0 \times m$  by definition, and we conclude that:

$$L \subseteq \bigcup_{x, y \in \Sigma^{\leq n}} \bigcup_{u \in \Sigma^{\leq m \times N_0}} \text{Pl}(x) u \text{Pl}(y) \cup \text{Pl}(x) u \cup u \text{Pl}(x) \quad . \quad (1)$$



## 830 D Proofs for Section 5

831 *Proof (Proof of Corollary 51 as stated on page 10).* Because  $L \subseteq \downarrow_{\sqsubseteq_{\text{infix}}} L$ , the  
 832 right-to-left implication is trivial. For the left-to-right implication, let us assume  
 833 that  $L$  is a well-quasi-ordered language for the infix relation. Then  $L$  is included  
 834 in a finite union of products of chains for the prefix and suffix relations thanks  
 835 to Theorem 42:

$$L \subseteq \bigcup_{i=1}^n S_i \cdot P_i \quad .$$

836 Remark that if  $S_i$  is a chain for the suffix relation and  $P_i$  is a chain for the prefix  
 837 relation, then

$$\downarrow_{\sqsubseteq_{\text{infix}}} (S_i \cdot P_i) = (\downarrow_{\sqsubseteq_{\text{suffix}}} S_i) \cdot (\downarrow_{\sqsubseteq_{\text{prefix}}} P_i) \quad .$$

838 Indeed, any infix of a word in  $S_i P_i$  can be split into a suffix of a word in  $S_i$  and a  
 839 prefix of a word in  $P_i$ . Conversely, any such concatenations are infixes of a word  
 840 in  $S_i P_i$ .

841 As a consequence, we conclude that  $\downarrow_{\sqsubseteq_{\text{infix}}} L$  is itself included in a finite union  
 842 of products of chains. Furthermore, by definition of bounded languages,  $\downarrow_{\sqsubseteq_{\text{infix}}} L$   
 843 is also a bounded language. Hence, it is well-quasi-ordered by the infix relation  
 844 via Theorem 42.

▷ Back to p.10

845 *Proof (Proof of Lemma 58 as stated on page 12).* Let us assume that  $L$  is  
 846 infinite. The case when it is finite is similar, but will result in a finite word.

847 Because the alphabet  $\Sigma$  is finite, we can enumerate the words of  $L$  as  $(w_i)_{i \in \mathbb{N}}$ .  
 848 From  $(w_i)_{i \in \mathbb{N}}$ , we construct a sequence  $(u_i)_{i \in \mathbb{N}}$  by induction as follows:  $u_0 = w_0$ ,  
 849 and  $u_{i+1}$  is a word that contains  $u_i$  and  $w_i$ , which exists in  $L$  because  $L$  is  
 850 directed. Since  $L$  is well-quasi-ordered, one can extract an infinite set of indices  
 851  $I \subseteq \mathbb{N}$  such that  $u_i \sqsubseteq_{\text{infix}} u_j$  for all  $i \leq j \in I$ .

852 We can build a word  $w$  as the limit of the sequence  $(u_i)_{i \in I}$ . This word is  
 853 infinite or bi-infinite, and contains as infixes all the words  $u_i$  for  $i \in I$ . Because  
 854 every word of  $L$  is an infix of every  $u_i$  for a large enough  $I$ , one concludes that  
 855  $L$  is contained in the set of finite infixes of  $w$ . Conversely, every finite infix of  $w$   
 856 is an infix of some  $u_i$  by definition of the limit construction, hence belongs to  $L$   
 857 since  $u_i \in L$  and  $L$  is downwards closed.

aliaume: do we  
need wqo here?  
the proof should  
go through with-  
out it: the se-  
quence  $u_i$  is al-  
ready increasing  
for infix

▷ Back to p.12

858 *Proof (Proof of Lemma 59 as stated on page 12).*

859 Assume that  $w$  is ultimately uniformly recurrent. Consider a sequence of  
 860 words  $(w_i)_{i \in \mathbb{N}}$  that are finite infixes of  $w$ . Because  $w$  is ultimately uniformly  
 861 recurrent, there exists a bound  $N_0$  such that  $w_{\geq N_0}$  is uniformly recurrent. Let  
 862  $i < N_0$ , we claim that, without loss of generality, only finitely many words in  
 863 the sequence  $(w_i)_{i \in \mathbb{N}}$  can be found starting at the position  $i$  in  $w$ . Indeed, if  
 864 it is not the case, then we have an infinite subsequence of words that are all  
 865 comparable for the infix relation, and therefore a good sequence, because the  
 866 infix relation is well-founded. We can therefore assume that all words in the  
 867 sequence  $(w_i)_{i \in \mathbb{N}}$  are such that they start at a position  $i \geq N_0$ . But then they  
 868 are all finite infixes of  $w_{\geq N_0}$ , which is a uniformly recurrent word, whose set of  
 869 finite infixes is well-quasi-ordered (Theorem 53).

Conversely, assume that the set of finite infixes of  $w$  is well-quasi-ordered. Let us write  $\text{Rec}(w)$  the set of finite infixes of  $w$  that appear infinitely often. We can similarly define  $\text{Rec}(w_{\geq i})$  for any (infinite) suffix of  $w$ . The sequence  $R_i \triangleq \text{Rec}(w_{\geq i})$  is a descending sequence of downwards closed sets of finite words, included in the set of finite infixes of  $w$  by definition. Because the latter is well-quasi-ordered, there exists an  $N_0 \in \mathbb{N}$ , such that  $\bigcap_{i \in \mathbb{N}} R_i = R_{N_0}$ . Now, consider  $v \triangleq w_{\geq N_0}$ . By construction, every finite infix of  $v$  appears infinitely often in  $v$ . Given some finite infix  $u \sqsubseteq_{\text{infix}} v$ , we there exists a bound  $N_u$  on the distance between two consecutive occurrences of  $u$  in  $v$ . Indeed, if it is not the case, then there exists an infinite sequence  $(ux_iu)_{i \in \mathbb{N}}$  of infixes of  $v$ , such that  $x_i$  is a word of size  $\geq i$  and no shorter word  $uyu$  is an infix of  $ux_iu$ . Because the finite infixes of  $w$  (hence, of  $v$ ) are well-quasi-ordered, one can extract an infinite set of indices  $I \subseteq \mathbb{N}$  such that  $ux_iu \sqsubseteq_{\text{infix}} ux_ju$  for all  $i \leq j \in I$ . In particular,  $ux_iu \sqsubseteq_{\text{infix}} ux_ju$  for some  $j > |x_i|$ , which contradicts the fact that  $ux_ju$  coded two consecutive occurrences of  $u$  in  $v$ .

We have shown that for every finite infix  $u$  of  $v$ , there exists a bound  $N_u$  such that every two occurrences of  $u$  in  $v$  start at distance at most  $N_u$ . In particular, there exists a bound  $M_u$  such that every infix of  $v$  of size at least  $M_u$  contains  $u$ . We have proven that  $v$  is uniformly recurrent, hence that  $w$  is ultimately uniformly recurrent.

*Proof (Proof of Lemma 510 as stated on page 13).* Given a bi-infinite word  $w \in \Sigma^{\mathbb{Z}}$ , we can consider  $w_+ \in \Sigma^{\mathbb{N}}$  and  $w_- \in \Sigma^{\mathbb{N}}$  the two infinite words obtained as follows: for all  $i \in \mathbb{N}$ ,  $(w_+)_i = w(i)$  and  $(w_-)_i = w(-i)$ . Note that the two share the letter at position 0.

Assume that  $w_+$  and  $w_-$  are ultimately uniformly recurrent. Let us write  $\text{Infixes}(w)$  the set of finite infixes of  $w$ . Consider an infinite sequence of words  $(u_i)_{i \in \mathbb{N}}$  in  $\text{Infixes}(w)$ . If there is an infinite subsequence of words that are all in  $\text{Infixes}(w_+)$ , then there exists an increasing pair of indices  $i < j$  such that  $u_i \sqsubseteq_{\text{infix}} u_j$  because Theorem 53 applies to  $w_+$ . Similarly, if there is an infinite subsequence of words that are all in  $\text{Infixes}(w_-)$ , then there exists an increasing pair of indices  $i < j$  such that  $u_i \sqsubseteq_{\text{infix}} u_j$  because Theorem 53 applies to  $w_-$  (and the infix relation is compatible with mirroring). Otherwise, one can assume without loss of generality that all words in the sequence have a starting position in  $w_-$  and an ending position in  $w_+$ . In this case, let us write  $(k_i, l_i) \in \mathbb{N}^2$  the pair of indices such that  $u_i$  is the infix of  $w$  that starts at position  $-k_i$  of  $w$  (i.e.,  $k_i$  of  $w_-$ ) and ends at position  $l_i$  of  $w$  (i.e.,  $l_i$  of  $w_+$ ). Because  $\mathbb{N}^2$  is a well-quasi-ordering with the product ordering, there exists  $i < j$  such that  $k_i \leq k_j$  and  $l_i \leq l_j$ , in particular,  $u_i \sqsubseteq_{\text{infix}} u_j$ . We have proven that every infinite sequence of words in  $\text{Infixes}(w)$  is good, hence  $\text{Infixes}(w)$  is well-quasi-ordered.

Conversely, assume that  $\text{Infixes}(w)$  is well-quasi-ordered. In particular, the subset  $\text{Infixes}(w_+) \subseteq \text{Infixes}(w)$  is well-quasi-ordered. Similarly,  $\text{Infixes}(w_-)$  is well-quasi-ordered because the infix relation is compatible with mirroring. Applying Lemma 59, we conclude that both are ultimately uniformly recurrent words.

914 *Proof (Proof of Lemma 512 as stated on page 13).* Given a bi-infinite word  
 915  $w \in \Sigma^{\mathbb{Z}}$ , recall that we can consider  $w_+ \in \Sigma^{\mathbb{N}}$  and  $w_- \in \Sigma^{\mathbb{N}}$  the two infinite  
 916 words obtained as follows: for all  $i \in \mathbb{N}$ ,  $(w_+)_i = w(i)$  and  $(w_-)_i = w(-i)$ . Note  
 917 that the two share the letter at position 0.

918 To obtain the upper bound of  $\omega \cdot 3$ , we can consider the same argument  
 919 as for Lemma 511. We let  $N_0$  be such that  $w_{\geq N_0}$  and  $(w_-)_{\geq N_0}$  are uniformly  
 920 recurrent words. In any sequence of incomparable elements of  $\text{Infixes}(w)$ , there  
 921 are less than  $N_0^2$  elements that are found in  $(w_{<N_0})_{>-N_0}$ . Then, one has to pick  
 922 a finite infix in either  $w_{\geq N_0}$  or  $w_{\leq -N_0}$ . Because of Lemma 511, any sequence  
 923 of incomparable elements of these two infinite words has length bounded based  
 924 on the choice of the first element of that sequence. This means that the ordinal  
 925 width of  $\text{Infixes}(w)$  is at most  $\omega + \omega + N_0^2$ . We conclude that  $\text{w}(\text{Infixes}(w)) < \omega \cdot 3$ .

926 Let us briefly argue that the bound is tight. Indeed, one can construct a bi-  
 927 infinite word  $w$  by concatenating a reversed Thue-Morse sequence on a binary  
 928 alphabet  $\{a, b\}$ , a finite antichain of arbitrarily large size over a distinct alphabet  
 929  $\{c, d\}$ , and then a Thue-Morse sequence on a binary alphabet  $\{e, f\}$ . The ordinal  
 930 width of the set of infixes of  $w$  is then at least  $\omega \cdot 2 + K$ , where  $K$  is the size of the  
 931 chosen antichain, following the same argument as in the proof of Lemma 511,  
 932 using Lemma 55. ▷ Back to p.13

933 **Lemma D1** *Given an automatic sequence  $w \in \Sigma^{\mathbb{N}}$ , one can decide whether it* ▷ Proven p.27  
 934 *is ultimately uniformly recurrent.*

935 *Proof (Proof of Lemma D1 as stated on page 27).* We can rewrite this as a  
 936 question on the automatic sequence  $w$  as follows:

|  |                                   |
|--|-----------------------------------|
| $\exists N_0,$   | ultimately                        |
| $\forall i_s \geq N_0,$  | for every infix (start) $u$       |
| $\forall i_e > i_s,$   | for every infix (end) $u$         |
| $\exists k \geq 1,$  | there exists a bound              |
| $\forall j_s \geq N_0,$  | for every other infix (start) $v$ |
| $\forall j_e \geq j_s + k,$                                    | of size at least $k$              |
| $\exists l \geq 0,$  | there exists a position in $v$    |
| $\forall 0 \leq m < i_e - i_s,$                                | where $u$ can be found            |
| $j_s + m + l < j_e \wedge w(i_s + m) = w(j_s + m + l) \quad .$ |                                   |

937 Because  $w$  is computable by a finite automaton, one can reduce the above formula  
 938 to a regular language, for which it suffices to check emptiness, which is decidable.  
 939 ▷ Back to p.27

940 *Proof (Proof of Lemma 511 as stated on page 13).* Let  $N_0$  be a bound such  
 941 that  $w_{\geq N_0}$  is uniformly recurrent. Let us write  $\text{Infixes}(w)$  the set of finite infixes  
 942 of  $w$ . We prove that  $\text{w}(\text{Infixes}(w)) \leq \omega + N_0$ . Let  $u_1 \sqsubseteq_{\text{infix}} w$  be a finite word.

943 If  $u_1$  is an infix of  $w_{\geq N_0}$ , then there exists  $k \geq 1$  such that  $u_1$  is an infix of  
 944 every word of size at least  $k$ . In particular, there is finite bound on the length

of every sequence of incomparable elements starting with  $u_1$ . We conclude in particular that  $\text{Infixes}(w) \setminus \uparrow u_1$  has a finite ordinal width.

Otherwise,  $u_1$  can only be found *before*  $N_0$ . In this case, we consider a second element of a bad sequence  $u_2 \sqsubseteq_{\text{infix}} w$ , which is incomparable with  $u_1$  for the infix relation. If  $u_2$  is an infix of  $w_{\geq N_0}$ , then we can conclude as before. Otherwise, notice that  $u_1$  and  $u_2$  cannot start at the same position in  $w$  (because they are incomparable). Continuing this argument, we conclude that there are at most  $N_0$  elements starting before  $N_0$  at the start of any sequence of incomparable elements in  $\text{Infixes}(w)$ . We conclude that  $\mathfrak{w}(\text{Infixes}(w)) \leq \omega + N_0$ .

Let us now justify that this bound is tight. The Thue-Morse sequence over a binary alphabet  $\{a, b\}$  has ordinal width  $\omega$  from Lemma 55. Given a number  $N_0 \in \mathbb{N}$ , one can construct an arbitrarily long antichain of words for the infix relation by using a new letter  $c$ . When concatenating this (finite) antichain as a prefix of the Thue-Morse sequence, one obtains a new (infinite) word  $w$ . It is clear that the ordinal width of  $\text{Infixes}(w)$  is now at least  $\omega + N_0$ .

*Proof (Proof of Corollary 57 as stated on page 12).* It is always true that the ordinal height of a language over a finite alphabet is at most  $\omega$ . Let us now consider a well-quasi-ordered language  $L$  that is downwards closed for the infix relation. Applying Theorem 56, we can write  $L = \bigcup_{i=1}^n L_i$  where each  $L_i$  is the set of finite infixes of a finite, infinite or bi-infinite ultimately uniformly recurrent word  $w_i$ . We can then directly conclude that  $\mathfrak{w}(L_i)$  less than  $\omega$  (in the case of a finite word), less than  $\omega \cdot 2$  (in the case of an infinite word thanks to Lemma 511), or less than  $3 \cdot \omega$  (in the case of a bi-infinite word, thanks to Lemma 512). In any case, we have the bound  $\mathfrak{w}(L_i) < \omega \cdot 3$ .

Now,  $\mathfrak{w}(L) \leq \sum_{i=1}^n \mathfrak{w}(L_i) < \omega \cdot 3 < \omega^2$ . Finally, the inequality  $\mathfrak{o}(L) \leq \mathfrak{w}(L) \otimes \mathfrak{h}(L) < \omega \otimes \omega^2 = \omega^3$  allows us to conclude.

The tightness of the bounds is a direct consequence of Lemma 512, and by considering a finite union of these examples over disjoint alphabets (or even, by considering a binary alphabet and using unambiguous codes to separate the different components).

▷ Back to p.13

▷ Back to p.12

975 **E Proofs for Section 6**

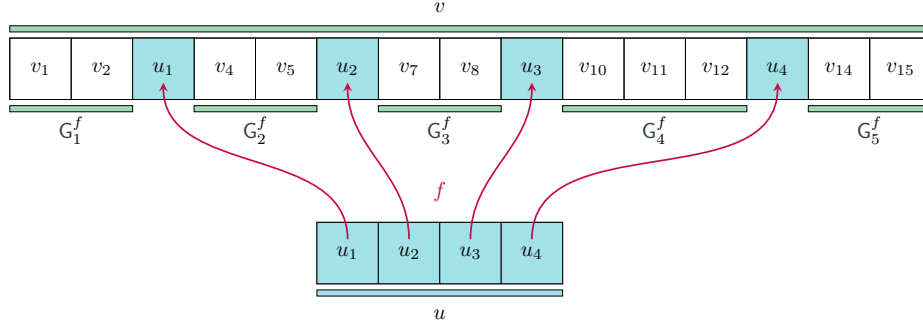


Fig. 6: The gap words resulting from a subword embedding between two finite words.

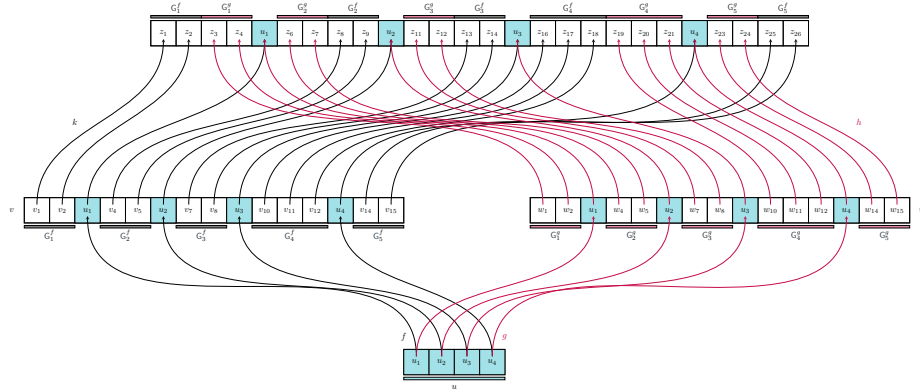


Fig. 7: We illustrate how embeddings  $f$  and  $g$  between runs of an amalgamation system can be glued together, seen on their canonical decomposition.

976 For this paper to be self-contained, we will also recall how runs of a finite  
 977 state automaton can be understood as an amalgamation system.

978 **Example E1 ([?, Section 3.2])** Let  $A = (Q, \delta, q_0, F)$  be a finite state automa-  
 979 tion over a finite alphabet  $\Sigma$ . Let  $\Delta$  be the set of transitions  $(q_1, a, q_2) \in Q \times \Sigma \times Q$ ,  
 980 and  $R \subseteq \Delta^*$  be the set of words over transitions that start with the initial state  
 981  $q_0$ , end in a final state  $q_f \in F$ , and such that the end state of a letter is the  
 982 start state of the following one. The canonical decomposition  $\text{can}$  is defined as

983 a morphism from  $\Delta^*$  to  $\Sigma^*$  that maps  $(q, a, p)$  to  $a$ . Because of the one-to-one  
 984 correspondence of steps of a run  $\rho$  and letters in its canonical decomposition, we  
 985 may treat the two interchangeably. Finally, given two runs  $\rho$  and  $\sigma$  of the au-  
 986 tomaton, we say that an embedding  $f \in \text{Hom}^*(\text{can}(\rho), \text{can}(\sigma))$  belongs to  $E(\rho, \sigma)$   
 987 when  $f$  is also defining an embedding from  $\rho$  to  $\sigma$  as words in  $\Delta^*$ .

988 The system  $(\Sigma, R, E, \text{can})$  is an amalgamation system, whose language is  
 989 precisely the language of words recognized by the automaton  $A$ .

990 *Proof.* By definition, the embeddings inside  $E(\rho, \sigma)$  are in of  $\text{Hom}^*(\text{can}(\rho), \text{can}(\sigma))$ ,  
 991 and they compose properly. Because  $\Delta = Q \times \Sigma \times Q$  is finite, it is a well-quasi-  
 992 ordering when equipped with the equality relation, and we conclude that  $\Delta^*$   
 993 with  $\leq^*$  is a well-quasi-order according to Higman's Lemma [?].

994 Let us now move to proving that the system satisfies the amalgamation  
 995 property. Given three runs  $\rho, \sigma, \tau \in R$ , and two embeddings  $f \in E(\rho, \sigma)$  and  
 996  $g \in E(\rho, \tau)$ , we want to construct an amalgamated run  $\sigma \vee \tau$ . Because letters in  
 997 the run  $\rho$  respect the transitions of the automaton (i.e., if the letter  $i$  ends in  
 998 state  $q$ , then the letter  $i + 1$  starts in state  $q$ ), then the gap word at position  $i$   
 999 starts in state  $q$  and ends in state  $q$  too. This means that for both embeddings  $f$   
 1000 and  $g$ , the gap words are read by the automaton by looping on a state. In par-  
 1001 ticular, these loops can be taken in any order and continue to represent a valid  
 1002 run. That is, we can even select the order of concatenation in the amalgamation  
 1003 for all  $0 \leq i \leq |\text{can}(\rho)|$  and not just for one separately.

1004 We conclude by remarking that the language of this amalgamation system  
 1005 is the set of  $\text{yield}(R)$ , because  $R$  is the set of valid runs of the automaton, and  
 1006  $\text{yield}(\rho)$  is the word read along a run  $\rho$ .

*Proof (Proof of Lemma 64 as stated on page 17).* Write  $u$  for  $G_\ell^f$  and  $v$  for  $G_\ell^g$ .  
 We may assume that both  $u$  and  $v$  are non-empty, as otherwise the lemma holds  
 trivially. Then, for all  $k \in \mathbb{N}$ , there exists a run with canonical decomposition

$$w_k = L_0 a_1 \cdots a_n L_n,$$

1007 where  $L_i \in \{vvu^k, vu^k v, u^k vv\}$  and specifically  $L_\ell = vu^k v$ .

1008 From Lemma 46, we may conclude that there are a finite number of words  
 1009  $x, y$ , and  $w$  such that each  $w_k$  is contained in a language  $P\downarrow(x)wP\downarrow(y)$ .

1010 As there is an infinite number of words  $w_k$ , we may fix  $x, y$ , and  $w$  and  
 1011 an infinite subset  $I \subseteq \mathbb{N}$  such that  $\{w_i \mid i \in I\} \subseteq P\downarrow(x)wP\downarrow(y)$ . This implies  
 1012 that either for infinitely many  $m \in \mathbb{N}$ ,  $u^m v \in P\downarrow(y)$  or for infinitely many  $m$ ,  
 1013  $vu^m \in P\downarrow(x)$ .

1014 In either case, we may conclude that either  $u \sqsubseteq_{\text{infix}} v$  or  $v \sqsubseteq_{\text{infix}} u$ : Let  $m, n \in$   
 1015  $\mathbb{N}$  such that  $m < n$  and  $u^m v, u^n v \in P\downarrow(y)$  (the case for  $vu^m$  and  $vu^n$  proceeding  
 1016 analogously). Without loss of generality, assume that  $|u^m|$  and  $|u^n|$  are multiples  
 1017 of  $|y|$ . We therefore find  $p \sqsubseteq_{\text{pref}} y, s \sqsubseteq_{\text{suff}} y$  such that  $u^m, u^n \in sy^*p$ , ergo  $ps = y$ .  
 1018 In other words, we can write  $u^m = (sp)^{m'}, u^n = (sp)^{n'}$ . As  $u^m v \in P\downarrow(y)$ , it  
 1019 follows that  $v$  is a prefix of some word in  $(sp)^*$ . Hence either  $v$  is a prefix of  $u$   
 1020 or  $u$  vice versa.

1021 *Proof (Proof of Lemma 65 as stated on page 17).* It is clear that Item i  $\Rightarrow$  Item ii  
 1022 because regular languages are recognized by finite automata, and finite automata  
 1023 are a particular case of amalgamation systems. The implication Item ii  $\Rightarrow$  Item iii  
 1024 is the content of Theorem 61. The implication Item iii  $\Rightarrow$  Item iv is Lemma 47.  
 1025 Finally, the implication Item iv  $\Rightarrow$  Item i is simply because a downwards closed  
 1026 language that is a finite union of products of chains is a regular language.

1027 Indeed, assume that  $L$  is downwards closed and included in a finite union  
 1028 of sets of the form  $P\downarrow(x)uP\downarrow(y)$  where  $x, y, u$  are possibly empty words. We can  
 1029 assume without loss of generality that for every  $n$ ,  $x^ny^n$  is in  $L$ , otherwise, we  
 1030 have a bound on the maximal  $n$  such that  $x^ny^n$  is in  $L$ , and we can increase  
 1031 the number of languages in the union, replacing  $x$  or  $y$  with the empty word  
 1032 as necessary. Let us write  $L' \triangleq \bigcup_{i=1}^k x_i^* u_i y_i^*$ . Then,  $L' \subseteq L$  by construction.  
 1033 Furthermore,  $L \subseteq \downarrow L'$ , also by construction. Finally, we conclude that  $L = \downarrow L'$   
 1034 because  $L$  is downwards closed. Now, because  $L'$  is a regular language, and  
 1035 regular languages are closed under downwards closure, we conclude that  $L$  is a  
 1036 regular language. ▷ Back to p.17

1037 *Proof (Proof of Theorem 61 as stated on page 15).* Assume that  $L$  is well-  
 1038 quasi-ordered by the infix relation, and obtained by an amalgamation system  
 1039  $(\Sigma, R, E, \text{can})$ .

1040 Let us consider the set  $M$  of minimal runs for the relation  $\leq_E$ , which is  
 1041 finite because the latter is a well-quasi-ordering. By Lemma 64, we know that  
 1042 for each minimal run  $\rho \in M$ , each gap language  $L_i^\rho$  of  $\rho$  is totally ordered by  
 1043  $\sqsubseteq_{\text{infix}}$ . Adapting the proof of language boundedness from [?, Section 4.2], we may  
 1044 conclude that  $L_i^\rho \subseteq P\downarrow(w)$  for some  $w \in L_i^\rho$ . As  $P\downarrow(w)$  is language bounded and  
 1045 this property is stable under subsets, concatenation and finite union, we can  
 1046 conclude that  $L$  is bounded as well. ▷ Back to p.15

1047 ⌈ Let us briefly recall that a rational transduction is a relation  $R \subseteq \Sigma^* \times \Gamma^*$   
 1048 such that there exists a finite state automaton that reads pairs of letters  $(a, b) \in$   
 1049  $(\Sigma \cup \{\epsilon\}) \times (\Gamma \cup \{\epsilon\})$  and recognizes  $R$ . A class of languages  $\mathcal{C}$  is closed under  
 1050 rational transductions if for every  $L \in \mathcal{C}$  and every rational transduction  $R$ , the  
 1051 language  $R(L) \triangleq \{v \in \Gamma^* \mid \exists u \in L, (u, v) \in R\}$  also belongs to  $\mathcal{C}$ .

1052 *Proof (Proof of Theorem 62 as stated on page 15).* We first show Item 3  $\Rightarrow$   
 1053 Item 1. We aim to make the inclusion test of Equation (1) of Theorem 42 effec-  
 1054 tive. Let  $R(n, m, N_0) \triangleq \bigcup_{x, y \in \Sigma^{\leq n}} \bigcup_{u \in \Sigma^{\leq m \times N_0}} P\downarrow(x)uP\downarrow(y) \cup P\downarrow(x)u \cup uP\downarrow(x)$ .  
 1055 For any concrete values of the bounds  $n, m$ , and  $N_0$ , this language is regular. The  
 1056 map  $L \mapsto L \cap \Sigma^* \setminus R(n, m, N_0)$  is a rational transduction because  $\Sigma^* \setminus R(n, m, N_0)$   
 1057 is regular. Since  $\mathcal{C}$  is closed under rational transductions, we can therefore re-  
 1058 duce the inclusion to emptiness of this language. However, we need to find these  
 1059 bounds first.

1060 To determine values for  $n$  and  $m$ , we first test if  $L$  is bounded. Since emptiness  
 1061 is decidable, we can apply the algorithm in [?, Section 4.2] to decide if  $L$  is  
 1062 bounded. If  $L$  is bounded, this algorithm yields words  $w_1, \dots, w_n$  such that  $L \subseteq$   
 1063  $w_1^* \dots w_n^*$  and therefore yields also the bounds in questions:  $n$  is the number of  
 1064 words, and  $m$  is the maximal length of a word  $w_i$  where  $1 \leq i \leq n$ . If  $L$  is not

1065 bounded, then  $L$  cannot be well-quasi-ordered by the infix relation because of  
 1066 Theorem 61 and we immediately return false.

1067 To determine the value for  $N_0$ , we then compute the downward closure (with  
 1068 respect to subwords) of  $L$ . This is effective and yields a finite-state automaton.  
 1069 Recall that  $N_0$  is the maximum number of repetitions of a word  $w_i$  that can  
 1070 not be iterated arbitrarily often. This value is therefore bounded above by the  
 1071 length of the longest simple path in this automaton.

1072 Item 1  $\Rightarrow$  Item 2. We just consider the transduction  $f$  that maps every word  
 1073  $w$  to  $\#w$  where  $\#$  is a fresh symbol. Then, for any language  $L \in \mathcal{C}$ ,  $L$  is well-  
 1074 quasi-ordered by prefix if and only if  $f(L)$  is well-quasi-ordered by infix.

1075 Item 2  $\Rightarrow$  Item 3. We consider the transduction  $R \triangleq \Sigma^* \times \{a, b\}^*$ . Then for  
 1076 any language  $L \in \mathcal{C}$ , the image of  $L$  through  $R$  is well-quasi-ordered by prefix if  
 1077 and only if  $L$  is empty.