A syntactic congruence for languages of birooted trees

Achim Blumensath[1], David Janin[2]

[1]TU Darmstadt,
Mathematik, AG Logik,
D-64289 Darmstadt, GERMANY
blumensath@mathematik.tu-darmstadt.de

[2]Université de Bordeaux,
LaBRI UMR 5800,
F-33405 Talence, FRANCE
janin@labri.fr

Abstract. The study of languages of labelled birooted trees, that is, elements of the free inverse monoid enriched by a vertex labelling, has led to the notion of quasi-recognisability. It generalises the usual notion of recognisability by replacing homomorphisms by certain prehomomorphism into finite ordered monoids, called adequate, that only preserve *some* products: the so-called disjoint ones. In this paper we study the underlying *partial algebra* setting and we define a suitable notion of a syntactic congruence such that (i) having a syntactic congruence of finite index captures MSO-definability; (ii) a certain order-bisimulation refinement of the syntactic congruence captures quasi-recognisability in the same way.

1 Introduction

When modelling systems and concepts, it is sometimes handy to consider objects that are composed of *overlapping* parts. A formal framework for the composition of such overlapping objects is given by the theory of inverse semigroups [19]. Recent applications of this theory include the modelling of music [8], interactive music systems [1, 14], and distributed systems [6]. As already observed in [16, 17] and developed further in [13], inverse semigroup theory conveys a notion of a higher dimensional string that seems particularly relevant in the context of such applications. This is especially clear in view of Stephen's representation theorem for inverse semigroups [26] that allows us to define graphical representations of elements of an inverse semigroup. These observations motivate the development of a formal language theory based on the theory of inverse semigroups.

However, the tools of classical formal language theory are not easily applicable for such a purpose. As already observed and studied in detail for languages in the free inverse monoid [25], that is, languages of *birooted trees,* the notion of algebraic recognisability by means of morphisms into finite monoids leads to a notion with rather weak expressive power. Further study of these languages in the context of tree walking automata has led to a strict hierarchy of classes of languages [12] ranging from *recognisable languages* (definable by means of finite monoids) to *logically definable languages* (definable in monadic second-order logic) and *regular languages* (definable with various notions of regular expressions).

It is known that the homomorphic image of an inverse monoid is an inverse monoid. When applied to inverse semigroups, the automata stemming from morphisms into finite monoids are reversible in a certain sense [21, 25]. Although leading to interesting studies of reversible computations (see [5] for instance), morphisms of inverse monoids preserve far too much structure to be used as a tool for language definability. The collapse of expressive power arises even in the absence of inverses themselves as illustrated by languages of *positive* birooted words [10] or languages of partially ordered graphs [4].

These observations lead us to the definition and the development of a more expressive notion of language definability, called *quasi-recognisability.* This notion is based on relaxing homomorphisms into *adequate premorphisms.* While a homomorphism of monoids preserves products, i.e., $\varphi(xy) = \varphi(x)\varphi(y)$, adequate premorphisms of ordered monoids are only required to be monotonic and submultiplicative, i.e., $\varphi(xy) \leq \varphi(x)\varphi(y)$, effectivity being ensured by additional preservation properties.

Applied first to languages of birooted words [10, 7], that is, subsets of the monoid of McAlister [22, 20], and then to languages of labelled birooted trees [9], this emerging notion of a quasi-recognisable language has been successfully related to classical automata theory. It has also been shown to 'essentially' capture definability by means of MSO-formulae [11, 9]. Hence, it provides new algebraic tools for the study of MSO-definable languages of finite trees [27].

At the moment the theory of quasi-recognisable languages is far from being fully understood. It turns out that the notion of an adequate premorphism is related to two fields of algebra: the theory of *partially ordered monoids* (adequate premorphisms are order preserving) and the theory of *partial algebras* (adequate premorphisms preserves a partial operation called the disjoint product).

In this paper we continue the study of languages of birooted trees and we make these connections with the theories of partial algebras and ordered monoids explicit. In short, we show that the theory of partial algebra provides effective tools to characterise MSO-definable languages of birooted trees. Additionally, the theory of partially ordered monoids provides, via the notion of quasi-recognisability, a more subtle description of (a large subclass of) these languages. This can be seen as a further example of the use of partially ordered monoids in algebraic language theory [23].

The paper is organised as follows. In the next section, we recall the basic notions from algebra we will need. We quickly review the definitions of a labelled birooted tree, the corresponding monoid, the operations of left and right projection, and the natural order on birooted trees. These notions are taken from the theory of inverse semigroups. We refer the reader to [9] for more details on the inverse monoid of labelled birooted trees. A thorough introduction to the theory of inverse semigroups can be found in [19].

The disjoint product of birooted trees – a notion that plays a key role in quasi-recognisability [9] – is reviewed in Section 3. This operation provides a link to the theory of partial algebras [2] which is exploited in this paper. In particular, we consider the notion of a closed $*$-congruence (cf. Definition 3.3). The closure property, which plays a central role in the theory of partial algebras, is characterised over birooted trees by means of their so-called *root type* (cf. Proposition 3.8).

In Section 4, we prove that every language of birooted trees admits a syntactic congruence, that is, a greatest closed $*$-congruence saturating the language (cf. Theorem 4.2). Then we prove that the syntactic congruence of a language has finite index if, and only if, the language is definable in monadic second-order logic (cf. Theorem 4.7). In this case we also obtain a linear time membership algorithm (cf. Theorem 4.8).

Quasi-recognisable languages, which capture finite boolean combinations of upward closed MSO-definable languages [9], are considered in Section 5. The notion of a bisimulation refinement of a closed $*$-congruence (cf. Definition 5.2) is defined and shown to exist. Quasi-recognisable languages are then characterised by means of the bisimulation refinement of their syntactic congruence (cf. Theorem 5.7).

Note: omitted proofs and further details are given in the appendices.

2 Birooted labelled trees

Throughout the paper we fix two finite alphabets $\mathbb{A} = \{a, b, c, \dots\}$ of *edge labels* and $\mathbb{F} = \{f, g, h, \dots\}$ of *vertex labels.* We assume that $\mathbb{A}$ is non-empty, while $\mathbb{F}$ is allowed to be empty. Let $\bar{\mathbb{A}} := \{\bar{a}, \bar{b}, \bar{c}, \dots\}$ be a disjoint copy of $\mathbb{A}$ and set $\widetilde{\mathbb{A}} = \mathbb{A} + \bar{\mathbb{A}}$. We define the *syntactic inverse mapping* $x \mapsto x^{-1}$ from $\widetilde{\mathbb{A}}$ to itself by $a^{-1} = \bar{a}$ and $\bar{a}^{-1} = a$, for $a \in \mathbb{A}$. This mapping is extended to $\widetilde{\mathbb{A}}^*$ by setting $1^{-1} := 1$ and $(au)^{-1} := u^{-1}a^{-1}$ for $a \in \widetilde{\mathbb{A}}$ and $u \in \widetilde{\mathbb{A}}^*$.

The *free group* $\mathrm{FG}(\mathbb{A})$ generated by $\mathbb{A}$ is the quotient of $\widetilde{\mathbb{A}}^*$ under the congruence generated by the equations $a\bar{a} = 1$ and $\bar{a}a = 1$ for $a \in \mathbb{A}$. Every equivalence class $[u] \in \mathrm{FG}(\mathbb{A})$ is uniquely determined by the unique word $\mathrm{red}(u)$ obtained from u by applying the rewriting rules $a\bar{a} \to 1$ and $\bar{a}a \to 1$ for $a \in \mathbb{A}$. In the sequel, we thus shall represent elements of $\mathrm{FG}(\mathbb{A})$ by their reduced form. In that case, the group product $u \cdot v$ of two elements $u, v \in \mathrm{FG}(\mathbb{A})$ is defined by $u \cdot v = \mathrm{red}(uv)$, and elements of $\mathrm{FG}(\mathbb{A})$ are partially ordered by the prefix order (defined over their reduced forms). Then, for every $u \in \mathrm{FG}(\mathbb{A})$, we have $u^{-1} \in \mathrm{FG}(\mathbb{A})$ and $u \cdot u^{-1} = 1 = u^{-1} \cdot u$, that is, the syntactic inverse u^{-1} of u coincides with its group inverse.

Definition 2.1 (Birooted labelled trees). A (vertex-labelled) birooted tree is a pair $x = \langle r, u \rangle$ where $r : \mathrm{FG}(\mathbb{A}) \to \mathbb{F} \cup \{\top\}$ is a partial mapping with a prefix-closed domain $\mathrm{dom}(r) \subseteq \mathrm{FG}(\mathbb{A})$ and $u \in \mathrm{dom}(r)$ is a distinguished vertex called the *output root.* The unit vertex $1 \in \mathrm{dom}(r)$ is called the *input root.*

In the case where the alphabet $\mathbb{F}$ contains at least two elements[1], the set of birooted trees is extended by a *zero element* 0. The set of all birooted trees is denoted by $\mathcal{B}(\mathbb{F}, \mathbb{A})$.

The *product* $x \cdot y$ of two non-zero birooted trees $x = \langle r, u \rangle$ and $y = \langle s, v \rangle$ is the birooted tree $\langle t, w \rangle$ where

$$\mathrm{dom}(t) = \mathrm{dom}(r) \cup u \cdot \mathrm{dom}(s)$$

and, for every $z \in \mathrm{dom}(t)$,

$$t(z) = \begin{cases} r(z) & \text{if } z \in \mathrm{dom}(r) - u \cdot \mathrm{dom}(s)\,, \\ s(u^{-1} \cdot z) & \text{if } z \in u \cdot \mathrm{dom}(s) - \mathrm{dom}(r)\,, \\ r(z) \wedge s(u^{-1} \cdot z) & \text{if } z \in \mathrm{dom}(r) \cap u \cdot \mathrm{dom}(s)\,. \end{cases}$$

The meet in the last clause of the definition is computed with respect to the trivial order on $\mathbb{F} \cup \{\top\}$ where $x \leq y$ iff $x = y$ or $y = \top$. The product is set to 0 if, for some z, the above meet does not exists, i.e., the labels of r and s at the respective places disagree. We extend the product to 0 by defining $x \cdot 0 = 0 = 0 \cdot x$ for all $x \in \mathcal{B}(\mathbb{F}, \mathbb{A})$. As usual, we may omit the dot and simply write xy instead of $x \cdot y$.

An example of labelled birooted trees and products is depicted Figure 1. The input/output roots are marked by dangling incoming/outgoing arrows. Directed edges are only labelled by letters of $\mathbb{A}$, letters of $\bar{\mathbb{A}}$ implicitly labelling the reverse of these edges.

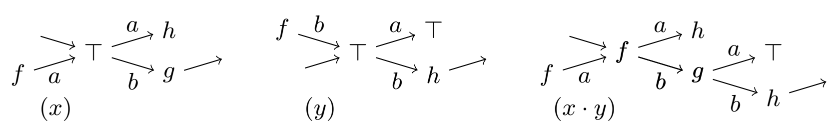

Fig. 1. Two compatible labelled birooted trees and their non-zero product.

As special cases of birooted trees there are *elementary* birooted trees that are either 0 or consist of a single vertex or a single edge. For all $f \in \mathbb{F}$ and $a \in \widetilde{\mathbb{A}}$, there are elementary birooted trees f and a that are defined as follows. The former have a single vertex labelled f, while the latter consist of a single edge labelled a that connects two vertices with label $\top$. The formal definitions are $f := \langle \{1 \mapsto f\}, 1 \rangle$ and $a := \langle \{1 \mapsto \top, a \mapsto \top\}, a \rangle$. The *neutral element* of $\mathcal{B}(\mathbb{F}, \mathbb{A})$ is the one vertex birooted tree with label $\top$ formally defined by $1 := \langle \{1 \mapsto \top\}, 1 \rangle$.

Most of the properties of birooted trees and their product follows from the following theorem. For the sake of completeness, let us recall that a monoid M is an *inverse monoid* when, for each $x \in M$, there exists a unique element $x^{-1} \in M$, called the *semigroup inverse* of x, such that $xx^{-1}x = x$ and $x^{-1}xx^{-1} = x^{-1}$.

[1] otherwise, there is no need of a zero

Theorem 2.2 ([9], [19]). *The set* $\mathcal{B}(\mathbb{F}, \mathbb{A})$ *equipped with the above product is an inverse monoid. It is the quotient of the free inverse monoid* $\mathrm{FIM}(\mathbb{F} + \mathbb{A})$ *by the identities* $ff = f$ *and* $fg = 0$*, for* $f, g \in \mathbb{F}$ *with* $f \neq g$*, and it is generated by the elementary birooted trees. For* $\mathbb{F} = \emptyset$*, the monoid* $\mathcal{B}(\emptyset, \mathbb{A})$ *is just the free inverse monoid* $\mathrm{FIM}(\mathbb{A})$ *itself.*

As both alphabets are fixed throughout the paper, we shall simply write $\mathcal{B}$ for the set of birooted trees over $\mathbb{F}$ and $\mathbb{A}$.

Below we review some basic notions and properties of the monoid $\mathcal{B}$ that follow from its definition and the fact that it is an inverse monoid. A detailed presentations of inverse semigroup theory can be found in [19].

First, note that $0^{-1} = 0$. The case of non-zero birooted trees is described in the following lemma.

Lemma 2.3 (Inverses and idempotent). *The inverse of a non-zero birooted tree* $x = \langle r, u\rangle$ *is* $x^{-1} = \langle r_u, u^{-1}\rangle$ *where*

$$\mathrm{dom}(r_u) = u^{-1} \cdot \mathrm{dom}(r) \quad \textit{and} \quad r_u(v) = r(u \cdot v)\,, \quad \textit{for } v \in \mathrm{dom}(r_u)\,.$$

Moreover, x *is idempotent if, and only if,* $u = 1$.

The next definition plays a fundamental role in our approach.

Definition 2.4 (Resets and co-resets). The *right projection,* or *reset,* of an element $x \in \mathcal{B}$ is the element $x^R := x \cdot x^{-1}$. Its *left projection,* or *co-reset,* is $x^L := x^{-1} \cdot x$.

We easily check that $x^R x = x = x x^L$, $x^R x^R = x^R$, and $x^L x^L = x^L$. Since idempotents are self inverse, this implies that the reset mapping $x \mapsto x^R$ and the co-reset mapping $x \mapsto x^L$ are both projections from the set of birooted trees into the set of idempotents.

Definition 2.5 (Natural order). The *natural order* $\leq$ on $\mathcal{B}$ is defined by

$$x \leq y \quad \text{:iff} \quad x = x^R \cdot y \quad (\text{equivalently } x = y \cdot x^L).$$

It can be shown that, as in any inverse monoid, a birooted tree $x \in \mathcal{B}$ is idempotent if, and only if, $x \leq 1$, that is, the idempotents are exactly the *subunits.* The set of idempotent elements of $\mathcal{B}$ is thus denoted by $U(\mathcal{B}) := \{x \in \mathcal{B} : x \leq 1\}$. In adequately ordered monoids, which we will introduce below, this is no longer the case, as there may be idempotents that are not subunits.

Every non-zero birooted tree $x = \langle s, u\rangle$ can be seen, as depicted in Figure 1, as a relational structure M_x with domain $\mathrm{dom}(s)$ over the signature $\mathbb{A}+\mathbb{F}+\{\mathrm{in}, \mathrm{out}\}$, where the symbols in $\mathbb{A}$ are interpreted as binary relations, those in $\mathbb{F}$ are interpreted as disjoint unary relations, and there are two distinguished vertices $\mathrm{in}, \mathrm{out} \in \mathrm{dom}(x)$ representing the input and output root. Encoding zero by the one vertex structure M_0 with non-empty interpretations of every symbol of $\mathbb{F}$ or $\mathbb{A}$, the natural ordered can then be characterised as follows.

Lemma 2.6. *Let x and y be two birooted trees. Then $x \leq y$ if and only if there is a morphism $\varphi : M_y \to M_x$ that preserves input and output root. For $x \neq 0$, this morphism is injective.*

Remark 2.7. Using such a representation of *all* birooted trees as structures, we can consider subsets of $\mathcal{B}$ that are *definable* in a given logic, say, in monadic second-order logic. The study of MSO-definable languages of birooted trees by means of algebraic methods is one of the main purpose of this paper.

3 The disjoint product

The *disjoint product,* introduced in [9], turns the algebra of birooted trees into a (finitely generated) *partial algebra* [2] over the signature consisting of the *partial* disjoint product, the *total* reset and co-reset operations, and (if necessary) the constant 0.

Definition 3.1 (Disjoint product). The partial *disjoint product* $x * y$ of two elements $x, y \in \mathcal{B}$ is equal to their usual product $x \cdot y$ if the following two conditions are satisfied:

(D1) $x \cdot y \neq 0$,
(D2) $x = \langle r, u \rangle$, $y = \langle s, v \rangle$, and $\mathrm{dom}(r) \cap u \cdot \mathrm{dom}(s) = \{u\}$.

Otherwise, the disjoint product is left undefined. We shall write $\exists x * y$ to denote both the existence of such a partial product and its value.

It is straightforward to check that the disjoint product is associative in the following sense: the disjoint product $x * (y * z)$ is defined if, and only if, $(x * y) * z$ is defined and, if this is the case, both products are equal.

Lemma 3.2 (Strong decomposition [9]). *Every birooted tree $x \in \mathcal{B}$ can be written as a combination of elementary birooted trees by disjoint products, reset and co-reset projections. This combination can be chosen to be of linear size in the number of vertices of x.*

The following definition is adapted from the theory of partial algebras (see [2] Sections 2.4–2.6) albeit with different terminology.

Definition 3.3. (a) A $*$-*congruence* over $\mathcal{B}$ is an equivalence relation $\simeq$ that is compatible with the disjoint product $*$ and the two projection operations ^{L} and ^{R} in the sense that

(P1) $x \simeq x'\,, \quad y \simeq y'\,, \quad \exists x * y\,, \quad \exists x' * y' \quad \text{implies} \quad x * y \simeq x' * y'\,,$

(P2) $x \simeq y \quad \text{implies} \quad x^L \simeq y^L \text{ and } x^R \simeq y^R\,,$

A $*$-congruence $\simeq$ is *closed* when it also satisfies the following property:

(P3) $x \simeq x' \text{ and } y \simeq y' \quad \text{implies} \quad \exists x * y \;\Leftrightarrow\; \exists x' * y'\,.$

(b) We call an equivalence relation on $\mathcal{B}$ *idempotent pure* if it does not identify an idempotent with a non-idempotent.

Proposition 3.4 (Lattice property [2]). *The set of $*$-congruences ordered by inclusion is a meet semi-lattice with intersection as meet. The set of closed $*$-congruences ordered by inclusion is a lattice with intersection as meet and the transitive closure of union as join.*

This proposition guarantees the existence of syntactic congruences, which will be defined and studied in the next section. We spend the rest of this section showing that there is a greatest idempotent-pure, closed $*$-congruence of finite index.

Definition 3.5 (Root types). The *root type* of a subunit $z \in U(\mathcal{B})$ is the set

$$\mathrm{rtp}(z) := \{f \in \mathbb{F} : zf = z\} \cup \{a \in \widetilde{\mathbb{A}} : za^R = z\}$$

where we omit the elements of $\mathbb{F}$ in the case that $|\mathbb{F}| \leq 1$.

Lemma 3.6. *For all $x, y \in \mathcal{B}$, we have*

$$\begin{aligned} \mathrm{rtp}(x^L) = \mathrm{rtp}(y^L) \quad &\textit{iff} \quad \exists x * z \Leftrightarrow \exists y * z\,, \quad \textit{for all } z \in \mathcal{B}\,, \\ \mathrm{rtp}(x^R) = \mathrm{rtp}(y^R) \quad &\textit{iff} \quad \exists z * x \Leftrightarrow \exists z * y\,, \quad \textit{for all } z \in \mathcal{B}\,. \end{aligned}$$

Definition 3.7 (Root equivalence). The *root equivalence* is defined by

$$x \approx_{\mathrm{rt}} y \quad \text{:iff} \quad \mathrm{rtp}(x^L) = \mathrm{rtp}(y^L) \text{ and } \mathrm{rtp}(x^R) = \mathrm{rtp}(y^R)\,.$$

The *strong root equivalence* $\approx_{\mathrm{srt}}$ is defined by

$$x \approx_{\mathrm{srt}} y \quad \text{:iff} \quad x \approx_{\mathrm{rt}} y \text{ and } x \in U(\mathcal{B}) \Leftrightarrow y \in U(\mathcal{B})\,.$$

Proposition 3.8. *The relation $\approx_{\mathrm{rt}}$ is the greatest closed equivalence. The relation $\approx_{\mathrm{srt}}$ is the greatest idempotent-pure closed $*$-congruence. Both equivalences have finite index.*

4 The syntactic congruence

In this section, we show that our framework induces a notion of a syntactic congruence that captures MSO-definability and that has a linear time membership algorithm.

Definition 4.1. A $*$-congruence $\simeq$ *saturates* a set $X \subseteq \mathcal{B}$ if

$$x \simeq y \quad \text{implies} \quad x \in X \Leftrightarrow y \in X\,, \quad \text{for all } x, y \in \mathcal{B}\,.$$

Theorem 4.2. *For every language $X \subseteq \mathcal{B}$ of birooted trees, there exists a greatest closed $*$-congruence $\simeq_X$ saturating X.*

Proof. By Proposition 3.4, the supremum of all closed $*$-congruences saturating X exists. This supremum still saturates X. □

Definition 4.3. We call the relation $\simeq_X$ of the previous lemma the *syntactic congruence* of X.

If the syntactic congruence has finite index, we obtain an effective membership algorithm in the same way as for languages of words. To present this algorithm we need to define quotients of partial algebras.

Definition 4.4 (Quotient algebra). Let $\simeq$ be a $*$-congruence over $\mathcal{B}$.

(a) The *quotient* $\mathcal{B}/\simeq$ is the partial algebra whose elements are the congruence classes of elements of $\mathcal{B}$ and whose operations are as follows: for $X, Y \in \mathcal{B}/\simeq$, we define

$$\begin{aligned} X^L &:= \{x^L \in \mathcal{B} : x \in X\}\,, \\ X^R &:= \{x^R \in \mathcal{B} : x \in X\}\,, \\ X * Y &:= \{z \in \mathcal{B} : z \simeq \exists x * y,\ x \in X,\ y \in Y\}\,, \end{aligned}$$

where $X * Y$ is only defined if the above set is nonempty.

(b) A *morphism of partial algebras* is a function $\varphi : A \to B$ between partial algebras A and B such that, for all $x, y \in A$,

$$\begin{aligned} \varphi(x^L) &= (\varphi(x))^L\,, \\ \varphi(x^R) &= (\varphi(x))^R\,, \\ \exists x * y \quad \text{implies} \quad \varphi(x * y) &= \exists \varphi(x) * \varphi(y)\,. \end{aligned}$$

(c) The canonical surjection $\theta_\simeq : \mathcal{B} \to \mathcal{B}/\simeq$ is the function mapping every element of $\mathcal{B}$ to its congruence class.

Lemma 4.5. *Let $\simeq$ be a $*$-congruence. The quotient algebra $\mathcal{B}/\simeq$ is well-defined and the canonical surjection $\theta_\simeq : \mathcal{B} \to \mathcal{B}/\simeq$ is a morphism of partial algebras.*

Proof. Essentially follows from standard results for partial algebras [2]. □

The relationship between the syntactic congruence and definability in monadic second-order logic [27] is as follows. Recall that M_x denotes the structure encoding a birooted tree x.

Lemma 4.6. *Let $\simeq$ be a $*$-congruence of finite index. For every class $X \in \mathcal{B}/\simeq$, there exists an* MSO*-formula φ_X such that*

$$M_x \models \varphi_X \quad \textit{iff} \quad x \in X\,, \quad \textit{for all } x \in \mathcal{B}\,.$$

Proof (Sketch). The claim follows from Lemma 3.2 and the fact that $\mathcal{B}/\simeq$ is finite: for every birooted tree $x \in \mathcal{B}$, the value $\theta_\simeq(x)$ can be defined in MSO by guessing a strong decomposition of x and a corresponding labelling in $\mathcal{B}/\simeq$, in almost the same way as the behaviour of a bottom up tree automaton over a finite tree can be described in MSO. □

Theorem 4.7. *A language $X \subseteq \mathcal{B}$ of birooted trees is definable in* MSO *if, and only if, its syntactic congruence $\simeq_X$ has finite index.*

Proof (Sketch). ($\Leftarrow$) As $\simeq_X$ saturates X, we can write $X = \bigcup_{x \in X} [x]_{\simeq_X}$. When the syntactic congruence has finite index, this union is finite. Consequently, the formula $\psi_X = \bigvee_{x \in X} \varphi_{[x]_{\simeq}}$ defines X, where $\varphi_{[x]_{\simeq}}$ are the formulae from Lemma 4.6.

($\Rightarrow$) If X is definable in MSO, we can use decomposition arguments (see [24] and [28]) to prove that $\simeq_X$ has finite index. Two trees with the same MSO-theory (up to a given quantifier rank) cannot be distinguished by X; hence they must be $\simeq_X$-equivalent. Since there are only finitely many such theories, it follows that $\simeq_X$ has only finitely many classes. □

Another consequence of Lemma 4.5 worth being mentioned concerns a membership decision algorithm.

Theorem 4.8. *Let $X \subseteq \mathcal{B}$ be a language whose syntactic congruence $\simeq_X$ has finite index. Given $\mathcal{B}/{\simeq_X}$ one can decide whether $x \in X$ in time linear in the size of the input $x \in \mathcal{B}$.*

Proof. By Lemma 3.2, the set of birooted trees is finitely generated as a partial algebra. It follows that, starting from the elementary birooted trees we can inductively compute the image $\theta_X(x)$ of a birooted tree $x \in \mathcal{B}$ under the canonical surjection $\theta_X : \mathcal{B} \to \mathcal{B}/{\simeq_X}$. The number of steps is linear in the size of x. This proves the claim under the usual assumption that every projection, disjoint product, or equality test in the partial algebra $\mathcal{B}/{\simeq_X}$ takes constant time. □

5 Application to quasi-recognisability

We aim now at a characterisation of quasi-recognisable languages via their syntactic congruence. We start by recalling the definition of quasi-recognisability.

Definition 5.1 (Quasi-recognisable languages [9]). (a) A function $\varphi : \mathcal{B} \to M$ from $\mathcal{B}$ into an ordered monoid M is an *adequate premorphism* if it satisfies the following conditions:

(M1) $x \leq y$ implies $\varphi(x) \leq \varphi(y)$,
(M2) $\varphi(xy) \leq \varphi(x)\varphi(y)$,
(M3) $\varphi(x^L) = (\varphi(x))^L$ and $\varphi(x^R) = (\varphi(x))^R$,
(M4) $\exists x * y$ implies $\varphi(x * y) = \varphi(x)\varphi(y)$.

(b) An ordered monoid M is *adequately ordered* if all subunits of M are idempotent and both $x^R = \min\{z \leq 1 : zx = x\}$ and $x^L = \min\{z \leq 1 : x = xz\}$ exist for every $x \in \mathcal{B}$.

(c) A language $X \subseteq \mathcal{B}$ is *quasi-recognisable* if there exists an adequate function $\varphi : \mathcal{B} \to M$ into a finite adequately ordered monoid M such that $X = \varphi^{-1}(\varphi(X))$.

It is proved in [9] that quasi-recognisable languages corresponds to finite Boolean combinations of upward closed (in the natural order) MSO-definable languages. We now aim at characterising quasi-recognisable languages by means of their syntactic congruence. We need two new notions for this characterisation: that a of $*$-bisimulation and that of an alternating chain.

Definition 5.2 ($*$-bisimulation). A closed $*$-congruence $\simeq$ is a $*$-*bisimulation* if it is idempotent pure and satisfies the property:

(P4) if $x \leq y \simeq z$ then there exists $x' \simeq x$ such that $x' \leq z$.

It is a *strong $*$-bisimulation* when, additionally, each $\simeq$-class is convex, i.e.,

(P5) if $x \leq y \leq z$ and $x \simeq z$ then $x \simeq y \simeq z$.

Lemma 5.3 (Lattice property). *The set of $*$-bisimulations ordered by inclusion forms a complete lattice with transitive closure of the union as join.*

Corollary 5.4. *Every closed $*$-congruence $\simeq$ contains a greatest $*$-bisimulation $\simeq^B$.*

Proof. The relation $\simeq_{\mathrm{m}} := \simeq \cap \approx_{\mathrm{srt}}$ is the greatest idempotent-pure closed $*$-congruence included in $\simeq$. We can define $\simeq^B$ as the transitive closure of the union of all $*$-bisimulations included in $\simeq_{\mathrm{m}}$. By Lemma 5.3, this is a $*$-bisimulation. □

Definition 5.5. We call the $*$-bisimulation $\simeq^B$ the *bisimulation refinement* of $\simeq$.

The second new notion for our characterisation is that of an alternating chain.

Definition 5.6. Let $\simeq$ be an equivalence relation on an ordered set M. An *alternating $\simeq$-chain* is an increasing sequence $x_0 \leq \cdots \leq x_n$ such that $x_i \not\simeq x_{i+1}$, for all $i < n$. The number n is called the *length* of the chain.

Theorem 5.7. *Let $X \subseteq \mathcal{B}$. The following properties are equivalent:*

(1) *The language X is quasi-recognisable.*
(2) *The bisimulation refinement $\simeq_X^B$ of the syntactic congruence $\simeq_X$ is a strong $*$-bisimulation of finite index.*
(3) *The syntactic congruence $\simeq_X$ has finite index and the length of alternating $\simeq_X$-chains is bounded.*

Proof (sketch). (1) ⇒ (3) Let $\simeq$ be the kernel of an adequate premorphism recognising X. The relation $\simeq \cap \approx_{\mathrm{srt}}$ is a $*$-congruence of finite index with bounded length of alternating chains. By Lemma 4.2, it is included in $\simeq_X$. Consequently, the same holds for $\simeq_X$.

(3) ⇒ (2) The fact that alternating $\simeq_X$-chains are bounded ensures that $\simeq_X^B$ is a strong $*$-bisimulation. Moreover, it has finite index since $\simeq_X$ does.

(2) ⇒ (1) For $X, Y \in \mathcal{B}/{\simeq_B^X}$, we define

$$X \preceq Y \quad \text{:iff} \quad \text{there are } x \in X \text{ and } y \in Y \text{ with } x \leq y$$

This relation is reflexive, transitive (by (P4)), and anti-symmetric (by (P5)). We can embed the resulting finite partial order $\langle \mathcal{B}/{\simeq_B^X}, \preceq \rangle$ into a finite adequately ordered monoid M via an adequate premorphism $\varphi : \mathcal{B} \to M$ whose kernel equals $\simeq_X$. □

6 Conclusion

We have characterised both recognisable and quasi-recognisable languages in terms of their syntactic congruences. It can be shown that the notion of strong $*$-bisimulation induces a refinement of quasi-recognisability: strong recognisability, which admits minimal recognisers.

As far as the membership problem is concerned, syntactic congruences and the corresponding quotients can be used, whether or not the considered language is quasi-recognisable.

Note that the bisimulation refinement may induce a non-elementary blowup in the index of the congruence. In return we hope that this refinement encodes more subtle properties of a language. We note however that we provide no effective algorithm to compute the bisimulation refinement of a closed $*$-congruence of finite index.

Although out of the scope of our presentation, it can also be shown that the class of MSO-definable languages is quite *robust* in the sense that it is closed under (non-zero) products, iterated products, inverses, upward and downward closures, reset and co-reset projections. How do these operations act on the syntactic algebras? An answer might lead to a better understanding of some subclasses of MSO-definable languages.

Last but not least, the monoid of labelled birooted trees studied here is a special case of so-called 0-E-unitary inverse monoids [19]. These monoids enjoy a similar graphical representation [26] with, especially, a natural order between non-zero elements characterised by means of injective morphisms. We suspect that the algebraic tools developed here could partially be adapted to this more general setting, perhaps in connection with languages of graphs as studied in [3].

References

1. F. Berthaut, D. Janin, and B. Martin. Advanced synchronization of audio or symbolic musical patterns: an algebraic approach. *International Journal of Semantic Computing*, 6(4):409–427, 2012.
2. P. Burmeister. *A Model Theoretic Oriented Approach to Partial Algebras.* Akademie-Verlag, 1986.
3. B. Courcelle and J. Engelfriet. *Graph structure and monadic second-order logic, a language theoretic approach*, volume 138 of *Encyclopedia of mathematics and its applications.* Cambridge University Press, 2012.
4. B. Courcelle and P. Weil. The recognizability of sets of graphs is a robust property. *Theoretical Comp. Science*, 342:173–228, 2005.
5. V. Danos and L. Regnier. Reversible, irreversible and optimal lambda-machines. *Theoretical Comp. Science*, 227(1-2):79–97, 1999.
6. A. Dicky and D. Janin. Modélisation algébrique du diner des philosophes. In *Modélisation des Systèmes Réactifs (MSR), in Journal Européen des Systèmes Automatisés (JESA Volume 47 - no 1-2-3/2013)*, Rennes, France, 2013.
7. D. Janin. Quasi-recognizable vs MSO definable languages of one-dimensional overlapping tiles. In *Mathematical Found. of Comp. Science (MFCS)*, volume 7464 of *LNCS*, pages 516–528, Bratislava, Slovakia, 2012.

8. D. Janin. Vers une modélisation combinatoire des structures rythmiques simples de la musique. *Revue Francophone d'Informatique Musicale (RFIM)*, 2, 2012.
9. D. Janin. Algebras, automata and logic for languages of labeled birooted trees. In *Int. Col. on Aut., Lang. and Programming (ICALP)*, volume 7966 of *LNCS*, pages 318–329, Riga, Latvia, 2013. Springer.
10. D. Janin. On languages of one-dimensional overlapping tiles. In *Int. Conf. on Current Thrends in Theo. and Prac. of Comp. Science (SOFSEM)*, volume 7741 of *LNCS*, pages 244–256, Spindlerûv Mlýn, Czech Republic, 2013. Springer.
11. D. Janin. Overlaping tile automata. In *8th International Computer Science Symposium in Russia (CSR)*, volume 7913 of *LNCS*, pages 431–443, Ekaterinburg, Russia, 2013. Springer.
12. D. Janin. Walking automata in the free inverse monoid. Research report RR-1464-12, LaBRI, Université de Bordeaux, 2013. (revised May 2013).
13. D. Janin. *Towards a higher dimensional string theory for the modeling of computerized systems*, volume 8327 of *LNCS*, pages 7–20. Springer, Novy Smokovec, Slovaquia, 2014.
14. D. Janin, F. Berthaut, and M. DeSainteCatherine. Multi-scale design of interactive music systems : the libTuiles experiment. In *10th Conference on Sound and Music Computing (SMC)*, Stockholm, Sweden, 2013.
15. D. Jongh and A. Troelstra. On the connection of partially ordered sets with some pseudo-boolean algebras. *Indagationes Mathematica*, 28:317–329, 1966.
16. J. Kellendonk. The local structure of tilings and their integer group of coinvariants. *Comm. Math. Phys.*, 187:115–157, 1997.
17. J. Kellendonk and M. V. Lawson. Tiling semigroups. *Journal of Algebra*, 224(1):140 – 150, 2000.
18. P. Körtesi, S. Radeleczki, and S. Szilágyi. Congruences and isotone maps on partially ordered sets. *Mathematica Pannonica*, 16(1):39–55, 2005.
19. M. V. Lawson. *Inverse Semigroups : The theory of partial symmetries*. World Scientific, 1998.
20. M. V. Lawson. McAlister semigroups. *Journal of Algebra*, 202(1):276 – 294, 1998.
21. S. W. Margolis and J.-E. Pin. Languages and inverse semigroups. In *Int. Col. on Aut., Lang. and Programming (ICALP)*, volume 172 of *LNCS*, pages 337–346. Springer, 1984.
22. D.B. McAlister. Inverse semigroups which are separated over a subsemigroups. *Trans. Amer. Math. Soc.*, 182:85–117, 1973.
23. J.-E. Pin. Algebraic tools for the concatenation product. *Theoretical Comp. Science*, 292(1):317–342, 2003.
24. S. Shelah. The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.
25. P. V. Silva. On free inverse monoid languages. *ITA*, 30(4):349–378, 1996.
26. J.B. Stephen. Presentations of inverse monoids. *Journal of Pure and Applied Algebra*, 63:81–112, 1990.
27. W. Thomas. Chap. 7. Languages, automata, and logic. In *Handbook of Formal Languages, Vol. III*, pages 389–455. Springer-Verlag, Berlin Heidelberg, 1997.
28. W. Thomas. Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In *Structures in Logic and Computer Science*, volume 1261 of *LNCS*, pages 118–143. Springer, 1997.

A Omitted proofs of Section 3

A.1 Proof of Proposition 3.4

Proof. This essentially follows from standard results for partial algebras [2]. We present the proof for the sake of completeness.

Let R_1 and R_2 be two $*$-congruences. Clearly, the intersection $R_1 \cap R_2$ is also a $*$-congruence. Furthermore, if both R_1 and R_2 are closed, then so is $R_1 \cap R_2$.

Let $R_1 \sqcup R_2 := (R_1 \cup R_2)^+$ be the transitive closure of the union $R_1 \cup R_2$. Since both R_1 and R_2 are reflexive, we have

$$R_1 \sqcup R_2 = (R_1 \cup R_2)^* = \bigcup \{(R_1 \cup R_2)^n : 0 \leq n < \omega\}\,.$$

It suffices to check that the relation $R_1 \sqcup R_2$ is a closed $*$-congruence. By definition it then follows that it is the least closed $*$-congruence containing both R_1 and R_2.

Let $(x, x'), (y, y') \in R_1 \sqcup R_2$. Since $R_1 \cup R_2$ is reflexive, we can find a number n such that $(x, x'), (y, y') \in (R_1 \cup R_2)^n$. We check Properties (P1)–(P3) by induction on n.

We start with (P1) and (P3). Assume that $\exists x * y$. If $n = 0$, then $x = x'$ and $y = y'$. Consequently, $\exists x' * y'$ and $(x * y, x' * y') \in R_1 \sqcup R_2$. Hence, suppose that $n = m+1$, for some m. Then there exists $x'' \in \mathcal{B}$ such that $(x, x'') \in (R_1 \cup R_2)^m$ and $(x'', x') \in R_i$ for some $i \in \{1, 2\}$. By inductive hypothesis, $\exists x * y$ implies $\exists x'' * y'$ and $(x * y, x'' * y') \in R_1 \sqcup R_2$. But since R_i satisfies (P3) with $(x'', x') \in R_i$ we also have $\exists x' * y'$. Hence, (P1) implies that $(x'' * y', x' * y') \in R_i \subseteq R_1 \sqcup R_2$. It follows, by transitivity of the relation $R_1 \sqcup R_2$, that $(x * y, x' * y') \in R_1 \sqcup R_2$.

The case of Property (P2) is treated similarly. We want to prove that we both have $(x^L, {x'}^L) \in R_1 \sqcup R_2$ and $(x^R, {x'}^R) \in R_1 \sqcup R_2$. If $n = 0$, then $x' = x$ and we are done. Hence, suppose that $n = m + 1$. Then there exists $z \in \mathcal{B}$ such that $(x, z) \in (R_1 \cup R_2)^m$ and $(z, x') \in R_i$ for some $i \in \{1, 2\}$. By inductive hypothesis, this implies that $(x^L, z^L) \in R_1 \sqcup R_2$ and $(x^R, z^R) \in R_1 \sqcup R_2$. Since R_i satisfies (P2) it also follows that $(z^L, {x'}^L) \in R_i \subseteq R_1 \sqcup R_2$ and $(z^R, {x'}^R) \in R_i \subseteq R_1 \sqcup R_2$. We conclude by transitivity. ◻

A.2 Proof of Lemma 3.6

Proof. ($\Leftarrow$) Assume that, for every $z \in \mathcal{B}$, we have $\exists z * x$ if, and only if, $\exists z * y$.

In the case where $x = 0$, the disjoint product $x * 1$ is undefined. Hence, $y * 1$ is also undefined. This implies that $y = 0$ and $\mathrm{rtp}(x^R) = \mathrm{rtp}(y^R)$.

In the remaining cases we may assume, by symmetry, that neither x nor y equals 0. Hence, $x = \langle r, u \rangle$ and $y = \langle s, v \rangle$. From the fact that $\exists a * x \Leftrightarrow a \notin \mathrm{dom}(s) \cap \widetilde{\mathbb{A}}$ for all $a \in \widetilde{\mathbb{A}}$, we immediately deduce that $\mathrm{rtp}(x^R) \cap \widetilde{\mathbb{A}} = \mathrm{rtp}(y^R) \cap \widetilde{\mathbb{A}}$.

In the case where $|\mathbb{F}| \geq 2$, we also have $\exists f * x \Leftrightarrow f \leq r(1)$ for all $f \in \mathbb{F}$. Hence, it follows that $\mathrm{rtp}(x^R) \cap \widetilde{\mathbb{F}} = \mathrm{rtp}(y^R) \cap \widetilde{\mathbb{F}}$. Observe that, in the case where $|\mathbb{F}| \leq 1$, the vertex labelling plays no role in the definition of the disjoint product.

(⇒) Suppose that $\mathrm{rtp}(x^R) = \mathrm{rtp}(y^R)$. A similar case study shows that for all $z \in \mathcal{B}$ we indeed have $\exists x * z$ if, and only if, $\exists y * z$.

Symmetrical arguments show that the statement $\mathrm{rtp}(x^R) = \mathrm{rtp}(y^R)$ is equivalent to the fact that, for every $z \in \mathcal{B}$, we have $\exists z * x$ if and only if $\exists z * y$. □

A.3 Proof of Proposition 3.8

Proof. Clearly, $\approx_{\mathrm{rt}}$ is an equivalence and it satisfies (P3), by Lemma 3.6. Hence, it is closed. It follows by Lemma 3.6 that it is the greatest such relation.

It is easy to see that $\approx_{\mathrm{srt}}$ is the greatest idempotent-pure closed equivalence. It therefore remains to show that it is a $*$-congruence.

Property (P2) easily follows from the observation that, for all idempotents $x, y \in \mathcal{B}$, we have $x \approx_{\mathrm{srt}} y$ if, and only if, $\mathrm{rtp}(x) = \mathrm{rtp}(y)$. Indeed, suppose that $x \approx_{\mathrm{srt}} y$. Since $\approx_{\mathrm{srt}} \subseteq \approx_{\mathrm{rt}}$, this means in particular that $\mathrm{rtp}(x^R) = \mathrm{rtp}(y^R)$. Hence $x^R \approx_{\mathrm{srt}} y^R$ as both x^R and y^R are idempotent and idempotents are invariant under projections.

Property (P1) can be proved by a case distinction. Suppose that $x * y$ is defined.

- If both x and y are idempotent then
$$\mathrm{rtp}((x * y)^L) = \mathrm{rtp}((x * y)^R) = \mathrm{rtp}(x) \cup \mathrm{rtp}(y)\,.$$
- If x is idempotent and y is not idempotent then
$$\mathrm{rtp}((x * y)^L = \mathrm{rtp}(y^L) \quad \text{and} \quad \mathrm{rtp}((x * y)^R) = \mathrm{rtp}(x) \cup \mathrm{rtp}(y^R)\,.$$
- Similarly, if x is not idempotent and y is idempotent then
$$\mathrm{rtp}((x * y)^L) = \mathrm{rtp}(x^L) \cup \mathrm{rtp}(y) \quad \text{and} \quad \mathrm{rtp}((x * y)^R) = \mathrm{rtp}(x^R)\,.$$
- If neither x nor y is idempotent then
$$\mathrm{rtp}((x * y)^L) = \mathrm{rtp}(y^L) \quad \text{and} \quad \mathrm{rtp}((x * y)^R) = \mathrm{rtp}(x^R)\,.$$

To see that (P1) holds, consider elements $x \approx_{\mathrm{srt}} x'$ and $y \approx_{\mathrm{srt}} y'$ such that $x * y$ and $x' * y'$ are defined. By definition of $\approx_{\mathrm{rt}}$, we have $\mathrm{rtp}(x^L) = \mathrm{rtp}(x'^L)$, $\mathrm{rtp}(x^R) = \mathrm{rtp}(x'^R)$, $\mathrm{rtp}(y^L) = \mathrm{rtp}(y'^L)$ and $\mathrm{rtp}(y^R) = \mathrm{rtp}(y'^R)$. As $\approx_{\mathrm{srt}}$ is idempotent pure, it follows in all of the above cases that $\mathrm{rtp}((x*y)^L) = \mathrm{rtp}((x' * y')^L)$, $\mathrm{rtp}((x * y)^R) = \mathrm{rtp}((x' * y')^R)$, and $x * y \in U(\mathcal{B}) \Leftrightarrow x' * y' \in U(\mathcal{B})$. In other words, $x * y \approx_{\mathrm{srt}} x' * y'$. □

Remark A.1. (a) Note that, as soon as $\mathbb{A}$ is not a singleton, the relation $\approx_{\mathrm{rt}}$ is not a $*$-congruence. Although it satisfies axiom (P2), it does not satisfies axiom (P1). Indeed, given $x = a^L a^R$ and $y = a^L a a^R$, we have $x \approx_{\mathrm{rt}} y$ since $\mathrm{rtp}(x^L)$, $\mathrm{rtp}(x^R)$, $\mathrm{rtp}(y^L)$ and $\mathrm{rtp}(y^R)$ are all equal to $\{a, \bar{a}\}$. However, we have $\mathrm{rtp}((x * b)^R) = \{a, \bar{a}, b\}$ while $\mathrm{rtp}((y * b)^R) = \{a, \bar{a}\}$.

(b) Note that a $*$-congruence needs not to be idempotent pure. An example is the relation R defined from the relation $\approx_{\mathrm{srt}}$ by adding all pairs of birooted trees $x, y \in \mathcal{B}$ such that $\mathrm{rtp}(x^L)$, $\mathrm{rtp}(x^R)$, $\mathrm{rtp}(y^L)$ and $\mathrm{rtp}(y^R)$ contains $\widetilde{\mathbb{A}}$. Clearly, the relation R satisfies axiom (P2). Moreover, we do have $\approx_{\mathrm{srt}} \subset R \subset \approx_{\mathrm{rt}}$. But no trees in these newly added pairs can be used in a disjoint product. The relation R thus satisfies Properties (P1) and (P3) just for the same reasons the relation $\approx_{\mathrm{srt}}$ satisfies them.

B More details on Section 4

B.1 Computing the syntactic congruence by refinement

We provide here a strengthening of Lemma 4.2 that states the existence of the syntactic congruence in an arbitrary partial algebra over the signature of $\mathcal{B}$. Furthermore, it shows that the syntactic congruence can be computed as the greatest fixed point of a certain mapping. For finite partial algebras, this provides an effective algorithm to compute the syntactic congruence of X from any given closed $*$-congruence of finite index saturating X.

Lemma B.1. *Let $\mathcal{C}$ be a partial algebra with disjoint product and left and right projection. Assume that there is a greatest closed equivalence $\approx$ over $\mathcal{C}$. For every closed relation $R \subseteq \approx$, let $F(R)$ be the relation consisting of all pairs $(x, y) \in R$ such that, for every $z_1, z_2, z_3 \in \mathcal{C}$, the following properties are satisfied:*

(a) $(x^L, y^L) \in R$ *and* $(x^R, y^R) \in R$,
(b) $\exists x * z \in R \Leftrightarrow \exists y * z \in R$ *and* $\exists z * x \in R \Leftrightarrow \exists z * y \in R$, *for every* $z \in \mathcal{C}$.

The function F is monotonic with respect to the inclusion order. Furthermore, for every $X \subseteq \mathcal{C}$, there is the greatest closed $*$-equivalence $\simeq_X$ saturating X given by the transfinite equation:

$$\simeq_X = \bigcap_{\alpha} F^{\alpha}(\approx_X)$$

where

$$x \approx_X y \quad \text{:iff} \quad x \approx y \text{ and } x \in X \Leftrightarrow y \in X\,,$$

and $F^0(R) = \mathrm{id}$, $F^{\alpha+1}(R) = F(F^{\alpha}(R))$ and, $F^{\delta}(R) = \bigcap_{\alpha<\delta} F^{\alpha}(R)$, for limit ordinals δ.

Proof. We first check that, if R is an equivalence relation then so is $F(R)$. If R is closed, monotonicity of F then implies that $F(R)$ is also closed. The fixed-point theorem of Tarski thus ensures that F has a greatest fixed point $\sim$ which is given by the above formula.

By monotonicity, we have $\sim \subseteq \approx_X$. Hence, the relation $\sim$ saturates X. It remains to show that $\sim$ is a closed $*$-congruence. Indeed, given any closed $*$-congruence $\sim'$ saturating X, we can prove by induction on α that $\sim' \subseteq F^{\alpha}(\approx)$. Hence $\sim' \subseteq \sim$.

Since it is included in $\approx$, $\sim$ is a closed equivalence relation. It thus remains to show that $\sim$ is a $*$-congruence. It is straightforward to that Property (a) implies (P2) and Property (b) implies (P1). $\square$

Applied to the partial algebra $\mathcal{B}$ of birooted trees, this provides an alternative proof of Theorem 4.2. It also provides an effective way to compute the syntactic congruence of a recognisable language.

Corollary B.2. *Let $\simeq$ be a closed $*$-congruence of saturating a language $X \subseteq \mathcal{B}$. Let $\mathcal{B}/\simeq$ be the quotient and let $\varphi : \mathcal{B} \to \mathcal{B}/\simeq$ be the corresponding projection. Then we can compute the syntactic congruence $\simeq_X$ in time $O(n^3)$ where n is the index of $\simeq$.*

Proof. Let $\approx$ be the relation over $\mathcal{B}/\simeq$ defined by

$$Y \approx Z \quad :\text{iff} \quad y \approx_{\text{rt}} z \quad \text{for some}^2\ y \in Y \text{ and } z \in Z\,.$$

This relation is obviously the greatest closed equivalence over $\mathcal{B}/\simeq$. Therefore, we can use Lemma B.1 to obtain the greatest closed $*$-congruence $\sim_{\varphi(X)}$ over $\mathcal{B}/\simeq$ that saturates $\varphi(X)$. For every $x, y \in \mathcal{B}$, it follows that $x \simeq_X y$ if and only $\varphi(x) \sim_{\varphi(X)} \varphi(y)$. Since every equivalence relation over $\mathcal{B}/\simeq$ has index at most n, this relation can be computed in at most $O(n)$ iteration steps (the depth of the lattice of equivalence relation ordered by inclusion) and each step is taking time at most $O(n^2)$ (the size of the current relation times the number of possible applications of the separation rules (a) and (b)). $\square$

C Omitted proofs of Section 5

We review here some properties of the bisimulation refinement of a $*$-congruence that will eventually lead us to a proof of Theorem 5.7.

C.1 Proof of Lemma 5.3

Proof. Let R_1 and R_2 be two $*$-bisimulations over the set of birooted trees. By Proposition 3.4, we already know that the relation $R_1 \sqcup R_2 = (R_1 \cup R_2)^*$ is a closed $*$-congruence. Since R_1 and R_2 are idempotent pure, so is $R_1 \sqcup R_2$. Let us prove that the relation $R_1 \sqcup R_2$ is a $*$-bisimulation.

Consider elements $x, y, y' \in \mathcal{B}$ with $x \leq y$ and $(y, y') \in R_1 \sqcup R_2$. Then there is a number $n < \omega$ such that $(y, y') \in (R_1 \cup R_2)^n$. If $n = 0$, then $y' = y$ and we can take $x' = x$. Otherwise, there exists $m < \omega$ and $y'' \in \mathcal{B}$ such that $n = m + 1$ with $(y, y'') \in (R_1 \cup R_2)^m$ and $(y'', y') \in R_i$ for some $i \in \{1, 2\}$. By inductive hypothesis, there exists $x'' \in \mathcal{B}$ such that $(x, x'') \in R_1 \sqcup R_2$ and $x'' \leq y''$. As R_i satisfies Property (P4), we can find an element $x' \in \mathcal{B}$ such that $(x'', x') \in R_i \subseteq R_1 \sqcup R_2$ and $x' \leq y'$. By transitivity, it follows that $(x, x') \in R_1 \sqcup R_2$.

We have shown that the set of $*$-bisimulations forms a semi-lattice with $\sqcup$ as join. Since our proof clearly extends to an arbitrary number of $*$-bisimulations, this shows that the underlying semi-lattice is actually complete.

Since $\approx_{\mathrm{srt}}$ is the maximum idempotent-pure closed $*$-congruence and also a $*$-bisimulation, there is a greatest $*$-bisimulation. Consequently, the $*$-bisimulations ordered by inclusion form a complete lattice. □

Remark C.1. (a) Note that the meet of two $*$-bisimulations is not necessarily their intersection.

(b) In general, the lattice property does not seems to hold for strong $*$-bisimulations.

C.2 Properties of the bisimulation refinement

We give here a fixed point characterisation of the $*$-bisimulation refinement of a closed $*$-congruence. Before giving that characterisation, let us first note that the natural order is well-behaved in some sense with respect to projections and the disjoint product.

Lemma C.2. *Let $x, y, z \in \mathcal{B}$.*

(a) *For every $z \leq x^R$, there exists $x' \leq x$ such that $z = x'^R$.*
(b) *For every $z \leq x^L$, there exists $x' \leq x$ such that $z = x'^L$.*
(c) *For every $z \leq x * y$, there exist $x' \leq x$ and $y' \leq y'$ such that $z = x' * y'$.*

Proof. This easily follows from the embedding characterisation of the natural order stated in Lemma 2.6. □

We consider the following function F on binary relations:

$$F(R) := \big\{(x,y) \in R : \text{for every } x' \leq x \text{ there exists } y' \leq y \text{ with } (x',y') \in R\,, \\ \text{for every } y' \leq y \text{ there exists } x' \leq x \text{ with } (x',y') \in R\big\}\,.$$

Lemma C.3. *If R is an (idempotent-pure) closed $*$-congruence, then so is $F(R)$.*

Proof. Let R be a closed $*$-congruence. Clearly, $F(R)$ is a closed equivalence. To prove that is a $*$-congruence we have to check two properties.

For (P1), let $(x_1, y_1), (x_2, y_2) \in R$ and suppose that both $x_1 * x_2$ and $y_1 * y_2$ are defined. We have to show that $(x_1 * x_2, y_1 * y_2) \in F(R)$. Clearly, $(x_1 * x_2, y_1 * y_2) \in R$ as R satisfies (P1). Let $z \leq x_1 * x_2$. By Lemma C.2, there exists $x'_1, x'_2 \in \mathcal{B}$ such that $z = x'_1 * x'_2$, $x'_1 \leq x_1$, and $x'_2 \leq x_2$. But since $(x_1, y_1) \in F(R)$, by definition, there exists $y'_1 \in \mathcal{B}$ such that $(x'_1, y'_1) \in R$ and $y'_1 \leq y_1$. Similarly, there exists $y'_2 \in \mathcal{B}$ such that $(x'_2, y'_2) \in R$ and $y'_2 \leq y_2$. Since R satisfies (P1) and it is closed, we have $(x'_1 * x'_2, y'_1 * y'_2) \in R$. By monotonicity of the order, we have therefore found $z' = y'_1 * y'_2 \leq y_1 * y_2$ such that $(z, z') \in R$. By a symmetrical argument, if we assume that $z \leq y_1 * y_2$ then there exists $z' \leq x_1 * x_2$ such that $(z', z) \in R$. In other words, $(x_1 * x_2, y_1 * y_2) \in F(R)$. Hence, Property (P1) is satisfied.

For (P2), let $(x, y) \in F(R)$. We have to show that $(x^R, y^R) \in F(R)$. Clearly, $(x^R, y^R) \in R$ since R satisfies (P2). Let $z \leq x^R$. By Lemma C.2, there exists $x' \in \mathcal{B}$ such that $x' \leq x$ and $z = x'^R$. By definition of $F(R)$, there exists

$y' \in \mathcal{B}$ such that $y' \leq y$ with $(x', y') \in R$. But since R is a $*$-congruence, by (P2), it follows that $(x'^R, y'^R) \in R$. By monotonicity of the right projection, there therefore exists $z' = y'^R \leq y^R$ such that $(z, z') \in R$. By applying a symmetrical argument, if we assume that $z \leq y^R$ then there exists $z' \leq x^R$ such that $(z', z) \in R$. In other words, $(x^R, y^R) \in F(R)$. A similar argument shows that $(x^L, y^R) \in F(R)$. Hence, Property (P2) is satisfied.

Obviously, if R is idempotent pure then so is $F(R)$.

Lemma C.4 (Fixed point characterisation). *The bisimulation refinement $\simeq^B$ of a closed $*$-congruence $\simeq$ can be computed as the greatest fixed point*

$$\simeq^B = \bigcap_{\alpha} F^{\alpha}(\sim)$$

of F that is included in the relation $\sim := \simeq \cap \approx_{\mathrm{srt}}$.

Proof. It is straightforward to check that, if R is a $*$-bisimulation, then $R = F(R)$. Note that $\simeq^B$ is the greatest $*$-bisimulation included in $\simeq$ and, therefore, the greatest $*$-bisimulation included in $\sim$. As F is monotonic, the claim therefore follows by the fixed point theorem of Knaster–Tarski. □

The interest in (strong) $*$-bisimulations is that they provide a (partial order) preorder relation on the quotient. Strong $*$-bisimulations are order congruences [18].

Lemma C.5 (Induced preorder and order). *Let $\simeq$ be a $*$-bisimulation. Then the relation*

$$X \preceq_{\simeq} Y \quad \text{:iff} \quad x \leq y\,, \quad \textit{for some } x \in X \textit{ and } y \in Y$$

is a preorder on $\mathcal{B}/\simeq$. If $\simeq$ is strong, $\preceq_{\simeq}$ is a partial order.

Proof. Reflexivity of $\preceq_{\simeq}$ is immediate. For transitivity, we first show that $X \preceq_{\simeq} Y$ implies that, for every $y \in Y$, there is some $x \in X$ such that $x \leq y$. From this, transitivity follows easily.

Let $X, Y \in \mathcal{B}/\simeq$ with $X \preceq_{\simeq} Y$ and let $y \in Y$. By definition, there exists $x_0 \in X$ and $y_0 \in Y$ such that $x_0 \leq y_0$. As $y, y_0 \in Y$, we have $y_0 \simeq y$. Consequently, $x_0 \leq y_0 \simeq y$ implies, by (P4), that there is some $x \in X$ with $x \leq y$.

We have shown that $\preceq$ is a preorder. Assume now that the $*$-bisimulation $\simeq$ is strong. We have to prove that $\preceq_X$ is antisymmetric.

Let $X, Y \in \mathcal{B}/\simeq$ such that $X \preceq_{\simeq} Y \preceq_{\simeq} X$. Let $y \in Y$. Since $X \preceq_{\simeq} Y$, we can use the above claim to find some $x \in X$ with $x \leq y$. Similarly, $Y \preceq_{\simeq} X$ implies that there exists $y' \in Y$ such that $y' \leq x$. But this means that $x \leq y \simeq y' \leq x$. Hence, we have $x \simeq y$ by Property (P5). By definition of $\mathcal{B}/\simeq$, this implies that $X = Y$. □

The order of the quotient induced by a strong $*$-bisimulation allows us to prove that strong $*$-bisimulations are kernels of adequate premorphisms.

Definition C.6. Let $\simeq$ be a strong $*$-bisimulation. We define the *quasi-quotient* $M_\simeq$ *induced by* $\simeq$ as follows.

Let $S := \mathcal{B}/{\simeq}$ be the quotient of $\mathcal{B}$ under $\simeq$. For $x \in \mathcal{B}$, let $[x] := \{y \in \mathcal{B} : x \simeq y\}$ be the equivalence class of x. For $[x], [y] \in S$, we define

$$[x] \preceq [y] \quad \text{:iff} \quad x' \leq y' \quad \text{for some } x' \in [x] \text{ and } y' \in [y]\,.$$

The domain of $M_\simeq$ consists of all anti-chains of S, that is, all non-empty sets $\mathcal{X} \subseteq \mathcal{P}(S)$ whose elements are pairwise incomparable with respect to $\preceq$. The product $\mathcal{X} \cdot \mathcal{Y}$ of two anti-chains $\mathcal{X}, \mathcal{Y} \in M_\simeq$ is defined by

$$\mathcal{X} \cdot \mathcal{Y} := \mu\big((\bigcup \mathcal{X}) \cdot (\bigcup \mathcal{Y})\big)\,,$$

where $\mu : \mathcal{P}(\mathcal{B}) \to M_\simeq$ is defined by

$$\mu(X) := \{[z] \in S : z \text{ a maximal element of } X\}\,.$$

Elements of $M_\simeq$ are ordered by the relation

$$\mathcal{X} \leq \mathcal{Y} \quad \text{:iff} \quad (\forall X \in \mathcal{X})(\exists Y \in \mathcal{Y})[X \preceq Y]\,.$$

Theorem C.7 (Induced premorphism). *Let $\simeq$ be a strong $*$-bisimulation with index κ. The quasi-quotient $M_\simeq$ induced by $\simeq$ is an adequately ordered monoid of size $|M_\simeq| \leq 2^\kappa$. There exists an adequate premorphism $\varphi_\simeq : \mathcal{B} \to M_\simeq$ such that*

$$x \simeq y \quad \text{iff} \quad \varphi_\simeq(x) = \varphi_\simeq(y)\,, \quad \textit{for all } x, y \in \mathcal{B}\,.$$

Proof. We start by noting that we have seen in Lemma C.5 that the relation $\preceq$ is a partial order. This implies that the relation $\leq$ of $M_\simeq$ is indeed a partial order.

Furthermore, it is easy to check that the function μ in the definition of $M_\simeq$ is a well-defined mapping $\mathcal{P}(\mathcal{B}) \to M_\simeq$ and that it is surjective and monotonic. Moreover, we also have

$$\mu(\bigcup \mathcal{X}) = \mathcal{X}\,, \quad \text{for all } \mathcal{X} \in M_\simeq\,.$$

Since the natural order on birooted trees is stable under products and since $\mu(\bigcup \mathcal{X}) = \mathcal{X}$, it is routine to check that this product is associative. Hence, it turns $M_\simeq$ into a semigroup.

Since $\simeq$ is idempotent pure, this means that 1 is the maximal element of its class [1]. Therefore,

$$\mathcal{X} \cdot [1] = \mathcal{X} = [1] \cdot \mathcal{X}\,, \quad \text{for every } \mathcal{X} \in M_\simeq\,.$$

Consequently, $M_\simeq$ is a monoid.

We observe that the partial order $\leq$ is stable under product. Moreover, for every subunit $\mathcal{X} \leq [1]$ we have $M((\bigcup \mathcal{X}) \cdot (\bigcup \mathcal{X})) = M(\bigcup \mathcal{X})$ since $xy = x \wedge y$ for every $x, y \leq 1$. Hence $\mathcal{X} \cdot \mathcal{X} = \mathcal{X}$. In other words, subunits of $M_\simeq$ are idempotents.

It follows that the product of subunits is the meet and, since $M_{\simeq}$ is finite, it is therefore an adequately ordered monoid with $\mathcal{X}^L = \prod\{\mathcal{Z} \leq 1 : \mathcal{X} \cdot \mathcal{Z} = \mathcal{X}\}$ and $\mathcal{X}^R = \prod\{\mathcal{Z} \leq 1 : \mathcal{Z} \cdot \mathcal{X} = \mathcal{X}\}$.

To conclude the proof, we define $\varphi_{\simeq} : \mathcal{B} \to M_{\simeq}$ by $\varphi_{\simeq}(x) := \{[x]\}$ for every $x \in \mathcal{B}$. By definition, the kernel of $\varphi_{\simeq}$ equals $\simeq$. It remains to show that $\varphi_{\simeq}$ is an adequate premorphism. Clearly, it is monotonic (M1).

For (M2), let $x, y \in \mathcal{B}$. Then $[xy] \in \varphi_{\simeq}(x) \cdot \varphi_{\simeq}(y)$. Hence $\varphi_{\simeq}$ is submultiplicative.

For (M3), let $x \in \mathcal{B}$. By definition, $\varphi(x^R) \leq [1]$. Moreover, we easily check that all maximal elements of $[x^R] \cdot [x]$ belong to $[x]$. Hence $\varphi([x^R]) \cdot \varphi(x) = \varphi(x)$. Assume now that there is a subunit $\mathcal{Z} \leq [1]$ such that $\mathcal{Z} \cdot \varphi(x) = \varphi(x)$. This means that $M((\bigcup \mathcal{Z}) \cdot [x]) \subseteq M([x])$. Take $x' \in [x]$ maximal. This means that there exists $z \in \bigcup \mathcal{Z}$ such that $zx' = x'$. By definition of right projections in $\mathcal{B}$, it follows that $x'^R \leq z$. Therefore, $[x^R] \preceq [z]$ and, thus, $\{[x^R]\} \leq \{[z]\} \leq \mathcal{Z}$. In other words,

$$\varphi_{\simeq}(x^R) = \{[x^R]\} = (\varphi_{\simeq}(x))^R$$

A symmetrical argument shows that $\varphi_{\simeq}$ also preserves left projections, that is, the mapping $\varphi_{\simeq}$ satisfies Property (M3).

For (M4), since $\simeq$ is a closed $*$-congruence we have $[x] * [y] = [x] \cdot [y] = [xy]$, for all $x, y \in \mathcal{B}$ such that $x * y$ is defined. Hence $\varphi_{\simeq}(x * y) = \varphi_{\simeq}(x)\varphi_{\simeq}(y)$, that is, the mapping $\varphi_{\simeq}$ satisfies Property (M4). ◻

C.3 The alternation depth of a $*$-congruence

The following definition and the lemma that follows give a rather simple characterisation of those $*$-congruences whose bisimulation refinements are strong and of finite index.

Definition C.8 (Equivalence alternating depth). Let $\simeq$ be an equivalence relation on birooted trees.

(a) The $\simeq$-*depth* of a birooted tree $x \in \mathcal{B}$ is the least ordinal $d_{\simeq}(x) \leq \omega$ such that every alternating $\simeq$-chain $x_0 \leq \cdots \leq x_n = x$ ending in x has length $n \leq d_{\simeq}(x)$.

(b) The *alternation depth* $d(\simeq)$ of the relation $\simeq$ is defined by

$$d(\simeq) := \sup_{x \in \mathcal{B}} d_{\simeq}(x)$$

The link between the alternation depth of (syntactic) closed $*$-congruences and bisimulation refinements is stated in the following lemma.

Lemma C.9. *A closed $*$-congruence $\simeq$ has finite index and alternating depth if, and only if, its bisimulation refinement $\simeq^B$ is strong and has a finite index.*

Proof. ($\Rightarrow$) Assume that $\simeq$ has finite index and finite alternating depth. By Lemma C.4, we know that $\simeq^B = \bigcap F^n(\sim)$ where $\sim := \simeq \cap \approx_{\text{srt}}$. Let $x, y \in \mathcal{B}$

and let $n := d_{\simeq}(x) \vee d_{\simeq}(y)$ be the supremum of the $\simeq$-alternating depths of x and y. By induction on n one can show that,

$$(x, y) \in F^n(\sim) \quad \text{iff} \quad (x, y) \in F^m(\sim)\,, \quad \text{for all } m \geq n\,.$$

It follows that $\simeq^B = F^N(\sim)$ where $N := d(\simeq)$ is the alternation depth of $\simeq$. Since $|F^{n+1}(\sim)| \leq 2^{|F^{n+1}(\sim)|}$, for all $n < \omega$, this shows that $\simeq^B$ has finite index.

Moreover, it follows by a straightforward induction on N that $x \simeq^B y$ implies $d_{\simeq}(x) = d_{\simeq}(y)$, for all $x, y \in \mathcal{B}$. Suppose that $x \leq y \simeq_B z \leq x$. By (P4), there exists $x' \leq z$ such that $x' \simeq_B x$. We therefore have $x' \simeq^n x$ for some $n \leq N$. Since $d_{\simeq}(x') = d_{\simeq}(x)$ this implies that $x' \simeq z$. We claim that $z \simeq^{n+1} x$. Indeed, for every $z' \leq z$ we have $z' \leq x$ since $z \leq x$. Conversely, for every $z' \leq x$, there exists $z'' \leq x'$ with $z'' \simeq^n z'$. But, $x' \leq z$ implies that $z'' \leq z$. Hence, $z \simeq^B x$. It follows that $\simeq^B$ satisfies Property (P5). That is, $\simeq^B$ is a strong $*$-bisimulation.

($\Leftarrow$) Conversely, assume that $\simeq^B$ is a strong $*$-bisimulation with finite index. Since $\simeq^B \subseteq \simeq$, this implies that $\simeq$ has finite index. Let $x_0 \leq \cdots \leq x_{n-1}$ be an alternating $\simeq$-chain We claim that n is necessarily bounded by the index of $\simeq^B$. Otherwise, there are indices $i < j$ such that $x_i \simeq^B x_j$. Hence, $x_i \leq x_{i+1} \leq x_j$ implies that $x_i \simeq^B x_{i+1}$. Consequently, we also have $x_i \simeq x_{i+1}$ since $\simeq^B \subseteq \simeq$. This contradicts the fact that the chain is alternating. We have shown that the alternation depth of $\simeq$ is bounded by the index of $\simeq^B$. □

Remark C.10. It could be the case that for every closed $*$-congruence $\simeq$, the relation $\simeq$ has finite index if, and only if, its bisimulation refinement $\simeq^B$ has finite index. Of course, this does not hold for arbitrary equivalence relations on ordered sets. consider the set ω with the standard order and the relation $\simeq$ of having the same parity. But we suspect that this could be true in the meet-lattice structure of $\mathcal{B}$ with the natural order.

C.4 Complete proof of Theorem 5.7

(3) $\Rightarrow$ (2) follows by Lemma C.9 and (2) $\Rightarrow$ (1) was proved in Theorem C.7. Hence, it remains to prove (1) $\Rightarrow$ (3).

Lemma C.11. *The syntactic congruence $\simeq_X$ of a quasi-recognisable language X of birooted trees has finite index and alternation depth.*

Proof. Suppose that X is quasi-recognisable. Let $\varphi : \mathcal{B} \to M$ be an adequate premorphisms into a finite adequately ordered monoid M recognising X and let $\simeq_\varphi$ be the kernel of φ. Clearly, the relation $\sim := \simeq_\varphi \cap \approx_{\mathrm{rt}}$ is a closed $*$-congruence of finite index saturating X. Consequently, $\sim \subseteq \simeq_X$.

For a contradiction, suppose that the relation $\simeq_X$ does not have finite alternation depth. Then there exist alternating $\simeq_X$-chains of arbitrary length. As the relation $\approx_{\mathrm{rt}}$ has finite index, this also means that there are such chains whose elements are $\approx_{\mathrm{rt}}$-equivalent, i.e., with the same root types. Fix an alternating $\simeq_X$-chain $x_0 \leq \cdots \leq x_n$ where all x_i are $\approx_{\mathrm{rt}}$-equivalent and n is greater than the index of $\sim$. Then there are indices $i < j$ such that $x_i \sim x_j$. In particular, we have $x_i \simeq_\varphi x_j$. By monotonicity of φ, it follows that $x_i \simeq_\varphi x_{i+1}$. As all elements have the same root type, this implies that $x_i \sim x_{i+1}$. Since $\sim$ is included in $\simeq_X$, it follows that $x_i \simeq_X x_{i+1}$. This contradicts the assumption that $x_0, \dots, x_n$ is alternating. $\square$

D Strong recognisability and minimal recognisers

All the material developed above also leads us to the notion of a strongly quasi-recognisable language that, although equi-expressive, admits minimal recognisers.

D.1 The category AdP

A priori, the disjoint product is only defined over $\mathcal{B}$. Hence there is not a defined notion of an adequate premorphism between *arbitrary* adequately ordered monoids. A simple remedy is to enrich the notion of an adequately ordered monoid by a given disjoint product operation.

Definition D.1 (Enriched adequately ordered monoid). An *enriched adequately ordered monoid* is an adequately ordered monoid M equipped with a partial product $*$ that is compatible with the usual product in the sense that, if $x * y$ is defined, then $x * y = xy$.

The notion of an adequate premorphism can be extended as follows:

Definition D.2 (Adequate premorphism). An *adequate premorphism* is a mapping $\varphi : M \to N$ between two enriched adequately ordered monoids M and N such that

(M1) $x \leq y$ implies $\varphi(x) \leq \varphi(y)$,
(M2) $\varphi(xy) \leq \varphi(x)\varphi(y)$,
(M3) $\varphi(x^L) = (\varphi(x))^L$ and $\varphi(x^R) = (\varphi(x))^R$,
(M4) $\exists x * y$ implies $\exists \varphi(x) * \varphi(y)$ and $\varphi(x * y) = \varphi(x) * \varphi(y)$.

Remark D.3. By default, for every adequate premorphism $\varphi : \mathcal{B} \to M$, the disjoint product in M is defined as follows: the product xy of two elements $x, y \in M$ is disjoint, which is denoted by $x * y$ when there exists $x' \in \varphi^{-1}(x)$ and $y' \in \varphi^{-1}(x)$ such that $\exists x' * y' \in \mathcal{B}$.

Lemma D.4. *The identity* $\mathrm{id} : M \to M$ *is an adequate premorphism for every adequately ordered monoid* M*. The composition* $\varphi_2 \circ \varphi_1 : M_1 \to M_3$ *of two adequate premorphisms* $\varphi_1 : M_1 \to M_2$ *and* $\varphi_2 : M_2 \to M_3$ *is again an adequate premorphism. Consequently, we obtain a category* **AdP** *of adequately ordered monoids with adequate premorphisms.*

The question of finding a syntactic adequately ordered monoid recognising a given language $X \subseteq \mathcal{B}$ amounts to finding a terminal object in the subcategory of **AdP** that is induced by the adequately ordered monoids that recognise X.

However, it seems that such a terminal object does not exist. Indeed, the kernel $\simeq_\varphi$ defined by an adequate premorphism φ is a *convex* $*$-congruence, i.e., $x \leq y \leq z$ and $x \simeq_\varphi z$ implies $x \simeq_\varphi y$. In general, there may not exist a greatest convex relation saturating a language as illustrated, for instance, by the language $X = \{a, c, d\}$ on the four element set $S = \{a, b, c, d\}$ where the order is defined by $a < b < c$ while d is incomparable.

D.2 The category StAdP

The techniques developed in the proof of Theorem C.7 lead us to the following strengthening of the notion of quasi-recognisability and a corresponding refinement of the category **AdP** where syntactic recognisers exist.

Definition D.5 (Strongly adequate premorphisms). An adequate premorphism $\varphi : \mathcal{B} \to M$ is *strongly adequate* if, additionally, its satisfies the following properties.

(M5) $\varphi(x) \leq 1$ implies $x \leq 1$.
(M6) $\exists x * y$ implies $\exists x' * y'$, for all $x' \in [x]_\varphi$ and $y' \in [y]_\varphi$.
(M7) $\varphi(x) \leq \varphi(y)$ implies $x' \leq y$, for some $x' \in [x]_\varphi$.
(M8) $\varphi(x) \leq \varphi(y)\varphi(z)$ implies $\varphi(x) \leq \varphi(y'z')$, for some $y' \in [y]_\varphi$ and $z' \in [z]_\varphi$.

Remark D.6. Property (M5) ensures that the kernel of φ is idempotent pure. Property (M6) ensures that it is closed. Property (M7) ensures that φ is *strongly monotonic* in the sense of [15]. Property (M8) ensures that every element in the submonoid $\langle\varphi(\mathcal{B})\rangle$ of M generated by the image of φ is compatible in some sense with the ordered monoid structure of $\mathcal{B}$. This last axiom, slightly technical, seems crucial for the normalisation Lemma D.8 below.

Lemma D.7. *Strongly adequate premorphisms and adequately ordered monoids (with given disjoint products) form a subcategory* **StAdP** *of the category* **AdP** *of adequately ordered monoids and adequate premorphisms.*

Lemma D.8. *Let $\varphi : \mathcal{B} \to M$ be a strongly adequate morphism with finite codomain M, let $\simeq_\varphi$ be its kernel, and let $M_{\simeq_\varphi}$ be the quasi-quotient induced by $\simeq_\varphi$. Then $\simeq_\varphi$ is a strong $*$-bisimulation and there is a strongly adequate premorphism $\iota : M \to M_{\simeq_\varphi}$ such that $\varphi_\simeq = \iota \circ \varphi$ where, for every $a, b \in M$, the disjoint product $a * b := ab$ is defined in M by $\exists a * b$ if $\exists x * y$ for some*[3] *$x \in \varphi^{-1}(a)$ and $y \in \varphi^{-1}(b)$.*

Proof. The fact that φ is an adequate premorphism guarantees that its kernel $\simeq_\varphi$ is a $*$-congruence. Property (M5) ensures that it is idempotent pure. Property (M6) ensures it is closed, and Property (M7) ensures (P4). Consequently, it is a

[3] or, by (M6), for all

$*$-bisimulation. Property (P5) is immediate from the fact that φ is monotonic. Indeed, given elements $x, y, z \in \mathcal{B}$ with $x \le y \simeq_\varphi z \le x$, we have $\varphi(x) \le \varphi(y)$, $\varphi(y) = \varphi(z)$ and $\varphi(z) \le \varphi(x)$. By transitivity, it follows that $\varphi(x) \le \varphi(y) \le \varphi(x)$. Thus, $x \simeq_\varphi y$. Therefore, $\simeq_\varphi$ is a strong $*$-bisimulation. Hence we can use Theorem C.7 to obtain the desired premorphism $\varphi_{\simeq_\varphi} : \mathcal{B} \to M_{\simeq_\varphi}$.

For $x \in \mathcal{B}$, set $[x] := \varphi^{-1}(\varphi(x)) \in \mathcal{B}/{\simeq_\varphi}$. Let $\iota : M \to M_{\simeq_\varphi}$ be defined by

$$\iota(a) := \mu(\varphi^{-1}(a^{\downarrow}))$$

where $a^{\downarrow} := \{b \in M : b \le a\}$ and μ is the function from Definition C.6.

The monotonicity of μ implies (M1), i.e., the monotonicity of ι. Property (M8) of φ ensures that ι is submultiplicative (M2). Since the mapping ι is clearly idempotent pure (M5) and closed (M6) arguments similar to the ones in the proof of Theorem C.7 ensure that Properties (M3) and (M4) are satisfied. Properties (M7) and (M8) for ι are inherited from the same properties of φ.

In other words, ι is a strongly adequate premorphism from M into $M_{\simeq_\varphi}$. Moreover, for every $x \in \mathcal{B}$, by definition of $\varphi_{\simeq_\varphi}$, we have $\iota(\varphi(x)) = \varphi_{\simeq_\varphi}(x)$. Hence, $\varphi_{\simeq_\varphi} = \iota \circ \varphi$. □

We obtain the following characterisation of quasi-recognisable languages.

Theorem D.9. *A language $X \subseteq \mathcal{B}$ of birooted trees is quasi-recognisable if, and only if, it is quasi-recognisable by a strongly adequate premorphism.*

Proof. Assume that X is a quasi-recognisable language. By Theorem 5.7, it is recognised by the strongly adequate premorphism $\varphi_{\simeq^B_X}$. As strongly adequate premorphisms are restrictions of adequate premorphisms, the claim follows.

Theorem D.10. *Let $X \subseteq \mathcal{B}$ be a quasi-recognisable language. The strongly adequate premorphism $\varphi_{\simeq^B_X} : \mathcal{B} \to M_{\simeq^B_X}$ of Theorem C.7 is weakly terminal in the category of all strongly adequate premorphisms recognising X (considered as a subcategory of the comma category $(\mathcal{B}, \mathbf{StAdP})$).*

Proof. Let $\varphi : \mathcal{B} \to M$ be a strongly adequate premorphism recognising X. We have to prove that there is a strongly adequate premorphism $\psi : M \to M_{\simeq^B_X}$ such that $\varphi_{\simeq^B_X} = \psi \circ \varphi$.

Let $\simeq$ be the kernel of φ and let $\varphi_\simeq : \mathcal{B} \to M_\simeq$ and $\iota : M \to M_\simeq$ be the strongly adequate premorphisms given by Lemma D.8. As $\simeq$ is a strong $*$-bisimulation included in the syntactic congruence $\simeq_X$ of the language X, we have $\simeq \subseteq \simeq^B_X$. Hence, we can define a map $\eta : M_\simeq \to M_{\simeq_X}$ by

$$\eta(\mathcal{X}) := \mu_{\simeq_X}(\bigcup \mathcal{X})$$

that lifts the inclusion surjective mapping $\mathcal{B}/{\simeq_\varphi} \to \mathcal{B}_{\simeq^B_X}$ to a strongly adequate premorphism $M_\simeq \to M_{\simeq^B_X}$. The composition $\psi := \eta \circ \iota$ is the desired premorphism (see Figure 2). □

When X is quasi-recognisable, the monoid $M_{\simeq^B_X}$ can therefore be considered as the *syntactic adequately ordered monoid* of X.

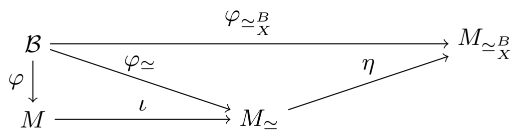

Fig. 2. Minimal strong recogniser.

Remark D.11. Note that we only prove that $M_{\simeq^B_X}$ is *weakly terminal* in the sense that the strongly adequate premorphism $\psi = \eta \circ \iota$ found in the above proof may not be unique. This comes from the fact that, in general, the adequate premorphism φ may not be surjective.

However, in the category of the partial algebras of the form $\varphi/\simeq$ ordered by $\preceq_{\simeq}$ as defined in Lemma C.5 for every kernel $\simeq$ of a strongly adequate premorphism $\varphi : \mathcal{B} \to M$ recognizing X, and the related (monotonic) partial algebra morphism that becomes surjective, we easily check that the object $\mathcal{B}/\simeq^B_X$ is terminal in the usual sense.