Locality and Centrality: The Variety ZG

Antoine Amarilli
LTCI, Télécom Paris, Institut polytechnique de Paris, France

Charles Paperman
LINKS, CRIStAL, Université de Lille, INRIA, France

Abstract

We study the variety $\mathbf{ZG}$ of monoids where the elements that belong to a group are *central*, i.e., commute with all other elements. We show that $\mathbf{ZG}$ is *local*, that is, the semidirect product $\mathbf{ZG} * \mathbf{D}$ of $\mathbf{ZG}$ by definite semigroups is equal to $\mathbf{LZG}$, the variety of semigroups where all local monoids are in $\mathbf{ZG}$. Our main result is thus: $\mathbf{ZG} * \mathbf{D} = \mathbf{LZG}$. We prove this result using Straubing's delay theorem, by considering paths in the category of idempotents. In the process, we obtain the characterization $\mathbf{ZG} = \mathbf{MNil} \vee \mathbf{Com}$, and also characterize the $\mathbf{ZG}$ languages, i.e., the languages whose syntactic monoid is in $\mathbf{ZG}$: they are precisely the languages that are finite unions of disjoint shuffles of singleton languages and regular commutative languages.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases regular language, variety, locality

Acknowledgements We thank Jean-Éric Pin and Jorge Almeida for their fruitful advice.

1 Introduction

In this paper, we study a variety of monoids called $\mathbf{ZG}$. It is defined by enforcing that the elements of the monoid that belong to a group are *central*, i.e., commute with all other elements of the monoid. The notation $\mathbf{ZG}$ thus stands for *Zentral Group*, inspired by the classical notion of centrality in group theory. We can also define $\mathbf{ZG}$ with the equation $x^{\omega+1}y = yx^{\omega+1}$ on all elements x and y, where ω is the idempotent power of the monoid.

The variety $\mathbf{ZG}$ has been introduced by Auinger [5] as a subvariety of interest of the broader class $\mathbf{ZE}$ of semigroups where the idempotent elements are central. The study of $\mathbf{ZE}$ was initiated by Straubing [21]. Straubing shows in particular that the variety $\mathbf{MNil}$ (called simply $\mathbf{V}$ in the paper) of regular languages generated by finite languages is exactly the variety of aperiodic monoids in $\mathbf{ZE}$. From this, a systematic investigation of the subclasses of $\mathbf{ZE}$ was started by Almeida and pursued by Auinger: see [2, page 211] and [5, 4].

Our specific motivation to explore $\mathbf{ZG}$ comes from our study of the *dynamic membership problem for regular languages*. In this problem [18], we want to handle update operations on an input word while maintaining the information of whether it belongs to a fixed regular language. In a companion paper also submitted to ICALP'21 [3], we identify a variant of $\mathbf{ZG}$ as a plausible tractability boundary characterizing the languages for which every update operation can be handled in $O(1)$. Specifically, this variant can be defined as the so-called *semidirect product* of $\mathbf{ZG}$ by definite ($\mathbf{D}$) semigroups, which we denote $\mathbf{ZG} * \mathbf{D}$.

This semidirect product operation on varieties, which we use to define $\mathbf{ZG} * \mathbf{D}$, intuitively corresponds to composing finite automata via a kind of cascade operation. Its study is the subject of a large portion of semigroup theory, inspired by the classical study of the semidirect product in group theory. There are also known results to understand specifically the semidirect product *by* $\mathbf{D}$. For instance, the *Derived category theorem* [28] studies it as a decisive step towards proving the decidability of membership to an arbitrary semidirect product, i.e., deciding if a given monoid belongs to the product. The product by $\mathbf{D}$ also arises naturally in several other contexts: the dotdepth hierarchies [22], the circuit complexity of regular languages [23], or the study of the successor relations in first-order logic [26, 27, 14].

Understanding this product with $\mathbf{D}$ is notoriously complicated. For instance, it requires specific dedicated work for some varieties like $\mathbf{J}$ or $\mathbf{Com}$ [13, 12, 25]. Also, this product does not preserve the decidability of membership, i.e., Auinger [6] proved that there are varieties $\mathbf{V}$ such that membership in $\mathbf{V}$ is decidable, but the analogous problem for $\mathbf{V} * \mathbf{D}$ is undecidable. For the specific case of the varieties $\mathbf{ZG}$, $\mathbf{ZE}$, or even $\mathbf{MNil}$, we are not aware of prior results describing their semidirect product with $\mathbf{D}$.

Locality. Existing work has nevertheless identified some cases where the $*\mathbf{D}$ operator can be simplified to a much nicer *local operator*, that preserves the decidability of membership and is easier to understand. For any semigroup S, the *local monoids* of S are the subsemigroups of S of the shape eSe with e an idempotent element of S. For a variety $\mathbf{V}$, we say that a semigroup belongs to $\mathbf{LV}$ if all its local monoids are in $\mathbf{V}$. It is not hard to notice that the variety $\mathbf{V} * \mathbf{D}$ is always a subvariety of $\mathbf{LV}$, i.e., that every monoid in $\mathbf{V} * \mathbf{D}$ must also be in $\mathbf{LV}$. In some cases, we can show a *locality result* stating that the other direction also holds, so that $\mathbf{V} * \mathbf{D} = \mathbf{LV}$. In those cases we say that the variety $\mathbf{V}$ is *local*. The locality of the variety of monoids $\mathbf{DA}$ [2] is a famous result that has deep implications in logic and complexity [26, 9, 11] and has inspired recent follow-up work [17]. Locality results are also known for other varieties, for instance the variety of *semi-lattice monoids* (monoids that are both idempotent and commutative) [15, 8], any *sub-varieties of groups* [22, Theorem 10.2], or the *$\mathcal{R}$-trivial variety* [20, 19, 24]. This suggests an angle of attack to understanding the variety $\mathbf{ZG} * \mathbf{D}$: establishing a locality result of this type for $\mathbf{ZG}$.

Contributions. Our main result in this paper is to show the locality of the variety $\mathbf{ZG}$:

▶ **Theorem 1.1.** *We have* $\mathbf{LZG} = \mathbf{ZG} * \mathbf{D}$.

In the process of showing this result, we obtain a characterization of $\mathbf{ZG}$-congruences, i.e., congruences $\sim$ on Σ^* where the quotient $\Sigma^*/\sim$ is a monoid of $\mathbf{ZG}$. We show that they are always refined by a so-called *n-congruence*, which identifies the number of occurrences of the *frequent letters* (the ones occurring $> n$ times in the word) modulo n, and also identifies the exact subword formed by the *rare letters* (the ones occurring $\leq n$ times). Thanks to this (Theorem 3.4), we also obtain a characterization of the languages of $\mathbf{ZG}$, i.e., the languages whose syntactic monoid is in $\mathbf{ZG}$: they are exactly the finite unions of disjoint shuffles of singleton languages and commutative languages (Corollary 3.5). We also characterize $\mathbf{ZG}$ as a variety of monoids: $\mathbf{ZG} = \mathbf{MNil} \vee \mathbf{Com}$, for $\mathbf{MNil}$ defined in [21] and $\mathbf{Com}$ the variety of commutative monoids.

Paper structure. We give preliminaries in Section 2 and formally define the variety $\mathbf{ZG}$. We then give in Section 3 our characterizations of $\mathbf{ZG}$ via the so-called n-congruence. We then define in Section 4 the varieties $\mathbf{ZG} * \mathbf{D}$ and $\mathbf{LZG}$ used in Theorem 1.1, which we prove in the rest of the paper. We first introduce the framework of Straubing's delay theorem used for our proof in Section 5, and rephrase our result as a claim (Claim 5.6) on paths in the category of idempotents. We then study in Section 6 how to pick a sufficiently large value of n as a choice of our n-congruence, show in Section 7 two lemmas on paths in the idempotent category, and finish the proof in Section 8. We conclude in Section 9.

2 Preliminaries

For a complete presentation of the basic concepts (automata, monoids, semigroups, groups, etc.) the reader can refer to the book of J. E. Pin [16] or to the more recent lecture notes [29].

All semigroups, groups, and monoids that we consider are finite.

Semigroups and varieties. For a semigroup S, we call $x \in S$ *idempotent* if $x^2 = x$. We call the *idempotent power* of $x \in S$ the unique idempotent element which is a power of x. (This means that x is idempotent iff it is its own idempotent power.) Now, the *idempotent power of S* is an integer ω such that for any element $x \in S$, the element x^ω is the idempotent power of x. We write $x^{\omega+k}$ for any $k \in \mathbb{Z}$ to mean $x^{\omega+k'}$ where k' is the remainder of k in the integer division by ω.

We will use notions from formal language theory for some of our definitions. We denote by Σ an alphabet and by Σ^* the set of all finite words on Σ. We denote by ϵ the empty word. For $w \in \Sigma^*$, we denote by $|w|$ the length of w. For $u, v \in \Sigma^*$, we say that u is a *subword* of v if there is $0 \leq n \leq |u|$ and $1 \leq i_1 < \cdots < i_n \leq |v|$ such that $u = v_{i_1} \cdots v_{i_n}$. For $w \in \Sigma^*$ and $a \in \Sigma$, we denote by $|w|_a$ the number of occurrences of a in w. A *language* L is a subset of Σ^*. A *variety of (regular) languages* is a class of regular languages which is closed under Boolean operations, left and right derivatives, and inverse homomorphisms. A *non-erasing variety* is closed under the same operations, except we only require closure under inverse *non-erasing* homomorphisms, that is, those that do not map any letter of Σ to ϵ.

A *variety of monoids* (resp., *variety of semigroups*) is a class of monoids (resp., *semigroups*) closed under direct product, quotient, and submonoid (resp., *subsemigroup*). We recall that Eilenberg's theorem [10] gives a one-to-one correspondence between varieties of languages and varieties of monoids. A similar one-to-one correspondence exists between non-erasing varieties of languages and varieties of semigroups. In the following, we abuse notation and identify varieties of monoids with varieties of languages following this correspondence. To be more precise, we say that a monoid M *recognizes* a language L if there exists a morphism $\eta : \Sigma^* \to M$ such that $L = \eta^{-1}(\eta(L))$. Eilenberg's correspondence then says that, when considering a variety $\mathbf{V}$ of monoids, the languages recognized by monoids in $\mathbf{V}$ belong to the corresponding variety of languages. Eilenberg's correspondence extends to a correspondence between variety of semigroups and non-erasing variety of languages.

Congruences. A *finite index congruence* on a finite alphabet Σ is a congruence on Σ^* that has a finite number of equivalence classes. For a given finite index congruence $\sim$, the quotient $\Sigma^*/\sim$ is a finite monoid, whose law corresponds to concatenation over Σ^* and whose neutral element is the class of the empty word. The *syntactic monoid* of a regular language over Σ^* is the quotient by the syntactic congruence for the language, which is a finite index congruence because the language is regular. Letting $\mathbf{V}$ be a variety of monoids, we say that a finite index congruence $\sim$ on Σ is a $\mathbf{V}$-*congruence* if the quotient $\Sigma^*/\sim$ is a monoid in $\mathbf{V}$. For a given $\mathbf{V}$-congruence $\sim$, the map $\eta : \Sigma^* \to \Sigma^*/\sim$, defined by associating each word with its equivalent class, is an onto morphism within a monoid of $\mathbf{V}$. Hence, each equivalence class is a language of $\mathbf{V}$, since it is recognized by $\Sigma^*/\sim$.

The variety ZG. In this paper, we study the variety of monoids $\mathbf{ZG}$ defined by the equation: $x^{\omega+1}y = yx^{\omega+1}$ for all $x, y \in M$. Intuitively, this says that the elements of the form $x^{\omega+1}$ are central, i.e., commute with all other elements. This clearly implies the same for elements of the form $x^{\omega+k}$ for any $k \in \mathbb{Z}$, as we will implicitly use throughout the paper:

▶ **Claim 2.1.** *For any monoid M in $\mathbf{ZG}$, for $x, y \in M$, and $k \in \mathbb{Z}$, we have: $x^{\omega+k}y = yx^{\omega+k}$.*

Note that these elements are precisely the monoid elements that are within a (possibly trivial) subgroup. This motivates the name $\mathbf{ZG}$, which stands for "Zentral Group": it follows the traditional notation $Z(\cdot)$ for central subgroups, and extends the variety **ZE** introduced in [1, p211] which only requires idempotents to be central. Thus, we have $\mathbf{ZG} \subsetneq \mathbf{ZE}$, and non-commutative groups are examples of monoids that are in **ZE** but not in **ZG**.

By Eilenberg's theorem, **ZG** then defines a variety of regular languages, namely the languages whose syntactic monoid is in **ZG**. Note that any regular commutative language is in **ZE** and in **ZG** (as is clear from the equation), and any finite language is also in **ZE** and in **ZG** (their unique group element is a zero so it commutes with everything).

3 Characterizations of ZG

In this section, we present our characterizations of **ZG**, which we will use to prove Theorem 1.1. We will show that **ZG** is intimately linked to a congruence on words called the n-congruence. Intuitively, two words are identified by this congruence if the subwords of the *rare letters* (occurring less than n times) are the same, and the numbers of occurrences of the *frequent letters* (occurring more than n times) are congruent modulo n. Formally:

▶ **Definition 3.1** (Rare and frequent letters, n-congruence). *Fix an alphabet Σ and a word $w \in \Sigma^*$. Given a threshold $n \in \mathbb{N}$, we call $a \in \Sigma$* rare *in w if $|w|_a \leq n$, and* frequent *in w if $|w|_a > n$. We define the* rare subword *$w_{\leq n}$ to be the subword of w obtained by keeping only the rare letters of w, i.e., the subword of w where we keep precisely the letters of the (possibly empty)* rare alphabet *$\{a \in \Sigma \mid |w|_a \leq n\}$.*

*For $n > 0$, the n-*congruence *$\sim_n$ is defined by writing $u \sim_n v$ for $u, v \in \Sigma^*$ iff:*
- *The rare subwords are equal: $u_{\leq n} = v_{\leq n}$;*
- *The rare alphabets are the same: for all $a \in \Sigma$, we have $|u|_a > n$ iff $|v|_a > n$;*
- *The number of occurrences modulo a are the same: for all $a \in \Sigma$ such that $|u|_a > n$ (and $|v|_a > n$), we have that $|u|_a$ and $|v|_a$ are congruent modulo n.*

We first remark that two n-equivalent words are also m-equivalent for any divisor m of n:

▶ **Claim 3.2.** *For any alphabet Σ, for any $n > 0$, for any $m > 0$, if m is a multiple of n then the m-congruence refines the n-congruence.*

What is more, observe that n-congruences are a particular case of **ZG**-congruences:

▶ **Claim 3.3.** *For any alphabet Σ and $n > 0$, the n-congruence over Σ^* is a* **ZG***-congruence.*

Proof sketch. Considering any equivalence class of the n-congruence, we can enforce in **ZG** the (commutative) conditions on the number of occurrences of the frequent letters, and interleave this with the requirement on the rare subword. ◀

The goal of this section is to show the following result. Intuitively, it states that **ZG**-congruences are always refined by a sufficiently large n-congruence. Formally:

▶ **Theorem 3.4.** *Consider any* **ZG***-congruence $\sim$ over Σ^* and consider its associated monoid $M := \Sigma^*/\sim$. Let $n := (|M| + 1) \cdot \omega$ with ω the idempotent power of M. Then the congruence $\sim$ is refined by the n-congruence on Σ.*

Before proving this result, we spell out some of its consequences. The most important one is that Theorem 3.4 implies a characterization of languages in **ZG**, which is similar to the one obtained by Straubing in [21] for the variety **MNil**. To define **MNil**, define a *nilpotent semigroup* S to be a semigroup satisfying the equation $x^\omega y = yx^\omega = x^\omega$, and let S^1 be the monoid obtained from S by adding an *identity element* 1 to S (i.e., an element with $1x = x1 = x$ for all $x \in S$) if S does not have one. The variety **MNil** is generated by semigroups of the form S^1 for S a nilpotent semigroup. It was shown in [21] that the languages of **MNil** are *disjoint monomials*, that is, Boolean combinations of languages of the shape $B^*a_1B^*a_2\cdots a_kB^*$ with $B \cap \{a_1, \ldots, a_k\} = \emptyset$.

Our analogous characterization for **ZG** is the following, obtained via Theorem 3.4:

▶ **Corollary 3.5.** *Any* **ZG** *language L can be expressed as a finite union of languages of the form $B^*a_1B^*a_2\cdots a_kB^* \cap K$ where $\{a_1, \ldots, a_k\} \cap B = \emptyset$ and K is a regular commutative language.*

Equivalently, we can say that every language of **ZG** is a finite union of *disjoint shuffles* of a singleton language (containing only one word) and of a regular commutative language, where the *disjoint shuffle* operator interleaves two languages (i.e., it describes the sets of words that can be achieved as interleavings of one word in each language) while requiring that the two languages are on disjoint alphabets. We sketch the proof of Corollary 3.5:

Proof sketch. We know that the syntactic congruence of a **ZG** language is a **ZG**-congruence, so by Theorem 3.4 it is refined by an n-congruence; and the equivalence classes of an n-congruence can be expressed as stated. ◀

This corollary also implies a characterization of the variety of monoids **ZG**. To define it, we use the *join* of two varieties **V** and **W**, denoted by $\mathbf{V} \vee \mathbf{W}$, which is the variety of monoids generated by the monoids of **V** and those of **W**. Alternatively, the join is the smallest variety containing both varieties. We then have:

▶ **Corollary 3.6.** *The variety* **ZG** *is generated by commutative monoids and monoids of the shape S^1 with S a nilpotent semigroup. In other words, we have:* $\mathbf{ZG} = \mathbf{MNil} \vee \mathbf{Com}$.

Proof. Clearly **ZG** contains both **Com** and **MNil**. Furthermore, by Corollary 3.5, any language in **ZG** is a union of intersections of a language in **MNil** and a language in **Com**. Hence, it is in the variety generated by them, concluding the proof. ◀

On a different note, we will also use Theorem 3.4 to show a technical result that will be useful later. It intuitively allows us to regroup and move arbitrary elements:

▶ **Corollary 3.7.** *For any monoid M in* **ZG**, *letting $n \geq (|M| + 1) \cdot \omega$, for any element m of M and elements $m_1, \ldots, m_n$ of M, we have*

$$m \cdot m_1 \cdot m \cdot m_2 \cdot m \cdots m \cdot m_n \cdot m \cdot m_n \cdot m = m^{n+1} \cdot m_1 \cdots m_n.$$

Having spelled out the consequences of Theorem 3.4, we turn to its proof. It crucially relies on a general result about **ZG** that we will use in several proofs, and which is shown simply by manipulating equations:

▶ **Lemma 3.8.** *Let M be a monoid of* **ZG**, *let ω be the idempotent power, and let $x, y \in M$. Then we have: $(xy)^\omega = x^\omega y^\omega$.*

We can then sketch the proof of Theorem 3.4:

Proof sketch. From Lemma 3.8, we can rewrite any **ZG**-congruence to a normal form, where frequent letters are moved to the end of the word. Looking at this form, we can show that n-equivalence implies equivalence by the **ZG**-congruence. ◀

4 Defining $\mathrm{ZG} * \mathrm{D}$ and LZG, and Result Statement

We have given our characterizations of **ZG** and presented some preliminary results. We now move to the definition of **LZG** and $\mathbf{ZG} * \mathbf{D}$, to show that $\mathbf{LZG} = \mathbf{ZG} * \mathbf{D}$ (Theorem 1.1).

$\mathbf{ZG} * \mathbf{D}$. We denote by $\mathbf{D}$ the variety of the *definite semigroups*, i.e., the semigroups satisfying the equation $yx^\omega = x^\omega$. The variety of semigroups $\mathbf{ZG} * \mathbf{D}$ is intuitively defined by taking the *semidirect product* of monoids in **ZG** and semigroups in **D**. Although we will not use directly its definition in this paper, we recall it for completeness. Given two semigroups S and T, a *semigroup action* of S on T is defined by a map $\mathrm{act} : S \times T \to T$ such that $\mathrm{act}(s_1, \mathrm{act}(s_2, t)) = \mathrm{act}(s_1 s_2, t)$ and $\mathrm{act}(s, t_1 t_2) = \mathrm{act}(s, t_1)\mathrm{act}(s, t_2)$. We then define the product $\circ_{\mathrm{act}}$ on the set $T \times S$ as follows: for all s_1, s_2 in S and t_1, t_2 in T, we have: $(t_1, s_1) \circ_{\mathrm{act}} (t_2, s_2) := (t_1 \mathrm{act}(s_1, t_2), s_1 s_2)$. The set $T \times S$ equipped with the product $\circ_{\mathrm{act}}$ is a semigroup called the *semidirect product* of S by T, denoted $T \circ_{\mathrm{act}} S$. The variety $\mathbf{ZG} * \mathbf{D}$ is then the variety generated by the semidirect products of monoids in **ZG** and semigroups in **D**. Remark that this operation is equivalent to the *wreath product of varieties*. Furthermore we could equivalently replace **D** by the variety of *locally trivial semigroups*. We refer to [22] for a detailed presentation on this subject.

LZG. Last, we introduce the variety **LZG**. This is the variety of semigroups S such that, for every idempotent e of S, the submonoid eSe of elements that can be written as ese for some $s \in S$ is in **ZG**. In other words, a semigroup is in **LZG** iff it satisfies the following equation: for any x, y and z in S, we have:

$$(z^\omega x z^\omega)^{\omega+1}(z^\omega y z^\omega) = (z^\omega y z^\omega)(z^\omega x z^\omega)^{\omega+1}.$$

Again, following Eilenberg's theorem, we also see **LZG** as a non-erasing variety of languages.

Main result. Our main result, stated in Theorem 1.1, is that $\mathbf{ZG} * \mathbf{D}$ and **LZG** are actually the same variety. To prove this result, we will first present the general framework of Straubing's delay theorem in the next section and show the easy inclusion $\mathbf{ZG} * \mathbf{D} \subseteq \mathbf{LZG}$, before embarking with the actual proof.

5 Straubing's Delay Theorem

To show our main result, we will use Straubing's delay theorem from [22]. We first give some prerequisites to recall this result. To this end, let us first define a general notion of *category*:

▶ **Definition 5.1.** *A* finite category *on a set of objects O is a finite multiset C over $O \times O$ of* arrows*, each arrow going from an object to another object (possibly itself), equipped with a composition law: for any arrows $a, b \in C$ such that we can write $a = (o, o')$ and $b = (o', o'')$, the composition law gives us ab which must be an arrow of the form $ab = (o, o'')$. Further, this composition law must be associative. What is more, we require that for any object o, there exists an arrow (o, o) which is the identity for all elements with which it can be combined (hence these arrows are in particular unique).*

We now define the notion of *idempotent category* of a semigroup. The idempotent category of a language is then defined as that of its syntactic semigroup.

▶ **Definition 5.2** (Idempotent category). *Let S be a semigroup. The* idempotent category S_{E} *of S is the finite category defined as follows:*

- *The objects of S are the idempotents of S.*
- *For any idempotents e and f and any element x of S such that $x \in eSf$, we have an arrow labeled by x going from e to f, which we will denote by (e, x, f).*

The composition law of the category is $(e, x, f)(f, y, g) = (e, xy, g)$. Note that it is clearly associative thanks to the associativity of the composition law on S.

Let us now study the idempotent category of a semigroup S in more detail. Let $\mathrm{Arrows}(S_{\mathrm{E}}) = \{(e, x, f) \mid x \in eSf\}$ be the set of arrows of the idempotent category. For brevity, we denote this set simply as B.

A *path* of S_{E} is a nonempty word of B^* whose sequence of arrows is *valid*, i.e., the end object of each arrow except the last one is equal to the starting object of the next arrow. Because S_{E} is a category, each path is equivalent to an element of the category, i.e., composing the arrows of the path according to the composition law of the category will give one arrow of the category, whose starting and ending objects will be the *starting object of the path* (i.e., that of the first arrow) and the *ending object of the path* (i.e., of the last arrow). Two paths are *coterminal* if they have the same starting and end object. Two paths p_1 and p_2 are S_{E}*-equal* if they evaluate to the same category element, which we write $p_1 \equiv p_2$. Note that if two paths are S_{E}-equal then they must be coterminal. A *loop* is a path whose starting and ending objects are the same.

A *congruence* is an equivalence relation $\sim$ over B^* which satisfies *compositionality*, i.e., it is compatible with the concatenation of words in the following sense: for any words x, y, z, and t of B^*, if $x \sim y$ and $z \sim t$, then $xz \sim yt$. Note that the relation is also defined on words of B^* that are not valid, i.e., do not correspond to paths; and compositionality also applies to such words.

▶ **Definition 5.3** (Compatible congruence). *A congruence $\sim$ on B^* is* compatible *with S_{E} iff for any two coterminal paths p_1 and p_2 of S_{E} such that $p_1 \sim p_2$, then $p_1 \equiv p_2$. In other words, $\sim$ is compatible with S_{E} iff, on words of B^* that are coterminal paths, it refines S_{E}-equality.*

Recall the notion of a $\mathbf{ZG}$-congruence from Section 2. We are now ready to state Straubing's delay theorem. The theorem applies to any variety, but we state it specifically for $\mathbf{ZG}$ for our purposes. The theorem gives us an alternative characterization of $\mathbf{ZG} * \mathbf{D}$:

▶ **Theorem 5.4** (Straubing's delay theorem (Theorem 5.2 of [22])). *A language L is in $\mathbf{ZG} * \mathbf{D}$ iff, writing S_{E} the idempotent category of L and defining $B := \mathrm{Arrows}(S_{\mathrm{E}})$ as above, there exists a $\mathbf{ZG}$-congruence on B^* which is compatible with S_{E}.*

Using our notion of n-congruence, via Theorem 3.4 and Claim 3.3, we rephrase it again:

▶ **Corollary 5.5.** *A language L is in $\mathbf{ZG} * \mathbf{D}$ iff, writing S_{E} and B as above, there exists an n-congruence on B^* which is compatible with S_{E}.*

Before moving on to the full proof of our main theorem (Theorem 1.1), we conclude the section by noticing that the Straubing delay theorem implies the easy direction of our result, namely, if L is in $\mathbf{ZG} * \mathbf{D}$ then L is in $\mathbf{LZG}$. This easy direction follows directly from [28], but we provide a self-contained argument in Appendix C for completeness.

In the rest of this paper, we show the much harder direction, i.e., if L is in $\mathbf{LZG}$ then L is in $\mathbf{ZG} * \mathbf{D}$. To prove this, using Corollary 5.5, it suffices to show:

▶ **Claim 5.6.** *Let S be a semigroup of $\mathbf{LZG}$, write S_{E} its idempotent category and $B := \mathrm{Arrows}(S_{\mathrm{E}})$. There is $n > 0$ such that the n-congruence on B^* is compatible with S_{E}.*

This result then implies, by our rephrasing of Straubing's result (Corollary 5.5), that S is in $\mathbf{ZG} * \mathbf{D}$. So in the rest of this paper we prove Claim 5.6. The proof is structured in three sections. First, in Section 6, we carefully choose the value n in the congruence to be "large enough". Second, in Section 7, we show auxiliary results about paths in the category of idempotents. Third, in Section 8, we conclude the proof, first by an induction on the number of rare arrow occurrences, then by a decomposition of the category using ear decompositions of multigraphs.

6 Choosing the Congruence

In this section, we define our choice of the value of n to prove Claim 5.6. Intuitively, we need to choose n to be large enough so that the "gap" between the number of occurrences of rare letters and of frequent letters can be made sufficiently large:

▶ **Definition 6.1.** *For Σ an alphabet, $u \in \Sigma^*$, $n \in \mathbb{N}$, and $m > 0$, we say that n is an* m-distant rare-frequent threshold *if, letting $\Sigma_r := \{a \in \Sigma \mid |u|_a \leq n\}$ be the rare letters for n, then their* total *number of occurrences in u plus 1 multiplied by m is less than n, formally $\left(1 + \sum_{a\in\Sigma_r} |u|_a\right) \times m \leq n$.*

In other words, as every frequent letter occurs strictly more than n times, this guarantees that every frequent letter occurs strictly more than $(r + 1)m$ times, where r is the total number of rare letters. This means that there must some contiguous subword containing no rare letter where the frequent letter occurs $> m$ times. This will prove useful in pumping arguments. Specifically, we will want an m-distant rare-frequent threshold with $m := |S|$.

Of course, we cannot pick an n which will be an m-distant rare-frequent threshold for any path u: there will always be paths u where the number of arrow occurrences is close to n. However, remember that n-equivalence implies n'-equivalence for all n' that divide n (Claim 3.2). This suggests that, by choosing a large and composite enough n, we can ensure that, given any pair of paths, we can pick a divisor n' which is a m-distant rare-frequent threshold. We will also want to ensure in the sequel that n' is also always a multiple of the idempotent power ω and of $|S| + 1$. Let us formally state that such a choice of n exists:

▶ **Lemma 6.2.** *For any $m \geq 1$, for any semigroup S, letting S_{E} be the category of idempotents of S, there exists an integer $n \geq 2$ with the following property: for any paths u_1, u_2 in S_{E}, there exists a divisor n' of n which is a multiple of $\omega \times (|S| + 1)$ (for ω the idempotent power of S) such that n' is an m-distant rare-frequent threshold for u_1 and for u_2.*

This value of n will be the one we choose in our proof of Claim 5.6. Note that the result applies to arbitrary semigroups, not only those of $\mathbf{ZG}$. Before we prove Lemma 6.2, we first observe for later that if a path has an m-distant rare-frequent threshold (actually it suffices to have a 1-distant rare-frequent threshold) then the frequent arrows in this path for this threshold must form a so-called *union of strongly connected components (SCCs)*:

▶ **Definition 6.3.** *Given S_{E} and a subset B' of its arrows B, we say that B'* is a union of SCCs *if, letting G be the directed graph on the objects of S_{E} formed of the arrows of B', then all connected components of G are strongly connected.*

▶ **Claim 6.4.** *Fix S and S_{E} and B, let $w \in B^*$ be a path of S_{E}, and let $n' > 1$ be a 1-distant rare-frequent threshold of w. Then the set of frequent arrows of w for n' is a union of SCCs.*

Proof sketch. Any frequent arrow occurs $> n'$ times in w, so w must contain $\geq n'$ occurrences of a return path. Now, by 1-distance, not all of these paths can contain a rare arrow. ◀

Hence, in the rest of this section, we prove Lemma 6.2. Let us show an abstract result that will give our choice of value n:

▶ **Claim 6.5.** *For any $d > 0$ and $k > 0$ and $m \geq 1$, there exists $n \geq m$ such that for any d-tuple T of integers, there exists an $n'' \geq m$ such that, letting $n' := n''k$, we have that n' divides n and that $m \sum_{i \in F} T_i \leq n'$, where $F = \{i \mid T_i \leq n'\}$.*

Intuitively, d is the cardinality of the alphabet B, k is $\omega \times (|S| + 1)$ (ensuring that we always work with multiples of that value), m enforces a sufficiently large gap (it is the parameter of a distant rare-frequent threshold), and n is the threshold that we will choose, to ensure the existence of a suitable threshold n'. Let us sketch the proof of Claim 6.5:

Proof sketch. By choosing a sufficiently large n, we can ensure that we have $2^d + 1$ divisors of n that are candidate thresholds (are multiple of k) and are sufficiently far apart. As there are only 2^d possible partitions in rare and frequent alphabets, the pigeonhole principle ensures that two divisors achieve the same partition. Taking the larger one then ensures that the gap between rare and frequent letter occurrences is sufficiently large. ◀

With Claim 6.5, it is now easy to show Lemma 6.2; details are given in the appendix.

7 The Loop Insertion and Prefix Substitution Lemmas

We now show several auxiliary results on the category of idempotents to be used in the sequel. We first show some combinatorial results on paths and loops. We then use them to establish two technical claims on paths: the *loop insertion lemma*, making it possible to insert any loop of frequent arrows to the power n' (with n' a sufficiently distant rare-frequent threshold) without affecting equivalence; and the *prefix substitution lemma*, which we can use to replace a prefix of frequent arrows by another up to inserting a loop later in the path.

Recall that S_{E} denotes the idempotent category of the semigroup S in **LZG** that we study, and B denotes the set of arrows of S_{E}.

Basic combinatorial results. We first apply the definition of **ZG** to the local monoid to get a trivial result about the commutation between loops:

▶ **Claim 7.1.** *Let x and y be two coterminal loops of S_{E}, let $k \in \mathbb{Z}$, and let ω be an idempotent power of S. We have: $x^{\omega+k}y \equiv yx^{\omega+k}$*

We then show that frequent loops can be "recombined" without changing the category image, simply by equation manipulation:

▶ **Claim 7.2.** *For x, x' two coterminal paths in S_{E} and y, y' coterminal paths in S_{E} such that xy and $x'y'$ are valid loops, we have: $(xy)^\omega(x'y')^\omega \equiv (xy')^\omega(x'y)^\omega(xy)^\omega(x'y')^\omega$.*

The previous lemma implies that we can freely change the initial part of a path, even if it not a loop, when there is a coterminal path under an ω with which we can swap it. We show this again by equation manipulation, and it will be crucial for the prefix substitution lemma that we show later in the section:

▶ **Claim 7.3.** *For x, x' two coterminal paths in S_{E} and y, y' coterminal paths in S_{E} such that xy and $x'y'$ are valid loops, and for any path t coterminal with y, the following equation holds: $xt(xy)^\omega(x'y')^\omega \equiv x't(xy)^\omega xy'(x'y')^{\omega-1}$.*

Loop insertion lemma. We now argue that, when we have a sufficiently distant rare-frequent threshold n', we can insert any arbitrary loop raised to the power n' without changing the category element to which a path evaluates:

▶ **Lemma 7.4** (Loop insertion lemma). *Let p be a path, and assume that n' is a rare-frequent threshold for p which is $|S|$-distant and a multiple of ω. Let $p = rt$ be a decomposition of p (with r or t potentially empty), let o be the object between r and t (i.e., the final object of r, or the initial object of t if r is empty), and let p' be a loop over o that only uses frequent arrows. Then $p \equiv r(p')^{n'} t$ (note that they are also n'-equivalent by construction).*

We only sketch the proof of this result, which is proved in the appendix.

Proof sketch. Intuitively, we show that, at any object o along a path, any idempotent x that can be achieved as a loop of frequent arrows of the form $q_x^{n'}$ on o can be added without affecting equivalence. This clearly preserves n'-equivalence by definition, so we must only argue that it only preserves equivalence. This claim implies Lemma 7.4, as spawning the loop $(p')^{n'}$ when indicated is then absorbed by one of these idempotents.

To establish the claim, we first show that, for any prefix u of frequent arrows, we can spawn a loop starting by u with some arbitrary return path. This is by induction on the length of u. The base case of a prefix a of length 1 is shown by the pigeonhole principle: we consider all occurrences of a, and as it is frequent and the threshold is $|S|$-distant we can apply the pigeonhole principle to two occurrences of a separated only by frequent arrows, which we then iterate to form the loop. The induction step is shown by spawning a loop with a shorter prefix, then spawning the missing arrow of the prefix within the first loop, and recombining. Thanks to this, all necessary loops can be spawned, proving the claim. ◀

Prefix substitution lemma. We finally show that we can freely change any prefix of frequent arrows of a path, up to inserting a loop of frequent arrows elsewhere:

▶ **Lemma 7.5.** *Let $p = xry$ be a path, and assume that n' is a rare-frequent threshold for p which is $|S|$-distant and a multiple of ω. Let x' be a path coterminal with x. Assume that every arrow in x and in x' is frequent. Assume that some object o in the SCC of frequent arrows of the initial object of r occurs again in y, say as the intermediate object of $y = y_1 y_2$. Then there exists $y' = y_1 y'' y_2$ for some loop y'' consisting only of frequent arrows such that $p \equiv x'ry'$ and such that $p \sim_{n'} x'ry'$.*

This claim is shown in the appendix: it uses Claim 6.4 to argue that frequent arrows are a union of SCCs, and crucially relies on Claim 7.3.

8 Concluding the Proof of $\mathbf{LZG} \subseteq \mathbf{ZG} * \mathbf{D}$: Claim 5.6

We are now ready to prove the second direction of Theorem 1.1, namely Claim 5.6. We have fixed the semigroup S in $\mathbf{LZG}$, its category of idempotents S_{E}, and $B := \mathrm{Arrows}(S_{\mathrm{E}})$. We take the n given by Lemma 6.2. Our goal is to show that the n-congruence on B^* is compatible with S_{E}. To do so, let u_1 and u_2 be two coterminal paths that are n-equivalent. We must show that $u_1 \equiv u_2$, i.e., the two paths u_1 and u_2 evaluate to the same category element in S_{E}.

To do so, we use the guarantee on n ensured by Lemma 6.2 to pick an n' with which to work. Considering the path u_1, the lemma ensures that there is a divisor n' of n which is a multiple of $\omega \times (|S| + 1)$ and is an $|S|$-distant rare-frequent threshold for u_1. As n' divides n and u_1 and u_2 are n-equivalent, by Claim 3.2 we know that they are also n'-equivalent. So we will only consider the n'-congruence, denoted $\sim$, from now on. We know that u_1 and u_2 are n'-equivalent, that n' is a multiple of $\omega \times (|S|+1)$, and that n' is an $|S|$-distant rare-frequent threshold for u_1 and for u_2. Recall that, following Definition 3.1, now that we have fixed the threshold n', we call an arrow of B *rare* in u_1 in u_2 if it occurs $\leq n'$ times, and *frequent* otherwise.

We will show that $u_1 \equiv u_2$, and in fact will show that $p_1 \equiv p_2$ for pairs of paths p_1, p_2 more generally. More specifically, we establish the following claim by finite induction on r: let p_1 and p_2 be two coterminal paths that are n'-equivalent and which contain r rare arrows each. Then $p_1 \equiv p_2$. Showing this for all r establishes in particular that $u_1 \equiv u_2$.

Base case: all arrows in p_1 and p_2 are frequent. The base case of the induction is:

▶ **Claim 8.1.** *Let p_1 and p_2 be two coterminal paths that are n'-equivalent and which contain no rare arrows. Then $p_1 \equiv p_2$*

The claim is shown in Appendix F.1, so we only sketch the proof here. We consider the multigraph G of all arrows occurring in p_1 and p_2 (note that these arrows for p_1 and p_2 must be the same). We prove that $p_1 \equiv p_2$ by another induction, this time on the number of frequent arrows, i.e., the number of edges of the multigraph G. Formally, we show the following by finite induction on the integer η: let q_1 and q_2 be two coterminal paths that are n'-equivalent, where all rare arrows have 0 occurrences, and where there are $\leq \eta$ different frequent arrows. Then $q_1 \equiv q_2$. Showing this for all η establishes in particular that $p_1 \equiv p_2$.

The base case is $\eta = 0$, in which case q_1 and q_2 must be empty and the claim is trivial, so what matters is the induction step on G. Assume that the claim is true for any q_1 and q_2 such that G has $\leq \eta$ edges. Consider q_1 and q_2 such that G has $\eta + 1$ edges. Recall that G is strongly connected: indeed, we know, as all arrows are rich, that G is a union of SCCs, and p_1 (or p_2) witnesses that G is connected, so we know that G is strongly connected. The induction case is shown using a decomposition result on strongly connected multigraphs following the notion of *ear decomposition*:

▶ **Lemma 8.2.** *Let G be a strongly connected nonempty directed multigraph. We have:*

- *G is a simple cycle; or*
- *G contains a simple cycle $u_1 \to \cdots \to u_n \to u_1$ with $n \geq 1$, where all vertices $u_1, \ldots, u_n$ are pairwise disjoint, such that all intermediate vertices $u_2, \ldots, u_{n-1}$ only occur in the edges of the cycle, and such that the removal of the cycle leaves the graph strongly connected (note that the case $n = 1$ corresponds to the removal of a self-loop); or*
- *G contains a simple path $u_1 \to \cdots \to u_n$ with $n \geq 2$ where all vertices are pairwise distinct, such that all intermediate vertices $u_2, \ldots, u_{n-1}$ only occur in the edges of the path, and such that the removal of the path leaves the graph strongly connected (note that the case $n = 2$ corresponds to the removal of a single edge).*

This is a known result [7], but we give a self-contained proof in Appendix F.1. We use it to distinguish three cases in the induction step of the induction on G, which we now sketch.

The first case is when G is a simple cycle. In this case n'-equivalence ensures that the cycle is taken by q_1 and q_2 some number of times with the same remainder modulo n', so they evaluate to the same element because n' is a multiple of ω.

The second case is when G contains a simple cycle only connected to a single object. This time, we argue as in the previous case that the number of occurrences of the cycle must have the same remainder, and we can use Corollary 3.7 to merge all the occurrences together. However, to eliminate them, we need to use Lemma 7.5, to modify q_1 and q_2 to have the same prefix (up to and including the cycle occurrences), while preserving equivalence. This allows us to consider the rest of the paths (which contains no occurrence of the cycle), apply the induction hypothesis to them, and conclude by compositionality. A technicality is that we must ensure that removing the common prefix does not make some arrows insufficiently frequent relative to the distant rare-frequent threshold. We avoid this using Lemma 7.4 to spawn sufficiently many copies of a suitable loop.

The third case is when G contains a simple path π connecting two objects. The reasoning is similar, but we also use Lemma 7.4 to spawn a loop involving a return path for π and a path that is parallel to π (i.e., does not share any arrows with it). The return path in this loop can then be combined with π to form a loop, which we handle like in the previous case.

Induction case: some arrows are rare Let us now show the induction step for the outer induction, namely, the one on the number of occurrences of rare arrows. We assume the claim of the outer induction for $r \in \mathbb{N}$. Consider two paths p_1 and p_2 that are n'-equivalent and that contain $r+1$ rare letters. Let us partition them as $p_1 = q_1 a s_1$ and $p_2 = q_2 a s_2$ where q_1 and q_2 all consist of frequent arrows, and a is the first rare arrow of p_1 and p_2 (note that n'-equivalence implies that the first rare arrow is the same in both paths). In this case, q_1 and q_2 are two coterminal paths consisting only of frequent arrows (or they are empty), and s_1 and s_2 are two coterminal paths (possibly empty) with r rare letter occurrences.

The full proof is given in appendix; we only sketch it. By Claim 6.4, either the SCC at the origin of a occurs again in the rest of the path, or it does not. In the first case, we argue with Lemma 7.5 that the prefix q_1 can be substituted for q_2 without affecting equivalence, reason on the rest of the path by induction hypothesis, and conclude by compositionality. In the second case, n'-equivalence intuitively ensures that q_1 and q_2, and s_1 and s_2, must each be n'-equivalent, so we can apply the induction hypothesis to each of them. This concludes the induction step of the outer induction.

Concluding the proof. We have established by induction that p_1 and p_2 evaluate to the same category element in all cases. This implies that n-equivalence for our choice of n is compatible with S_{E}, so by Corollary 5.5 we know that L is in $\mathbf{ZG} * \mathbf{D}$. Thus, $L \in \mathbf{LZG}$ implies that $L \in \mathbf{ZG} * \mathbf{D}$. We have therefore established the locality result $\mathbf{LZG} = \mathbf{ZG} * \mathbf{D}$, concluding the proof of Claim 5.6 and hence of Theorem 1.1.

9 Conclusion

In this paper, we have given a characterization of the languages of $\mathbf{ZG}$, and proved that the variety $\mathbf{ZG}$ is local. The methodology seems to be adaptable enough to tackle $\mathbf{ZG} \cap \mathbf{A} = \mathbf{MNil}$ as well, but this would require a careful analysis of the proofs that we devote to future work.

The case of $\mathbf{ZE}$ is more complicated. As proved by Almeida [1], we have $\mathbf{ZE} = \mathbf{G} \vee \mathbf{Com}$, that is, $\mathbf{ZE}$ is the variety of monoids generated by both groups and commutative languages. Now, commutative languages are a specific example of a non-local variety, while $\mathbf{G}$ is a local variety. This being said, we do not know of general results showing the preservation or non-preservation of locality under such operators. Interestingly, however, one can check by computation that the counter-example language in $\mathbf{LCom}$ but not in $\mathbf{Com} * \mathbf{D}$ (illustrating that $\mathbf{Com}$ is not local), namely $e^*af^*be^*cf^*$, is in $\mathbf{LZG}$.

We hope that extending our approach to a study of locality for centrally defined varieties in general could lead to such general results on the interplay of join operations and of the locality or non-locality for arbitrary varieties.

References

1 Jorge Almeida. *Finite semigroups and universal algebra*, volume 3. World Scientific, 1994.

2 Jorge Almeida. A syntactical proof of locality of DA. *Int. J. Algebra Comput.*, 6(2):165–178, 1996. doi:10.1142/S021819679600009X.

3 Antoine Amarilli, Louis Jachiet, and Charles Paperman. Dynamic membership for regular languages. Available online: https://a3nm.net/publications/amarilli2021dynamic.pdf. Also submitted to ICALP'21, 2021.

4 K. Auinger. Join decompositions of pseudovarieties involving semigroups with commuting idempotents. *Journal of Pure and Applied Algebra*, 170(2):115–129, 2002.

5 Karl Auinger. Semigroups with central idempotents. In *Algorithmic problems in groups and semigroups*, pages 25–33. Springer, 2000.

6 Karl Auinger. On the decidability of membership in the global of a monoid pseudovariety. *Int. J. Algebra Comput.*, 20(2):181–188, 2010. doi:10.1142/S0218196710005571.

7 Jørgen Bang-Jensen and Gregory Z Gutin. *Digraphs: theory, algorithms and applications*. Springer Science & Business Media, 2008.

8 J.A. Brzozowski and Imre Simon. Characterizations of locally testable events. *Discrete Mathematics*, 4(3):243–271, 1973. doi:10.1016/s0012-365x(73)80005-6.

9 Luc Dartois and Charles Paperman. Alternation hierarchies of first order logic with regular predicates. In *FCT*, 2015.

10 Samuel Eilenberg. *Automata, languages, and machines. Vol. B.* Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1976. With two chapters ("Depth decomposition theorem" and "Complexity of semigroups and morphisms") by Bret Tilson, Pure and Applied Mathematics, Vol. 59.

11 Nathan Grosshans, Pierre McKenzie, and Luc Segoufin. The power of programs over monoids in DA. In *MFCS*, 2017.

12 Robert Knast. A semigroup characterization of dot-depth one languages. *RAIRO Theor. Informatics Appl.*, 17(4):321–330, 1983. doi:10.1051/ita/1983170403211.

13 Robert Knast. Some theorems on graph congruences. *RAIRO Theor. Informatics Appl.*, 17(4):331–342, 1983. doi:10.1051/ita/1983170403311.

14 Manfred Kufleitner and Alexander Lauser. Quantifier alternation in two-variable first-order logic with successor is decidable. In *STACS*, 2013.

15 Robert McNaughton. Algebraic decision procedures for local testability. *Mathematical Systems Theory*, 8(1):60–76, 1974. doi:10.1007/bf01761708.

16 J.-E. Pin. *Varieties of formal languages*. Foundations of Computer Science. Plenum Publishing Corp., New York, 1986. With a preface by M.-P. Schützenberger, Translated from the French by A. Howie. doi:10.1007/978-1-4613-2215-3.

17 Thomas Place and Luc Segoufin. Decidable characterization of FO2(<, +1) and locality of DA. abs/1606.03217, 2016.

18 Gudmund Skovbjerg Frandsen, Peter Bro Miltersen, and Sven Skyum. Dynamic word problems. *JACM*, 44(2):257–271, 1997.

19 Benjamin Steinberg. A modern approach to some results of Stiffler. In *Semigroups and Languages*. World Scientific, 2004. doi:10.1142/9789812702616_0013.

20 Price Stiffler. Chapter 1. extension of the fundamental theorem of finite semigroups. *Advances in Mathematics*, 11(2):159–209, 1973. doi:10.1016/0001-8708(73)90007-8.

21 Howard Straubing. The variety generated by finite nilpotent monoids. *Semigroup Forum*, 24(1):25–38, 1982. doi:10.1007/bf02572753.

22 Howard Straubing. Finite semigroup varieties of the form V*D. *Journal of Pure and Applied Algebra*, 36:53–94, 1985.

23 Howard Straubing. *Finite automata, formal logic, and circuit complexity*. Birkhäuser Boston Inc., Boston, MA, 1994.

24 Howard Straubing. A new proof of the locality of R. *International Journal of Algebra and Computation*, 25(01n02):293–300, 2015. doi:10.1142/s0218196715400111.

25 Denis Thérien and Alex Weiss. Graph congruences and wreath products. *J. Pure Appl. Algebra*, 36(2):205–215, 1985. doi:10.1016/0022-4049(85)90071-4.

26 Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC*, 1998.

27 Denis Thérien and Thomas Wilke. Temporal logic and semidirect products: An effective characterization of the until hierarchy. *SIAM J. Comput.*, 31(3):777–798, 2001. doi:10.1137/S0097539797322772.

28 Bret Tilson. Categories as algebra: an essential ingredient in the theory of monoids. *J. Pure Appl. Algebra*, 48(1-2):83–198, 1987. doi:10.1016/0022-4049(87)90108-3.

29 Jean Éric Pin. Mathematical foundations of automata theory. https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf, 2019.

A Proofs for Section 2 (Preliminaries)

▶ **Claim 2.1.** *For any monoid M in* **ZG**, *for $x, y \in M$, and $k \in \mathbb{Z}$, we have: $x^{\omega+k}y = yx^{\omega+k}$.*

Proof. This is simply because we can always write $x^{\omega+k}$ as $(x^{\omega+k})^{\omega+1}$, because the latter is equal to $x^{(\omega+k)\times(\omega+1)}$ which is indeed equal to $x^{\omega+k}$. Thus, by setting $x' := (x^{\omega+k})$ and applying the equation, we conclude. ◀

B Proofs for Section 3 (Characterizations of ZG)

B.1 Miscellaneous Results

▶ **Claim 3.2.** *For any alphabet Σ, for any $n > 0$, for any $m > 0$, if m is a multiple of n then the m-congruence refines the n-congruence.*

Proof. The claim is trivial for $m = n$, so we assume $m > n$. As $m > n$, if two words u and v have the same rare alphabet for m, then they have the same rare alphabet for n, because the number of occurrences of all rare letters for m is the same, so the same ones are also rare for n. Furthermore, if they have the same rare subword for m, the restriction of this same rare subword to the rare letters for n yields the same word. Last, we show that the number of occurrences modulo n are the same. For the letters that were frequent for m, this is the case because their number of occurrences is congruent modulo m, hence modulo n because n divides m. For the letters that were not frequent for m, this is because their number of occurrences has to be the same because the rare subwords for m were the same. ◀

▶ **Claim 3.3.** *For any alphabet Σ and $n > 0$, the n-congruence over Σ^* is a* **ZG***-congruence.*

Proof. Let E be an equivalence class of the n-congruence, which we see as a language of Σ^*, and let us show that E is a language of **ZG**. Let $\Sigma = A \sqcup B$ the partition of Σ in rare and frequent letters for the class E, let u be the word over A^* associated to the class E, and let $\vec{k}$ be the $|B|$-tuple describing the modulo values for E. We know that the singleton language $\{u\}$ is a language of **ZG**, because it is finite. Hence, the language $U = B^*u_1 \cdots B^*u_nB^*$ is also in **ZG**, because it is the inverse inverse of $\{u\}$ by the morphism that erases the letters of B and is the identity on A. Similarly, the language C of words of B^* where the modulo values of each letter are as prescribed by $\vec{k}$ and where every letter occurs at least n times is a language of **ZG**, because it is commutative. For the same reason, the language C' of words of Σ^* whose restriction to B are in C is also a language of **ZG**, because it is the inverse image of C by the morphism that erases the letters of A and is the identity on B. Now, we remark that $E = C' \cap U$, so E is in **ZG**, concluding the proof. ◀

B.2 Proofs of the Characterizations (Consequences of Theorem 3.4)

▶ **Corollary 3.5.** *Any* **ZG** *language L can be expressed as a finite union of languages of the form $B^*a_1B^*a_2 \cdots a_kB^* \cap K$ where $\{a_1, \ldots, a_k\} \cap B = \emptyset$ and K is a regular commutative language.*

Proof. Fix a language L in **ZG**, and consider the syntactic congruence $\sim$ of L: it is a **ZG**-congruence. By Theorem 3.4, there exists $n \in \mathbb{N}$ such that $\sim$ is refined by a n-congruence $\sim'$. Now, by definition of the syntactic congruence, the set of words of Σ^* that are in L is a union of equivalence classes of $\sim$, hence of $\sim'$. This means that L can be expressed as the union of the languages corresponding to these classes.

Now, an equivalence class of the n-congruence $\sim'$ can be expressed as the shuffle of two languages: the singleton language containing the rare word defining the class, and the language that imposes that all frequent letters are indeed frequent (so the rare alphabet is as required) and that the modulo of their number of occurrences is as specified. The second language is commutative, and the disjointness of rare and frequent letters guarantees that the shuffle is indeed disjoint.

Thus, we have shown that L is a union of disjoint shuffles of a singleton language and a regular commutative language. The form stated in the corollary is equivalent, i.e., it it the shuffle of the singleton language $\{a_1 \cdots a_k\}$ and of the commutative language obtained by restricting K to the subalphabet B. ◀

▶ **Corollary 3.7.** *For any monoid M in* **ZG**, *letting $n \geq (|M| + 1) \cdot \omega$, for any element m of M and elements $m_1, \ldots, m_n$ of M, we have*

$$m \cdot m_1 \cdot m \cdot m_2 \cdot m \cdots m \cdot m_n \cdot m \cdot m_n \cdot m = m^{n+1} \cdot m_1 \cdots m_n.$$

Proof. We consider the free monoid M^*. Let $\eta : M^* \to M$ be the onto morphism defined by $\eta(m_1 \cdot m_2) = \eta(m_1) \cdot \eta(m_2)$. Let $\sim$ be the congruence it induces over M^*, i.e., for $u, v \in M^*$, we have $u \sim v$ if $\eta(u) = \eta(v)$. Remark that, as M is in **ZG**, the congruence $\sim$ is a **ZG**-congruence by definition. Hence, by Theorem 3.4, $\sim$ is refined by a n-congruence where $n = (|M| + 1)\omega$. Now, consider the two words in the equation to be shown above: they are words of M^*. As the letter m over is then a frequent letter, we know that the two words are indeed n-congruent, which concludes. ◀

B.3 Proving Lemma 3.8

We show the lemma by establishing two claims:

▶ **Claim B.1.** *We have: $(xy)^\omega = x^\omega y^\omega (xy)^\omega$*

Proof. We have $(xy)^\omega = xy(xy)^{\omega-1}$: as $(xy)^{\omega-1}$ is central the right-hand-side is equal to: $x(xy)^{\omega-1}y$. By injecting an $(xy)^\omega$ in the latter, we obtain:

$$(xy)^\omega = x(xy)^{\omega-1}(xy)^\omega y.$$

Applying this equality ω times gives:

$$(xy)^\omega = (x(xy)^{\omega-1})^\omega (xy)^\omega y^\omega.$$

Now, note that we can expand $(x(xy)^{\omega-1})^\omega$, commuting the $(xy)^{\omega-1}$ to regroup the x into x^ω and regroup the $(xy)^{\omega-1}$ into $((xy)^{\omega-1})^\omega$ which is equal to $(xy)^\omega$, so that the first factor of the right-hand-side is equal to $x^\omega(xy)^\omega$

Thus, by commuting, we obtain: $(xy)^\omega = x^\omega y^\omega (xy)^\omega$, the desired result. ◀

▶ **Claim B.2.** *We have: $x^\omega y^\omega = x^\omega y^\omega (xy)^\omega$.*

Proof. We have $x^\omega y^\omega = x^{\omega-1} x y^{\omega-1} y$, so by the equation of **ZG** we get:

$$x^\omega y^\omega = x^{\omega-1} y^{\omega-1} xy.$$

Now, we have $x^{\omega-1} y^{\omega-1} = x^{\omega-1} x^\omega y^{\omega-1} y^\omega$, and by the equation of **ZG** we have:

$$x^{\omega-1} y^{\omega-1} = x^{\omega-1} y^{\omega-1} x^\omega y^\omega.$$

Inserting the second equality in the first, we have:

$$x^\omega y^\omega = x^{\omega-1} y^{\omega-1} x^\omega y^\omega xy.$$

Now, applying this equality ω times gives $x^\omega y^\omega = (x^{\omega-1}y^{\omega-1})^\omega x^\omega y^\omega (xy)^\omega$. As the first factor of the right-hand side is equal to $x^\omega y^\omega$, we get $x^\omega y^\omega = x^\omega y^\omega (xy)^\omega$, the desired result. ◀

Putting Claims B.1 and B.2 together immediately establishes Lemma 3.8.

B.4 Concluding the Proof of Theorem 3.4

To conclude the proof of Theorem 3.4, we will now show a kind of "normal form" for **ZG**-congruences, by arguing that any word can be rewritten to a word where frequent letters are moved to the end of the word, without breaking equivalence for the **ZG**-congruence. This relies on Lemma 3.8 and allows us to get to the notion of n-equivalence. Specifically:

▶ **Claim B.3.** *Let $\sim$ be a **ZG**-congruence on Σ. Let $n := (|M| + 1) \cdot \omega$ where M is the monoid associated to $\sim$. Then, for all $w \in \Sigma^*$, for every letter $a \in \Sigma$ which is frequent in w (i.e., $|w|_a > n$), writing w' the restriction of w to $\Sigma \setminus \{a\}$, and writing $w'' := w' a^{|w|_a}$, we have: $w \sim w''$.*

Proof. Define n as in the claim statement, and let $\mu : \Sigma^* \to M = \Sigma^*/{\sim}$ be the morphism associated to $\sim$. Remark that by definition, for any words u, v, we have $u \sim v$ iff $\mu(u) = \mu(v)$.

Let us take an arbitrary w and $a \in \Sigma$ such that a is frequent in w. We can therefore write $w = w_1 a w_2 a \cdots w_m a w_{m+1}$ with $m = |w|_a > n > |M|$. Furthermore, letting $x_l = \mu(v_1 \cdots v_l a)$ for each $1 \leq l \leq m$, as $m > M$ we know by the pigeonhole principle that there exist $1 \leq i < j \leq m$ such that $x_i = x_j$. But we have $x_j = x_i z \mu(a)$ where $z = \mu(v_{i+1} a \cdots v_j)$. By applying the equation ω times, we have that $x_i = x_i (z\mu(a))^\omega$.

Now, by Lemma 3.8, we have $(z\mu(a))^\omega = z^\omega \mu(a)^\omega$. This is equal to $z^\omega \mu(a)^\omega \mu(a)^\omega$, and by now applying the lemma in reverse we conclude that $(z\mu(a))^\omega = (z\mu(a))^\omega \mu(a)^\omega$. Finally, we obtain $x_i = x_i (z\mu(a))^\omega = x_i (z\mu(a))^\omega \mu(a)^\omega = x_i \mu(a)^\omega$.

Now, the equation of **ZG** ensures that $\mu(a)^\omega$ is central, so we can commute it in $\mu(w)$ and absorb all occurrences of $\mu(a)$ in $\mu(w)$, then move it at the end, while keeping the same μ-image. Formally, from $x_i = x_i \mu(a)^\omega$, we have

$$\mu(w) = \mu(w_1)\mu(a) \cdots \mu(w_i)\mu(a)\mu(a)^\omega \mu(w_{i+1})\mu(a) \cdots \mu(w_m)\mu(a)\mu(w_{m+1}),$$

and we commute $\mu(a)^\omega$ to merge it with all $\mu(a)$ and then commute the resulting $\mu(a)^{\omega + |w|_a}$.

As $|w|_a \geq \omega$, this μ-image is the same as the one that we obtain from $w' a^{|w|_a}$, with w' as defined in the statement of the claim. This establishes that $w \sim w''$ and concludes the proof. ◀

We can now conclude the proof of Theorem 3.4:

Proof of Theorem 3.4. Let $\sim$ be a **ZG**-congruence on Σ^*, M its associated monoid, and fix $n := (|M| + 1) \cdot \omega$ as in the theorem statement. Let u and v be two n-congruent words of Σ^*, we need to prove that they are indeed $\sim$-equivalent. Let $\Sigma' = \{a_1, \ldots, a_r\}$ be the subset of letters in Σ that are frequent in u (hence in v, as they are n-congruent). By successive applications of Claim B.3 for every frequent letter in Σ', starting with u, we know that $u \sim u_{\leq n} a_1^{|u|_{a_1}} \cdots a_r^{|u|_{a_r}}$. Likewise, we have $v \sim v_{\leq n} a_1^{|v|_{a_1}} \cdots a_r^{|v|_{a_r}}$. Now, we know that ω divides n. Thus, as for any $1 \leq i \leq r$, the values $|u|_{a_i}$ and $|v|_{a_i}$ are greater than n and congruent modulo n (by definition of the n-congruence), we have $a_i^{|u|_{a_i}} \sim a_i^{|v|_{a_i}}$. We also know by definition of the n-congruence that $u_{\leq n} = v_{\leq n}$. All of this together establishes that $u \sim v$. Thus, the n-congruence indeed refines the $\sim$-congruence, concluding the proof. ◀

C Proofs for Section 5 (Straubing's Delay Theorem)

In this appendix, we give the self-contained proof of the easy direction of our main result, namely:

▶ **Claim C.1.** *We have* $\mathbf{ZG} * \mathbf{D} \subseteq \mathbf{LZG}$.

Proof. If L is in $\mathbf{ZG} * \mathbf{D}$, then by Theorem 5.4, there exists a $\mathbf{ZG}$-congruence $\sim$ compatible with S_{E}. Let us now show that L is in $\mathbf{LZG}$ by showing that, writing S the syntactic semigroup of L, for any idempotent e, the local monoid eSe is in $\mathbf{ZG}$. Let e be an idempotent. By definition of S_{E}, the local monoid eSe is isomorphic to the subset of arrows of S_{E} going from e to e, with their composition law. Let us denote this subset by B_e. Define $\sim_e$ to be the specialization of the relation $\sim$ to B_e. Remark that $N = B_e^*/\sim_e$ is a submonoid of $B^*/\sim$, and is hence in $\mathbf{ZG}$ because $B^*/\sim$ is, and $\mathbf{ZG}$ is a variety. Remark that since all words in B_e^* are valid paths in S_{E}, the local monoid $H := eSe$ defines a congruence $\sim_2$ over B_e^* where two paths are equivalent if they evaluate to the same monoid element. We know that $\sim$, hence $\sim_e$, refines this congruence $\sim_2$. Hence, H is a quotient of N. Thus, eSe is a quotient of $B_e^*/\sim_e$, which is a submonoid of a monoid in $\mathbf{ZG}$, concluding the proof. ◀

D Proofs for Section 6 (Choosing the Congruence)

D.1 Proving Claim 6.4 and Claim 6.5

▶ **Claim 6.4.** *Fix S and S_{E} and B, let $w \in B^*$ be a path of S_{E}, and let $n' > 1$ be a 1-distant rare-frequent threshold of w. Then the set of frequent arrows of w for n' is a union of SCCs.*

Proof. Consider G the directed graph of Definition 6.3. Let us assume by way of contradiction that G has a connected component which is not strongly connected. This means that there exists an edge (u, v) of G such that there is no path from u to v in G. Consider any frequent arrow a in S_{E} achieving the edge (u, v) of G. As a is frequent is w, we know that a occurs $> n'$ times in w, hence w contains $\geq n'$ paths from the final object v of a back to the initial object u of a. As there is no path from v to u in G, each one of these paths must contain an arrow of B which is rare in w.

Hence, the total number of rare arrows in w is at least n'. But the 1-distant rare-frequent threshold condition imposes that the total number of rare arrow occurrences in w is $\leq n' - 1$. We have thus reached a contradiction. ◀

▶ **Claim 6.5.** *For any $d > 0$ and $k > 0$ and $m \geq 1$, there exists $n \geq m$ such that for any d-tuple T of integers, there exists an $n'' \geq m$ such that, letting $n' := n''k$, we have that n' divides n and that $m \sum_{i \in F} T_i \leq n'$, where $F = \{i \mid T_i \leq n'\}$.*

Proof. Let us take $n := k \times ((md)^{2^d+1}!)$, which ensures $n \geq m$. This ensures that n has as divisors $kmd, k(md)^2, \ldots, k(md)^{2^d+1}$: these are all our possible choices of values for n'. Now take any d-tuple T. For any possible choice of n'', the *rare set* is the subset of coordinates of T having values $\leq kn''$. By the pigeonhole principle, we can choose two of the divisors above, say $(md)^i$ and $(md)^j$ with $i < j$, having the same rare set, i.e., letting $F_i = \{i' \mid T_{i'} \leq (md)^i k\}$ and $F_j = \{i' \mid T_{i'} \leq (md)^j k\}$, we have $F_i = F_j$.

Let us set $n'' := (md)^j$, and let $n' := n''k$. By construction, we have $n' \geq m$, and n' divides n. Now, consider the sum $m \times \sum_{i' \in F_j} T_{i'}$. As $F_j = F_i$, we know that for every $i' \in F_j$, we have $T_{i'} \leq (md)^i k$. Thus, the sum is at most d times this value because T is a d-tuple, and multiplying by m we know that the sum $m \times \sum_{i' \in F_j} T_{i'}$ is at most $(md) \times (md)^i k$, hence it is $\leq (md)^{i+1} k$, so it is $\leq n'k$ for $n' := (md)^j$ because $i < j$. This concludes the proof. ◀

D.2 Concluding the proof of Lemma 6.2

We are now ready to show Lemma 6.2:

Proof of Lemma 6.2. Fix S, S_{E}, and B, and the desired m. Take n to be as given by Claim 6.5 with $d := 2\,|B|$, with $k := \omega \times (|S| + 1)$, with ω being the idempotent power of the semigroup, and with m being the desired m plus 1. Now consider any pair u_1, u_2 of S_{E}. Let T be the d-tuple of the letter occurrences of u_1, followed by those of u_2. The statement of Claim 6.5 ensures that there exists $n' > 1$ such that $n'k$ divides n and such that the total number of rare arrows in u_1 plus in u_2 is $\leq (n'k)/(|S| + 1)$. As $n' \geq |S| + 1$ and $k \geq |S| + 1$, we have $n'k \geq |S|\,(|S| + 1)$, so we have $(n'k)/(|S| + 1) \leq (n'k)/\,|S| - 1$. Hence, the total number of rare arrows u_1 plus in u_2 is $\leq (n'k)/\,|S| - 1$. So the same is true of the rare arrows in u_1, and of the rare arrows in u_2. By contrast, the frequent arrows in u_1 occur $> n'k$ times, and the same is true of the frequent arrows in u_2. Hence, by Definition 6.1, $n'k$ is an $|S|$-distant rare-frequent threshold for u_1 and for u_2. Now, note that $n'k$ is a multiple of k, hence of $\omega \times (|S| + 1)$. Thus, we have achieved all desired conditions and showed the result. ◀

E Proofs for Section 7 (The Loop Insertion and Prefix Substitution Lemmas)

E.1 Proof of Basic Combinatorial Results

▶ **Claim 7.1.** *Let x and y be two coterminal loops of S_{E}, let $k \in \mathbb{Z}$, and let ω be an idempotent power of S. We have: $x^{\omega+k} y \equiv y x^{\omega+k}$*

Proof. Recall that the equation of **ZG** implies: $x^{\omega+k} y \equiv y x^{\omega+k}$. By definition of **LZG**, the local monoid eSe, for e the initial and final object of x and y, is in **ZG**. As the loops x and y evaluate in the category to some arrow having e as starting and ending object, the previous equation then concludes the proof. ◀

▶ **Claim 7.2.** *For x, x' two coterminal paths in S_{E} and y, y' coterminal paths in S_{E} such that xy and $x'y'$ are valid loops, we have: $(xy)^\omega (x'y')^\omega \equiv (xy')^\omega (x'y)^\omega (xy)^\omega (x'y')^\omega$.*

Proof. Let us first show that:

$$(xy)^\omega (x'y')^\omega \equiv x'y(xy)^{\omega-1}(x'y')^{\omega-1}xy' \tag{1}$$

To show Equation 1, first rewrite $(xy)^\omega$ as $x(yx)^{\omega-1}y$ and likewise for $(y'x')^\omega$, to get:

$$(xy)^\omega (x'y')^\omega \equiv x(yx)^{\omega-1}yx'(y'x')^{\omega-1}y'$$

Then, we use Claim 7.1 to move $(yx)^{\omega-1}$, so the above is equal to:

$$xyx'(yx)^{\omega-1}(y'x')^{\omega-1}y'$$

We again rewrite $(yx)^{\omega-1}$ to $y(xy)^{\omega-2}x$, yielding:

$$xyx'y(xy)^{\omega-2}x(y'x')^{\omega-1}y'$$

We again use Claim 7.1 to move $(xy)^{\omega-2}$, merge it with the prefix xy, and move it back to its place, yielding:

$$x'y(xy)^{\omega-1}x(y'x')^{\omega-1}y'$$

We rewrite $(y'x')^{\omega-1}$ to $y'(x'y')^{\omega-2}x'$, yielding:

$$x'y(xy)^{\omega-1}xy'(x'y')^{\omega-2}x'y'$$

Again by Claim 7.1, we can merge $(x'y')^{\omega-1}$ with $x'y'$ and move it to finally get:

$$x'y(xy)^{\omega-1}(x'y')^{\omega-1}xy'$$

This establishes Equation 1.

Now, as $(xy)^{\omega}(x'y')^{\omega} \equiv (xy)^{2\omega}(x'y')^{2\omega}$, we can now apply Equation 1 ω times to the right-hand side and get:

$$(xy)^{\omega}(x'y')^{\omega} \equiv (x'y)^{\omega}(xy)^{\omega}(x'y')^{\omega}(xy')^{\omega} \tag{2}$$

As these elements commute (thanks to Claim 7.1), we have shown the desired equality. ◀

▶ **Claim 7.3.** *For x, x' two coterminal paths in S_{E} and y, y' coterminal paths in S_{E} such that xy and $x'y'$ are valid loops, and for any path t coterminal with y, the following equation holds: $xt(xy)^{\omega}(x'y')^{\omega} \equiv x't(xy)^{\omega}xy'(x'y')^{\omega-1}$.*

Proof. We apply Claim 7.2 to show the following equality on the left-hand-side:

$$xt(xy)^{\omega}(x'y')^{\omega} \equiv xt(xy')^{\omega}(x'y)^{\omega}(xy)^{\omega}(x'y')^{\omega}$$

By commutation of $(x'y)^{\omega}$ thanks to Claim 7.1, the right-hand-side is equal to:

$$(x'y)^{\omega}xt(xy')^{\omega}(x'y')^{\omega}(xy)^{\omega}$$

By expanding $(x'y)^{\omega} = x'(yx')^{\omega-1}y$, we get:

$$x'(yx')^{\omega-1}yxt(xy')^{\omega}(x'y')^{\omega}(xy)^{\omega}$$

By commutation of $(xy)^{\omega}$ and expanding it to $x(yx)^{\omega-1}y$, we get:

$$x'(yx')^{\omega-1}yx(yx)^{\omega-1}yxt(xy')^{\omega}(x'y')^{\omega}$$

Combining $(yx)^{\omega-1}$ with what precedes and follows, we get:

$$x'(yx')^{\omega-1}(yx)^{\omega+1}t(xy')^{\omega}(x'y')^{\omega}$$

By expanding $(x'y')^{\omega} = x'(y'x')^{\omega-1}y'$, and commuting $(yx')^{\omega-1}$ and $(yx)^{\omega+1}$, we get:

$$x't(xy')^{\omega}x'(yx')^{\omega-1}(yx)^{\omega+1}(y'x')^{\omega-1}y'$$

Now, we have $x'(yx')^{\omega-1} = (x'y)^{\omega-1}x'$, so we get:

$$x't(xy')^{\omega}(x'y)^{\omega-1}x'(yx)^{\omega+1}(y'x')^{\omega-1}y'$$

Commuting $(y'x')^{\omega-1}$ and doing a similar transformation, we get:

$$x't(xy')^{\omega}(x'y)^{\omega-1}(x'y')^{\omega-1}x'(yx)^{\omega+1}y'$$

Now, expanding $(yx)^{\omega+1}$, we get:

$$x't(xy')^{\omega}(x'y)^{\omega-1}(x'y')^{\omega-1}x'y(xy)^{\omega}xy'$$

Commuting $(x'y)^{\omega-1}$ and merging it with $x'y$, we finally get:

$$x't(xy')^{\omega}(x'y')^{\omega-1}(x'y)^{\omega}(xy)^{\omega}xy'$$

Note that $(x'y')^{\omega-1} \equiv (x'y')^{\omega}(x'y')^{\omega-1}$, so applying commutation we get:

$$x't(xy')^{\omega}(x'y')^{\omega}(x'y)^{\omega}(xy)^{\omega}(x'y')^{\omega-1}xy'$$

Now, applying Claim 7.2 in reverse (using commutation again, we can obtain):

$$x't(x'y')^{\omega}(xy)^{\omega}(x'y')^{\omega-1}xy'$$

And commuting $(x'y')^{\omega}$ and merging it yields:

$$x't(xy)^{\omega}(x'y')^{\omega-1}xy'$$

A final commutation of $(x'y')^{\omega-1}$ yields the desired right-hand-side, establishing the result. ◀

E.2 Proof of the Loop Insertion Lemma (Lemma 7.4)

▶ **Lemma 7.4** (Loop insertion lemma). *Let p be a path, and assume that n' is a rare-frequent threshold for p which is $|S|$-distant and a multiple of ω. Let $p = rt$ be a decomposition of p (with r or t potentially empty), let o be the object between r and t (i.e., the final object of r, or the initial object of t if r is empty), and let p' be a loop over o that only uses frequent arrows. Then $p \equiv r(p')^{n'}t$ (note that they are also n'-equivalent by construction).*

We first rephrase the claim to the following auxiliary result:

▶ **Claim E.1.** *Let p be a path, assume that n' is a rare-frequent threshold for p which is $|S|$-distant and a multiple of ω, let o be the object between r and t, and let X be the set of elements of the local monoid on o that can be achieved as a loop $q_x^{n'}$ on x with frequent arrows (i.e., the loop evaluates to an arrow (o, x, o)), noting that this implies that x is idempotent as n' is a multiple of ω. Then letting $q := \left(\prod_{x \in X} q_x^{n'}\right)$, we have that $p \equiv rqt$. (Note that they are also n'-equivalent by construction.)*

We explain why Claim E.1 implies the desired claim. Indeed, when taking $p = rt$ and taking $r(p')^{n'}t$, choosing for o the object between r and t, the sets X defined in the auxiliary result will be the same for both paths (because X only depends on o), so the rephrased claim implies that there is a loop q such that $rt \equiv rqt$ and $r(p')^{n'}t \equiv rq(p')^{n'}t$. Now, as $(p')^{n'}$ must correspond to an arrow of the form (o, x, o) for $x \in X$, it must be the same idempotent as one of the idempotents achieved by one of the loops in the definition of q, and as the local monoid is in **ZG** these idempotents commute and $q \equiv q(p')^{n'}$. Hence, we have $rqt \equiv rq(p')^{n'}t$. We know that $rqt \equiv rt$, and $rq(p')^{n'}t \equiv r(p')^{n'}t$. Thus we obtain $rt \equiv r(p')^{n'}t$. Thus, Lemma 7.4 is proved once we have shown Claim E.1.

Hence, all that remains is to show Claim E.1. We will do so by establishing a number of claims, all of which will have their proofs deferred to the appendix.

We first prove a preliminary claim which uses the pigeonhole principle to insert a loop containing an arbitrary arrow x in a word where x occurs more than $|S|$ times:

▶ **Claim E.2.** *Let $p = rt$ be a path, let o be the object between r and t, let x be an arrow starting at o, and assume that x occurs $> |S|$ times in p. Then we have $p \equiv r(xu)^\omega t$ for some return path u using only the arrows of p.*

Proof. As x occurs $k > |S|$ times in p, this provides a decomposition of p in the shape: $p = p_1 x p_2 x \cdots p_k x s$.

By the pigeonhole principle, there exists $i < j$ such that $p_1 x \cdots p_i x \equiv p_1 \cdots p_j x$. Hence, iterating, we obtain:

$$p_1 x \cdots p_j x \equiv p_1 x \cdots p_i x (p_{i+1} x \cdots p_j x)^\omega.$$

Moving the ω, we get:

$$p_1 x \cdots p_j x \equiv p_1 x \cdots p_{i-1} x p_i (x p_{i+1} x \cdots p_j)^\omega x.$$

This proves that p and $h(xu)^\omega g$ achieve the same category element by taking $u := p_{i+1} x \cdots p_j$, $h := p_1 x \cdots p_{i-1} x p_i$ and $g := p_{j+1} x \cdots p_k x s$. Remark that the terminal object of h is the same than the terminal object of r. Furthermore either h is a prefix of r or the converse. Assume first that $h = rw$ for some path w. Then, w and $(xu)^\omega$ belong to the local monoid of the terminal object of r which is in **ZG**. Since idempotents commute with all elements, we have $w(xu)^\omega \equiv (xu)^\omega w$ achieving that $rw(xu)^\omega g \equiv r(xu)^\omega wg \equiv r(xu)^\omega t$ since $wg = t$. The other case is symmetrical. This concludes the proof of Claim E.2. ◀

Let us extend this to a claim using the notion of distant rare-frequent threshold (this is where we use the fact that the threshold is distant):

▶ **Claim E.3.** *Let $p = rt$ be a path with an $|S|$-distant rare-frequent threshold n', let o be the object between r and t, and let x be any frequent arrow starting at o. Then we have $p \equiv r(xu)^\omega t$ for some return path u using only frequent arrows of p.*

Proof. As x is a frequent arrow and n' is $|S|$-distant, it occurs $> (\rho + 1)\,|S|$ times in p, where ρ is the total number of rare arrows of p. Hence, writing $p = p_1 a_1 \cdots p_\rho a_\rho p_{\rho+1}$ where the a_i are the rare arrows and the ρ_i are paths of frequent arrows, there must be a ρ_i containing $> |S|$ occurrences of x. Write $\rho_i = r' t'$ where the object between r and t is the initial object of x, which must exist as x occurs in ρ_i. Applying Claim E.2 to that decomposition, we have $\rho_i \equiv r'(xu)^\omega t'$ for some return path u using only arrows of ρ_i, hence only frequent arrows.

Now, as idempotents commute with all elements, similarly to the end of the proof of Claim E.2, we deduce that $p \equiv r(xu)^\omega t$. ◀

We then prove a generalization of the previous claim, going from a single frequent arrow to an arbitrary path of frequent arrows:

▶ **Claim E.4.** *Let $p = rt$ be a path with a rare-frequent threshold n' which is $|S|$-distant and a multiple of ω. Let o be the object between r and t, and let h be a path of frequent arrows starting at o. Then for any path h using only arrows of p beginning at the final object of r, we have $p \equiv r(hg)^\omega t$ for some return path g using only frequent arrows of p.*

Proof. We show the claim by induction on the length of h. The base case of the induction, with h of length 0, is trivial with g also having length 0.

For the inductive claim, write $h = h'a$. By induction hypothesis, there exists a g' using only frequent arrows of p such that:

$$p \equiv r(h'g')^\omega t.$$

Furthermore, by applying Claim E.3 to the decomposition $r' = rh'$ and $t' = g'(h'g')^{\omega-1}t$ and with the frequent arrow a we get a return path u using only frequent arrows of p such that:

$$r(h'g')^{\omega}t \equiv rh'(au)^{\omega}g'(h'g')^{\omega-1}t$$

So, iterating the ω power, and combining with the preceding equation, we get:

$$p \equiv rh'((au)^{\omega})^{\omega}g'(h'g')^{\omega-1}t.$$

Now, by applying $\omega - 1$ times Claim 7.1 to each $(au)^{\omega}$ except the first and to each loop going from after this $(au)^{\omega}$ to the position between an occurrence of h' and g', we get that:

$$p \equiv rh'(au)^{\omega}g'(h'(au)^{\omega}g')^{\omega-1}t.$$

Note the right-hand side is equal to: $r(h'(au)^{\omega}g')^{\omega}t$. So we have shown:

$$p \equiv r(h'(au)^{\omega}g')^{\omega}t.$$

So this establishes the inductive claim by taking $g := u(au)^{\omega-1}g'$. ◀

We can extend this to a claim about inserting arbitrary loops:

▶ **Claim E.5.** *Let $p = rt$ be a path with an $|S|$-distant rare-frequent threshold n', let o be the object between r and t, and let q and q' be loops on o using only frequent arrows. We have that $rqt \equiv rq(q')^{n'}(q'')^{n'}t$ for some loop q'' on o using only frequent arrows (note that the two are also n'-equivalent).*

Proof. We use Claim E.4 with $h := q'$. This gives us the existence of a return path g using only frequent arrows, which is then also a loop on o, such that $rqt \equiv rq(q'g)^{\omega}t$. Now, applying Lemma 3.8 to the local monoid on object o, we know that this evaluates to the same category element as: $rq(q')^{\omega}g^{\omega}t$. Hence, as n' is a multiple of ω, it evaluates to the same category element as $rq(q')^{n'}g^{n'}t$, which now preserves n'-equivalence and concludes the proof of Claim E.5. ◀

The only step left is to argue that Claim E.5 implies our rephrasing of the result that we wish to prove, Claim E.1. To do this, let o be the terminal object of r and let X be the set of idempotents definable from frequent arrows. For each idempotent $x \in X$, we can choose some loop q_x on frequent arrows that achieves it. Now, successive applications of Claim E.5 to each $x \in X$, and using the fact that the $(q')^{n'}$ and $(q'')^{n'}$ commute by Claim 7.1 (remember that n' is a multiple of ω), we know that $p = rt$ is n'-equivalent to, and evaluates to the same category element as, the path $r\left(\prod_{x\in X} q_x^{n'}(q_x')^{n'}\right)t$, for some q_x' for each q_x (corresponding to the q'' in that application of Claim E.5), which also only consists of frequent arrows. Now, since $(q_x')^{n'}$ is a loop on o consisting of frequent arrows, it also achieves an idempotent in the local monoid on o. As this monoid is in **ZG**, idempotents commute, and so these idempotents can all be combined with some idempotent $q_x^{n'}$ and absorbed by them. Thus, we get that rt is n'-equivalent to, and evaluates to the same category object as, the path $r\prod_{x\in X} q_x^{n'}t$. This concludes the proof of Claim E.1, and thus establishes our desired result, Lemma 7.4.

E.3 Proof of the Prefix Substitution Lemma (Lemma 7.5)

▶ **Lemma 7.5.** *Let $p = xry$ be a path, and assume that n' is a rare-frequent threshold for p which is $|S|$-distant and a multiple of ω. Let x' be a path coterminal with x. Assume that* *every arrow in x and in x' is frequent. Assume that some object o in the SCC of frequent arrows of the initial object of r occurs again in y, say as the intermediate object of $y = y_1 y_2$. Then there exists $y' = y_1 y'' y_2$ for some loop y'' consisting only of frequent arrows such that $p \equiv x'ry'$ and such that $p \sim_{n'} x'ry'$.*

Proof. As n' is an $|S|$-distant rare-frequent threshold, we know by Claim 6.4 that the frequent arrows occurring in p are a union of SCCs. Thus, there is a return path s for x' (i.e., $x's$ is a loop, hence xs also is) where s only consists of frequent arrows.

By our hypothesis on the initial object of r, we can decompose $y = y_1 y_2$ such that the terminal object of y_1 and initial object of y_2 is the initial object of r. Now, take p' to be the loop $s(xs)^\omega(x's)^\omega x$: note that all arrows of p' are frequent. Hence, by Lemma 7.4, xry evaluates to the same category element as, and is n'-equivalent to,

$$xry_1(p')^\omega y_2 = xry_1(s(xs)^\omega(x's)^\omega x)^{n'} y_2$$

By unfolding the power n', we get the following:

$$xry_1(p')^\omega y_2 = xry_1 s(xs)^\omega(x's)^\omega x(s(xs)^\omega(x's)^\omega x)^{n-1} y_2 = xry_1 s(xs)^\omega(x's)^\omega xz$$

where we write $z := (s(xs)^\omega(x's)^\omega x)^{n-1} y_2$ for convenience. We can therefore apply Claim 7.3 to obtain that:

$$(x(ry_1 s))(xs)^\omega(x's)^\omega xz \equiv (x'(ry_1 s))(xs)^\omega(xs)(x's)^{\omega-1} xz.$$

What is more, these two paths are clearly n'-equivalent, as they only differ in terms of frequent arrows (all arrows in x and x' being frequent) and the number of these arrows is unchanged by the transformation. This path is of the form given in the statement, taking $y' := y_1 s(xs)^\omega(xs)(x's)^{\omega-1} xz$ from which we can extract the right y''. This concludes the proof. ◀

F Proofs for Section 8 (Concluding the Proof of $\mathbf{LZG} \subseteq \mathbf{ZG} * \mathbf{D}$: Claim 5.6)

F.1 Proof of the Base Case: Claim 8.1

▶ **Claim 8.1.** *Let p_1 and p_2 be two coterminal paths that are n'-equivalent and which contain no rare arrows. Then $p_1 \equiv p_2$*

We first state and prove the lemma on graph decompositions:

▶ **Lemma 8.2.** *Let G be a strongly connected nonempty directed multigraph. We have:*

- *G is a simple cycle; or*
- *G contains a simple cycle $u_1 \to \cdots \to u_n \to u_1$ with $n \geq 1$, where all vertices $u_1, \ldots, u_n$ are pairwise disjoint, such that all intermediate vertices $u_2, \ldots, u_{n-1}$ only occur in the edges of the cycle, and such that the removal of the cycle leaves the graph strongly connected (note that the case $n = 1$ corresponds to the removal of a self-loop); or*
- *G contains a simple path $u_1 \to \cdots \to u_n$ with $n \geq 2$ where all vertices are pairwise distinct, such that all intermediate vertices $u_2, \ldots, u_{n-1}$ only occur in the edges of the path, and such that the removal of the path leaves the graph strongly connected (note that the case $n = 2$ corresponds to the removal of a single edge).*

We repeat here that the result is standard. The proof given below is only for the reader's convenience, and follows [7].

Proof. This result is showed using the notion of an ear decomposition of a directed multigraph. Specifically, following Theorem 7.2.2 of [7], for any nonempty strongly connected multigraph G, we can build a copy of it (called G') by the following sequence of steps, with the invariant that G' remains strongly connected:

- First, take some arbitrary simple cycle in G and copy it to G';
- Second, while there are some vertices of G that have not been copied to G', then pick some vertex v of G that was not copied, such that there is an edge (v', v) in G with v' a vertex that was copied. Now take some shortest path (hence a simple path) $v \to \cdots \to v''$ from v to the subset of the vertices of G that had been copied to G'. This path ends at a vertex v'' which may or may not be equal to v'. If $v'' \neq v'$, then we have a simple path $v' \to v \to \cdots \to v''$, which we copy to G'; otherwise we have a simple cycle, which we copy to G'. Note that, in both cases, all intermediate vertices in the simple path or simple cycle that we copy only occur in the edges of the path or cycle (as they had not been previously copied to G'). Further, G' clearly remains strongly connected after this addition.
- Third, once all vertices of G have been copied to G', take each edge of G that has not been copied to G' (including all self-loops), and copy it to G' (as a simple path of length 1). These additions preserve the strong connectedness of G'.

At the end of this process, G' is a copy of G.

Now, to show the result, take the graph G, consider how we can construct it according to the above process, and distinguish three cases:

- If the process stopped at the end of the first step, then G is a simple cycle (case 1 of the statement).
- If the process stopped after performing a copy in the second step, then considering the last simple path or simple cycle that we added, then it satisfies the conditions and its *removal* from G gives a graph which is still strongly connected (case 2 or case 3 of the statement).
- If the process stopped after performing a copy in the third step, then considering the last edge that we added, then it is a simple path of length 1 and its *removal* from G gives a graph which is still strongly connected (case 3 of the statement).

This concludes the proof. ◀

As explained in the body, we do an induction on the number of edges of the strongly connected multigraph G, whose base case is trivial. Here are the details of the three cases to consider in the induction step, following Lemma 8.2.

Case 1: G is a simple cycle. If G is a simple cycle, then distinguish the initial object of q_1 (hence, of q_2) as o_1, and let α be the category element corresponding to the cycle from o_1 to itself, and p the category element corresponding to the path from o_1 to the common terminal element of q_1 and q_2. We have: $q_1 = \alpha^{n_1} p$ and $q_2 = \alpha^{n_2} p$ with n_1 and n_2 being $\geq n'$ and having the same remainder modulo n'. By definition of x, there exists an idempotent e and some element $m \in eSe$ such that $x = (e, m, e)$. Hence, $q_1 = x^{n_1} p \equiv (e, m^{n_1}, e)p$ (resp. $q_2 = x^{n_2} p \equiv (e, m^{n_2}, e)p$). Since n' is a multiple of the idempotent power of S, we have $m^{n_1} \equiv m^{\omega + r}$ and $m^{n_2} \equiv m^{\omega + r}$ where r is the remainder modulo n'. Thus, we have $q_1 \equiv q_2$, concluding this case.

Case 2: G has a simple cycle. Recall that, in this case, we know that G has a simple cycle whose intermediate objects have no other incident edges and such that the removal of the simple cycle leaves the graph strongly connected. Let α be the simple cycle, starting from the only object e of the cycle having other incident edges. We can then decompose q_1 and q_2 to isolate the occurrences of the simple cycle (which must be taken in its entirety), i.e.:

$$q_1 = x_1\alpha x_2\alpha x_3 \cdots x_{t-1}\alpha x_t$$
$$q_2 = y_1\alpha y_2\alpha y_3 \cdots y_{t'-1}\alpha y_{t'}$$

We ensure that the edges of α do not occur elsewhere than in the α factors, except possibly in x_1, y_1 and in x_t, $y_{t'}$ if the paths q_1 and/or q_2 start/and or end in the simple cycle. However, in that case, we know that the prefixes of q_1 and q_2 containing this incomplete subset of the cycle must be equal (same sequence of arrows), and likewise for their suffixes. For this reason, it suffices to show the claim that q_1 and q_2 evaluate to the same category object under the assumption that both their initial and terminal objects are not intermediate vertices of the cycle: the claim then extends to the case when they can be (by adding the common prefixes and suffixes to the two paths that satisfy the condition, using the fact that a congruence is compatible with concatenation). Thus, in the rest of the proof for this case, we assume that the edges of α only occur in the α factors.

We will now argue that, to show that $q_1 \equiv q_2$, it suffices to show the same of two n'-equivalent coterminal paths from which all occurrences of the edges of the cycle have been removed and where all other edges still occur sufficiently many times. As this deals with paths where the underlying multigraph contains fewer edges, the induction hypothesis will conclude.

To do this, by Lemma 7.5, as x_1 and y_1 are coterminal and consist only of frequent arrows, and as the initial object of α occurs again in both paths, the path q_1 is n'-equivalent, and evaluates to the same category element as, some path:

$$q_1' = y_1\alpha x_2'\alpha x_3' \cdots x_{t''-1}'\alpha x_{t''}$$

For this reason, up to replacing q_1 by q_1', we can assume that $x_1 = y_1$.

Now, furthermore, $x_2, \ldots, x_{t-2}$ (resp. $y_2, \ldots, y_{t'-2}$) and α are coterminal cycles over the object e (which by definition corresponds to an idempotent of S). Hence, $\alpha x_2\alpha x_3 \cdots x_{t-1}\alpha = (e, mm_2mm_3 \cdots m_{t-1}m, e)$ where $\alpha = (e, m, e)$, $x_i = (e, m_i, e)$ for $2 \leq i \leq t-1$ and where m and all m_i's are in eSe, which is by hypothesis a monoid in **ZG**. Now, by Corollary 3.7, we know that $mm_2mm_3 \cdots m_{t-1}m = m^{t-1}m_2m_3 \cdots m_{t-1}$, because, as the arrows of α are frequent, the number of occurrences of α is $\geq n'$, and we have taken n' to be a multiple of $(|S|+1) \times \omega$ (where ω is the idempotent power of S), which is greater than $(|eSe|+1) \times k$, where k is the idempotent power of eSe (which divides ω, hence is $\leq \omega$). By applying the same reasoning to q_2, it suffices to show that the two following paths evaluate to the same category element, where $x_1 = y_1$:

$$x_1\alpha^{t-1}x_2x_3 \cdots x_{t-1}x_t$$
$$y_1\alpha^{t'-1}y_2y_3 \cdots y_{t'-1}y_{t'}$$

Now, because these two paths are n'-equivalent, we know that $t-1$ and $t'-1$ have the same remainder modulo n'. By the same reasoning as in case 1, they evaluate to the same category element as α^r, where r is the remainder. So it suffices to show that the two following paths evaluate to the same category element, with $x_1 = y_1$:

$$x_1\alpha^r x_2x_3 \cdots x_{t-1}x_t$$
$$y_1\alpha^r y_2y_3 \cdots y_{t'-1}y_{t'}$$

To ensure that the edges not in α still occur sufficiently many times, let β be any loop on e that visits all edges of G except the ones in α: this is doable because G is still strongly connected after the removal of α. Up to exponentiating β, we can assume that β traverses each edge sufficiently many times to satisfy the lower bound imposed by the requirement of n' being an $|S|$-distant rare-frequent threshold. By Lemma 7.4, it suffices to show that the following paths evaluate to the same category element:

$$q'_1 = x_1\alpha^r\beta^{n'}x_2x_3\cdots x_{t-1}x_t$$
$$q'_2 = y_1\alpha^r\beta^{n'}y_2y_3\cdots y_{t'-1}y_{t'}$$

Now, observe that both paths start by $x_1\alpha = y_1\alpha$, and the arrows of α do not occur in the rest of the paths. Now consider the paths $\beta^{n'}x_2x_3\ldots x_t$ and $\beta^{n'}y_2y_3\ldots y_{t'}$. They are paths that are coterminal, n'-equivalent because q_1 and q_2 were, where the frequent letters that are used are a strict subset of the ones used in p_1 and p_2, and where all other frequent letters occur sufficiently many times for n' to still be an $|S|$-distant rare-frequent threshold. Thus, by induction hypothesis, we know that these two paths evaluate to the same category element, so that q'_1 and q'_2 also do. This concludes case 2.

Case 3: G has a simple path. Recall that, in this case, we know that G has a simple path whose starting and ending objects have no other incident edges and such that the removal of the simple path leaves the graph strongly connected. We denote by $x \neq y$ the starting and ending objects of the path. Let π be the category element corresponding to the path. Since the removal of the path does not affect strong connectedness of the graph, there is a simple path from x to y sharing no edges with π; let κ be the category element to which this "return path" evaluates. Furthermore, there is a simple path ρ from y to x sharing no edges with π (this is because all intermediate objects of π only occur in the edges of π).

Like in the previous case, up to removing common prefixes and suffixes, it suffices to consider the case where q_1 and q_2 do not start or end in the intermediate vertices of π. For that reason, we can now isolate all occurrences of the edges of π, and write:

$$q_1 = x_1\pi x_2\pi x_3\cdots x_{t-1}\pi x_t$$
$$q_2 = y_1\pi y_2\pi y_3\cdots y_{t'-1}\pi y_{t'}$$

Like in the previous case, by Lemma 7.5, we can assume that $x_1 = y_1$.

By Lemma 7.4, we spawn a loop $(\rho\kappa)^{n'}$ after every occurrence of π without changing the category element and still respecting the n'-congruence. By expanding $(\rho\kappa)^{n'} = \rho\kappa(\rho\kappa)^{n'-1}$, it suffices to show that the following paths evaluate to the same category element, with $x_1 = y_1$:

$$q'_1 = x_1\pi\rho\kappa(\rho\kappa)^{n'-1}\cdots x_{t-1}\pi\rho\kappa(\rho\kappa)^{n'-1}x_t$$
$$q'_2 = y_1\pi\rho\kappa(\rho\kappa)^{n'-1}\cdots y_{t'-1}\pi\rho\kappa(\rho\kappa)^{n'-1}y_{t'}$$

We can now regroup the occurrences of $\pi\rho$, which are loops such that some edges (namely, the edges of π) only occur in these factors. This means that we can conclude as in case 2 for the cycle $\pi\rho$, as this cycle contains some edges that only occur there; we can choose β at the end of the proof to be a loop on x visiting all edges of G except those of π, which is again possible because G is still strongly connected even after the removal of π.

This establishes case 3 and concludes the induction step of the proof.

We have thus proved by induction that q_1 and q_2 evaluate to the same category element, in the base case of the outer induction where all edges of q_1 and q_2 are frequent.

F.2 Proof of the Induction Case

We now conclude the proof of the induction step for the outer induction. Recall from the body that we have taken two n'-equivalent paths $p_1 = q_1 a s_1$ and $p_2 = q_2 a s_2$ with a being the first of the $r + 1$ rare letters of the paths.

Remember that, as n' is an $|S|$-distant rare-frequent threshold for p_1 and p_2, then we know that the frequent arrows of p_1 form a union of SCCs (Claim 6.4); note that, thanks to n'-equivalence, the same is true of p_2 with the same SCCs. Let e be the source object of a, and consider the SCC C of frequent arrows that contains e. There are two cases, depending on whether some object of C occurs again in s_1 or not. Note that some object of C occurs again in s_1 iff the same is true of s_2, because which frequent arrow components occur again is entirely determined by the terminal objects of the rare arrows of s_1 and s_2, which are identical thanks to n'-equivalence.

Case 1: C occurs again after a. In this case, we apply Lemma 7.5, because q_1 and q_2 only consist of frequent arrows and some object of the SCC of the initial object of a occurs again in s_1. The claim tells us that there is a path:

$$p'_1 = q_2 a s'_1$$

which evaluates to the same category element as p_1 and is n'-equivalent to it. Hence, by compositionality, it suffices to show that s'_1 and s_2 evaluate to the same category element.

To apply the induction hypothesis, we simply need to ensure that n' is still a $|S|$-distant rare-frequent threshold for s'_1 and s_2. To do this, we need to ensure that the arrows that are frequent in p_1 and p_2 are still frequent there, and still satisfy the $|S|$-distant condition. Fortunately, we can simply ensure this by inserting a loop using Lemma 7.4. Formally, write $s'_1 = r_1 t_1$ where the intermediate object is the object of the SCC C that occurred in s_1 (the existence of such a decomposition is a consequence of the statement of Lemma 7.5), and write $s_2 = r_2 t_2$ in the same way (which we already discussed must be possible with s_2). Let p' be an arbitrary loop of frequent arrows where all arrows of C occur: this is possible because C is strongly connected. We know by Lemma 7.4 that $s'_1 = r_1 t_1$ and $r_1 (p)^{n'} t_1$ are both n'-equivalent and evaluate to the same category object: this is also true with $w_1 := r_1 (p)^{kn'} t_1$ for a sufficiently large k such that every frequent arrow of C occurs as many times as it did in p_1. Likewise, $s_2 = r_2 t_2$ and $w_2 := r_2 (p)^{kn'} t_2$ are both n'-equivalent and evaluate to the same category object, and we can take a sufficiently large k. So it suffices to consider w_1 and w_2.

Let us apply the induction hypothesis to them. They are two coterminal paths, and they are n'-equivalent because $w_1 \sim_{n'} p'_1 \sim_{n'} p_1$ and $w_2 \sim_{n'} p_2$ and by hypothesis $p_1 \sim_{n'} p_2$. What is more, the arrows that were rare in p_1 and p_2 are still rare for them, and they have r occurrences in total: this was true by construction of s_1 and s_2 and is true of s'_1 because $p_1 = q_1 a s_1 \sim_{n'} q_2 a s'_1$ and all arrows of q_2 are frequent so the rare subwords of s_1 and s'_1 are the same. The arrows that were frequent in p_1 and p_2 are still frequent in s_1 and s_2 and occur at least as many times as they did in p_1 and p_2 respectively: we have guaranteed this for the arrows of C using Lemma 7.4, and this is clear for the arrows outside of C as all their occurrences in p_1 and p_2 were in s_1 and s_2 respectively, and s'_1 has at least as many occurrences of every letter as s_1 does (this is a consequence of the statement of Lemma 7.5). This ensures that $s'_1 \sim_{n'} s_2$, and that n' is still an $|S|$-distant rare-frequent threshold for them.

Hence, by the induction hypothesis, we have $s'_1 \equiv s_2$, so that by compositionality we have $p_1 \equiv p_2$.

Case 2: C does not occur again after a. We first claim that $q_1 \equiv q_2$ by the base case of the outer induction. Indeed, first note that they are two coterminal paths. Now, every arrow x which is frequent in p_1 and p_2 is either in the SCC of the initial object of a or not. In the first situation, all the occurrences of x in p_1 must be in q_1, as any occurrence of x in s_1 would witness that we are in Case 2; and likewise all its occurrences in p_2 must be in q_2. In the second situation, all its occurrences in p_1 must be in s_1 and all its occurrences in p_2 must be in s_2, for the same reason. Thus, q_1 and q_2 contain no letter which was rare in p_1 and p_2, some of the frequent letters of p_1 and p_2 (those of the other SCCs) do not occur there at all, and the others occur there with the same number of occurrences. Thus indeed $q_1 \sim_{n'} q_2$, they contain no rare arrows, and n' is still an $|S|$-distant rare-frequent threshold for them. Thus, the base case of the outer induction concludes that they evaluate to the same category element.

We now claim that $s_1 \equiv s_2$ by the induction case of the outer induction. Indeed, they are again two coterminal paths. What is more, by the previous reasoning the arrows that are frequent in p_1 and p_2 either occur only in s_1 and s_2 or do not occur there at all. Thus, s_1 and s_2 contain r rare arrows (for the arrows that were already rare in p_1 and p_2), and the frequent arrows either occur in s_1 and s_2 with the same number of occurrences as in p_1 and p_2 or not at all. This implies that n' is still an $|S|$-distant rare-frequent threshold for s_1 and s_2. Thus, we have $s_1 \sim_{n'} s_2$ and the induction case of the outer induction establishes that $s_1 \equiv s_2$.

Thus, by compositionality, we know that p_1 and p_2 evaluate to the same category element. We have concluded both cases of the outer induction proof.