Chapter 9

Conclusions

9.1 Results and Focus

In this thesis, we have examined word operations defined by trajectories, and related problems. Our focus is on four areas: descriptional complexity, codes, language equations and iterated operations.

Chapter 4 considered the problem of state complexity for shuffle on trajectories. As shuffle on trajectories defines a very general family of operations on languages, examining the state complexity of shuffle on trajectories is a challenging problem. The work in Chapter 4 is the first research in state complexity which does not examine a particular operation, but a class of language operations. We have seen that the density of the set of trajectories proves useful for characterizing the state complexity of the resulting shuffle on trajectories operations. We have given upper and lower bounds on the state complexity of shuffle on trajectories for sets of trajectories which are slender, i.e., sets of trajectories which only contain a constant number of trajectories for any fixed length. In this case, we see that there is a substantial advantage over the case where the number of trajectories of any given length is not restricted.

In Chapter 5, we have introduced the operation of deletion along trajectories. Several natural deletion-like operations are particular cases of deletion along trajectories, including quotient, scattered deletion and sequential deletion. The most crucial theoretical aspect of deletion along trajectories is that it serves as an inverse to shuffle on trajectories. The closure properties of deletion along trajectories are different from those of shuffle on trajectories, and we have investigated these properties. The most interesting property is the similarity between the closure properties of deletion along trajectories and proportional removals. This similarity yields many non-regular sets of trajectories for which the associated deletion on trajectories operation preserves regularity.

Chapter 6 investigates classes of languages defined by shuffle on trajectories, and their relation to traditional classes of codes. This investigation has given much insight into the nature of code classes by shifting the focus from classes of languages to the language-theoretic properties of the associated sets of trajectories. We have addressed many natural lines of research in the theory of codes, in particular, questions of maximality, embedding of codes, finiteness properties and the relationship between convexity of languages and code-theoretic properties. While other general mechanisms for defining classes of code-like languages have been presented in the literature, we have found that shuffle on trajectories attains a desirable balance between expressive power on the one hand and the ability to obtain interesting results on the other. In particular, decidability results are obtained through the known closure properties of shuffle and deletion along trajectories.

Our consideration of these classes of codes in Chapter 6 has also led naturally to the definition of a binary relation on the set of all finite words for each set of trajectories. We have investigated algebraic properties of these binary relations; of particular interest is whether a set of trajectories defines a transitive binary relation. Decidability of these properties of the binary relation defined by a set of trajectories are also considered.

As deletion along trajectories constitutes an inverse to shuffle on trajectories, we can therefore investigate language equations involving both shuffle and deletion along trajectories. This is the focus of Chapter 7. We have investigated several different equation forms, and have obtained both positive and negative decidability results for the problem of determining whether an equation possesses a solution. We have also considered systems of equations with shuffle on trajectories, an important step in the investigation of language equations involving shuffle on trajectories.

Finally, Chapter 8 has investigated the questions raised by considering iterated versions of shuffle and deletion along trajectories operations. With a very natural definition of iterated shuffle on trajectories, denoted $(\sqcup\!\sqcup_T)^+(L)$, we can give, for all languages L and all sets of trajectories $T \subseteq \{0,1\}^*$ (resp., all $T \subseteq \{i,d\}^*$ with $T \supseteq i^*$), a characterization of the smallest language containing L and closed under $\sqcup\!\sqcup_T$ (resp., $\rightsquigarrow_T$). We have also studied explicit language equations, i.e., language equations of the form $X = \alpha$ where X is a variable and α is an expression involving $\sqcup\!\sqcup_T$. This study is a fundamental first step towards developing grammar systems involving shuffle on trajectories. In our study, we focus on characterization results for elementary explicit language equations involving shuffle on trajectories. For all $T \subseteq \{0,1\}^*$, we find that the equation $X = X \sqcup\!\sqcup_T X + L$ has least solution $(\sqcup\!\sqcup_T)^+(L)$.

9.2 Open Problems

In this section, we survey some of the more interesting open problems we have considered in this thesis. This is not a complete description of all open problems in this thesis.

Chapter 4 investigated the descriptional complexity of shuffle on trajectories. We concluded with the following conjecture:

Conjecture 9.2.1 *For all $T \subseteq \{0,1\}^*$ with $p_T(n) \in \Omega(2^n)$, there exists L_1, L_2 such that $sc(L_1 \sqcup\!\sqcup_T L_2) = 2^{\Omega(sc(l_1)sc(L_2))}$.*

In chapter 6, we investigated T-codes, which are a natural generalization of many classes of codes. From a theoretical standpoint, the most interesting open problem is characterizing those T for which all T-codes are finite:

Open Problem 9.2.2 *What are necessary and sufficient conditions on a set $T \subseteq \{0,1\}^*$ of trajectories such that $\mathcal{P}_T(\Sigma) \subseteq$ FIN?*

In particular, a characterization for Open Problem 9.2.2 which solves the following problem is of the most interest:

Open Problem 9.2.3 *Given a (regular, context-free) set $T \subseteq \{0, 1\}^*$ of trajectories, is it decidable whether $\mathcal{P}_T(\Sigma) \subseteq \text{FIN}$?*

One fundamental property of sets of trajectories which we have not resolved is freeness, which we discussed in Section 8.10.2.

Open Problem 9.2.4 *Let $T \subseteq \{0, 1\}^*$ be a regular (or context-free) set of trajectories. Is it decidable whether the semigroup $M = (\Sigma^*, \sqcup\!\sqcup_T)$ is a free semigroup?*

There are several open problems concerning iteration of shuffle and deletion on trajectories. The following problem is easily stated, but a solution does not seem obvious:

Open Problem 9.2.5 *For all $R \in \text{REG}$, does there exist $k \geq 1$ such that $(\rightsquigarrow)^+(R) \in \text{CF}_k$?*

9.3 Further Research Directions

In this section, we examine some research directions which we hope to explore in the future.

9.3.1 Confluence of ω_T

Given a transitive, reflexive binary relation ρ on Σ^*, we say that a language $L \subseteq \Sigma^*$ is *confluent with respect to* ρ if, for all $x, y \in L$, there exists $z \in L$ such that $x \, \rho \, z$ and $y \, \rho \, z$.

Ilie [73] has investigated the confluence property for arbitrary binary relations, and examined decidability problems for specific relations, including the prefix, suffix and factor orders. Investigating the decidability of the confluence problem for arbitrary ω_T remains an interesting research question.

9.3.2 Codes Defined by Multiple Sets of Trajectories

Let $n \geq 1$ and $T_i \subseteq \{0,1\}^*$ for $1 \leq i \leq n$. Define $\mathbf{T} = \{T_1, \ldots, T_n\}$. Consider the relation $\omega_{\mathbf{T}}$ on Σ^* defined by

$$x \,\omega_{\mathbf{T}}\, y \iff \bigwedge_{i=1}^{n} x \,\omega_{T_i}\, y$$

for all $x, y \in \Sigma^*$. Define $\mathcal{P}_{\mathbf{T}}(\Sigma)$ as the set of all anti-chains under $\omega_{\mathbf{T}}$. There has been interest in $\mathcal{P}_{\mathbf{T}_{\mathrm{ps}}}(\Sigma)$ for $\mathbf{T}_{\mathrm{ps}} = \{0^*1^*, 1^*0^*\}$, see Jürgensen and Konstantinidis [97, pp. 550–551] for references and a discussion of this class.

Further, we can define a related class $\mathcal{Q}^m_{\mathbf{T}}(\Sigma)$ as follows: $L \in \mathcal{Q}^m_{\mathbf{T}}(\Sigma)$ if and only if for all $L' \subseteq L$ with $|L'| \leq m$, $L' \in \cup_{i=1}^{n} \mathcal{P}_{T_i}(\Sigma)$. For $\mathcal{Q}^m_{\mathbf{T}}(\Sigma)$, other than $\mathbf{T}_{\mathrm{ps}}$, the classes $\mathbf{T}_{\mathrm{io}} = \{0^*1^*0^*, 1^*0^*1^*\}$ [139], as well as $\mathbf{T}_{k-\mathrm{io}} = \{(1^*0^*)^k 1^*, (0^*1^*)^k 0^*\}$ and $\mathbf{T}_{k-\mathrm{ps}} = \{(0^*1^*)^k, (1^*0^*)^k\}$ for $k \geq 1$ (see Long *et al.* [138, Defns. 5 and 6] or Long [137]) have received attention.

It is easily established that $\mathcal{Q}^2_{\mathbf{T}}(\Sigma) = \mathcal{P}_{\mathbf{T}}(\Sigma)$ and that $\mathcal{Q}^{n+i}_{\mathbf{T}}(\Sigma) = \mathcal{Q}^n_{\mathbf{T}}(\Sigma)$ for all $\mathbf{T}$ with $|\mathbf{T}| = n$ and for all $i \geq 0$. However, other problems related to these classes of languages do not seem to be so easy. For instance, the decidability of membership in $\mathcal{P}_{\mathbf{T}_{\mathrm{ps}}}(\Sigma)$ (see Ito *et al.* [76] or Jürgensen *et al.* [98]) relies intrinsically on the nature of the members of $\mathbf{T}_{\mathrm{ps}}$. The corresponding problem for $\mathbf{T}_{\mathrm{io}}$ also relies on the nature of the sets of trajectories involved [38]. It appears to be a very challenging problem to determine the decidability of membership in $\mathcal{P}_{\mathbf{T}}(\Sigma)$ for arbitrary $\mathbf{T} = \{T_1, \ldots, T_n\}$, where each T_i is regular. Kari *et al.* [108, Thm. 4.7] have solved a similar decision problem for two sets of trajectories in their framework of *bond-free property*. However, their approach does not seem to be adaptable to our situation.

9.3.3 Semantic Trajectory-Based Operations

By slightly expanding the notion of a trajectory, we find that trajectory-based operations have several interesting applications. Emerging uses of word operations, especially in modeling operations on strands of DNA, show promise for using shuffle and deletion on trajectories or related trajectory-based variants. Work has already been conducted in this area by Mateescu [144] and more recently by Kari *et al.* [108]. Both of these works focus on using the shuffle on trajectories model to operations on DNA.

However, there is another approach to adapting the trajectory-based framework to model situations such as operations on DNA strands. This approach considers what are known as *semantic* operations. In the paper which introduced shuffle on trajectories, Mateescu *et al.* make the following distinction between *syntactic* and *semantic* operations on words:

> We introduce and investigate new methods [shuffle on trajectories] to define parallel composition of words and languages. These methods are based on syntactic constraints on the shuffle operations. The constraints are referred to as syntactic constraints since they do not concern properties of the words that are shuffled, or properties of the letters that occur in these words.
>
> Instead, the constraints involve the general strategy to switch from one word to another word. Once such a strategy is defined, the structure of the words that are shuffled does not play any role.
>
> However, constraints that take into consideration the inner structure of the words that are shuffled together are referred to as semantic constraints. [147, p. 2]

With the current focus on operations on DNA, we see semantic operations not as a clumsy sibling to syntactic operations, but a challenging area of investigation. We feel that a suitable extension of the shuffle and deletion along trajectories model would provide much insight into the nature of semantic operations. A reasonable model exists which shares many similarities to the current shuffle on trajectories model, but adds sufficient semantic power to encompass many interesting semantic operations, including those investigated by several authors, including Daley *et al.* [29, 30, 31], Kari and Thierrin [116], Kari [107], Mateescu and Salomaa [149], Kudlek and Mateescu [127, 126], de Simone [34], Chen *et al.* [23] and many others; Mateescu *et al.* [146] summarize some of these semantic operations. We also note the possibility of extending the semantic framework proposed by Abdelwahed [1, 2], which is a model of combining processes in control of discrete event systems. Abdelwahed notes that, among others, the concept of shuffle on trajectories is "influential to the modelling paradigm [1, p. 8]" (called Interacting Discrete Event Systems–IDES) developed in the thesis.

It remains to be seen which results can be adapted from the syntactic case to the semantic case.

It is clear, however, that generalizations of some results–in particular, the equations considered by Daley *et al.* [29, 30]–will yield interesting results relating to bio-informatics.