# On the Expressivity of Recurrent Neural Cascades

**Nadezda A. Knorozova**[1] and **Alessandro Ronca**[2]
[1]Relational AI
[2]University of Oxford
nadezda.knorozova@relational.ai, alessandro.ronca@cs.ox.ac.uk

**Abstract**

Recurrent Neural Cascades (RNCs) are the recurrent neural networks with no cyclic dependencies among recurrent neurons. This class of recurrent neural networks is successfully used in practice. Besides training methods for a fixed architecture such as backpropagation, the cascade architecture naturally allows for constructive learning methods, where recurrent nodes are added incrementally one at a time, often yielding smaller networks. Furthermore, acyclicity amounts to a structural prior that even for the same number of neurons yields a more favourable sample complexity compared to a fully-connected architecture. A central question is whether the advantages of the cascade architecture come at the cost of a reduced expressivity. We provide new insights into this question. We show that the regular languages captured by RNCs with sign and tanh activation with positive recurrent weights are the *star-free* regular languages. In order to establish our results we develop a novel framework where the capabilities of RNCs are assessed by analysing which semigroups and groups a single neuron is able to implement. A notable implication of our framework is that RNCs can achieve the expressivity of all regular languages by introducing neurons that can implement groups.

## 1 Introduction

Recurrent Neural Cascades (RNCs) are the subclass of recurrent neural networks where recurrent neurons are cascaded. Namely, they can be layed out into a sequence so that every neuron has access to the state of the preceding neurons as well as to the external input; and, at the same time, it has no dependency on the subsequent neurons. RNCs have been successfully applied in many different areas, including information diffusion in social networks (Wang et al. 2017), geological hazard predictions (Zhu et al. 2020), automated image annotation (Shin et al. 2016), intention recognition (Zhang et al. 2018), and optics (Xu et al. 2020).

RNCs offer several advantages over fully-connected recurrent networks. First, RNCs have a more favourable *sample complexity*, or dually better *generalisation capabilities*. This comes for the reduced number of weights, half the one of a fully-connected recurrent network, which implies a smaller VC dimension (Koiran and Sontag 1998). Second, the acyclic structure of the cascade architecture naturally allows for so-called *constructive learning* techniques (Fahlman 1990; Reed and Marks II 1999). These techniques construct the network architecture dynamically during the training, often yielding smaller networks, faster training and improved generalisation. One such method is *recurrent cascade correlation*, which builds the architecture incrementally adding one recurrent neuron at a time (Fahlman 1990). RNCs emerge naturally here from the fact that every node does not depend on a node added later. RNCs also admit learning methods for fixed architectures, such as *backpropagation through time* (Werbos 1990), where only the weights are learned. For these methods the advantage of the cascade architecture comes from the reduced number of weights.

A central question is whether the advantages of the cascade architecture come at the cost of a reduced *expressivity* compared to the fully-connected architecture. The studies so far have shown that there exist regular languages that are not captured by RNCs with monotone activation such as tanh (Giles et al. 1995). However, an exact characterisation of their expressitivity is still missing. Furthermore, it is unclear whether the inability to capture all regular languages is a limitation of the cascade architecture, or rather of the considered activation functions. We continue this investigation and provide new insights into the capabilities of RNCs to capture regular languages.

## 2 Main Contributions

We develop an analysis of the capabilities of RNCs establishing the following expressivity results.

1. RNCs with sign or tanh activations capture the star-free regular languages. The expressivity result already holds when recurrent weights are restricted to be positive.
2. RNCs with sign or tanh activations and positive recurrent weights do not capture any regular language that is not star-free.
3. Allowing for negative recurrent weights properly extends the expressivity of RNCs with sign and tanh activations beyond the star-free regular languages.
4. We show that in principle the expressivity of RNCs can be extended gradually to all regular languages in a controlled way. It suffices to identify appropriate recurrent neurons. In particular, neurons that can implement finite simple groups. As a first step, we show that second-order sign and tanh neurons can implement the cyclic group of order two.

The first two points establish an important connection between recurrent neural cascades and the *star-free regular languages*. Specifically, they establish the importance of the sign of recurrent weights, and hence isolate the subclass $RNC_+$ of recurrent neural cascades with positive recurrent weights as a particularly important class. In fact, as a corollary of Points 1 and 2, the regular languages recognised by $RNC_+$ are exactly the star-free regular languages.

As a result of our investigation we develop a novel framework where recurrent neural networks are analysed through the lens of *Semigroup and Group Theory*. The framework is of independent interest, as its potential goes beyond our current results. The framework allows for establishing the expressivity of RNCs by analysing the capabilities of a single neuron from the point of view of which semigroups and groups it can implement. If a neuron can implement the so-called *flip-flop monoid*, then cascades of such neurons capture the star-free regular languages. To go beyond that, it is sufficient to introduce neurons that implement *groups*. Our framework can be readily used to analyse the expressivity of RNCs with neurons that have not been considered in this work. In particular, we introduce abstract flip-flop and group neurons, which are the neural counterpart of the flip-flop monoid and of any given group. To show expressivity results, it is sufficient to instantiate our abstract neurons. Specifically in this work we show how to instantiate flip-flop neurons with (first-order) sign and tanh, as well as a family of grouplike neurons with second-order sign and tanh. In a similar way, other results can be obtained by instantiating the abstract neurons with different activation functions.

## 3 Significance of the Results

Our expressivity results provide a more comprehensive understanding of the expressivity of recurrent neural cascades. Our analysis is fine-grained, and it highlights the role of different aspects of the architecture of a neural network such as the role of cyclicity and the sign of recurrent weights. Notably, our results establish an important connection between the subclass $RNC_+$ and the star-free regular languages. This makes $RNC_+$ a strong candidate for learning temporal patterns, since the star-free regular languages are a central class that corresponds to the expressivity of many well-known formalisms. Such formalisms include *star-free regular expressions* from where they take their name (Ginzburg 1968), *monadic first-order logic* on finite linearly-ordered domains (McNaughton and Papert 1971), *past temporal logic* (Manna and Pnueli 1991), and *linear temporal logic* on finite traces (De Giacomo and Vardi 2013). They are also the languages recognised by *counter-free automata* as well as *group-free automata* (Ginzburg 1968). On one hand, our result introduces an opportunity of employing $RNC_+$ for learning targets that one would describe in any of the above formalisms. For such targets, RNCs are sufficiently expressive and, compared to fully-connected recurrent neural networks, offer a more favorable sample complexity along with a wider range of learning algorithms. On the other hand, it places RNCs alongside well-understood formalisms with the possibility of establishing further connections and leveraging many existing fundamental results.

Our results establish a formal correspondence between continuous systems such as recurrent neural networks and discrete abstract objects such as automata, groups, and semigroups. Effectively they bridge recurrent neural networks with *algebraic automata theory*, cf. (Ginzburg 1968), two fields that developed independently and so far have not been considered to have any interaction.

## 4 Relevance to KR

The paper is relevant for KR since it studies the expressivity of recurrent neural networks in terms of formal languages, establishing connections with KR formalisms such as linear temporal logic and automata. Study of the expressivity is a major focus of the KR community, and it applies to neural formalisms in the same way it applies to logic-based formalisms. The paper employs KR methods such as automata theory and formal language theory.

## References


De Giacomo, G., and Vardi, M. Y. 2013. Linear temporal logic and linear dynamic logic on finite traces. In *IJCAI*.

Fahlman, S. E. 1990. The recurrent cascade-correlation architecture. In *NIPS*.

Giles, C.; Chen, D.; Sun, G.-Z.; Chen, H.-H.; Lee, Y.-C.; and Goudreau, M. 1995. Constructive learning of recurrent neural networks. *IEEE Transactions on Neural Networks*.

Ginzburg, A. 1968. *Algebraic Theory of Automata*. Academic Press.

Koiran, P., and Sontag, E. D. 1998. Vapnik-chervonenkis dimension of recurrent neural networks. *Discret. Appl. Math.*

Manna, Z., and Pnueli, A. 1991. Completing the temporal picture. *Theor. Comput. Sci.*

McNaughton, R., and Papert, S. A. 1971. *Counter-Free Automata*. The MIT Press.

Reed, R., and Marks II, R. J. 1999. *Supervised Learning in Feedforward Artificial Neural Networks*. MIT Press.

Shin, H.-C.; Roberts, K.; Lu, L.; Demner-Fushman, D.; Yao, J.; and Summers, R. M. 2016. Learning to read chest X-rays: RNC model for automated image annotation. In *CVPR*.

Wang, J.; Zheng, V. W.; Liu, Z.; and Chang, K. C.-C. 2017. Topological rnn for diffusion prediction. In *IEEE ICDM*.

Werbos, P. 1990. Backpropagation through time: What it does and how to do it. *Proc. of the IEEE*.

Xu, Z.; Sun, C.; Ji, T.; Manton, J. H.; and Shieh, W. 2020. Cascade recurrent neural network-assisted nonlinear equalization for a 100 Gb/s PAM4 short-reach direct detection system. *Optics Letters*.

Zhang, D.; Yao, L.; Zhang, X.; Wang, S.; Chen, W.; Boots, R.; and Benatallah, B. 2018. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. In *AAAI*.

Zhu, L.; Huang, L.; Fan, L.; Huang, J.; Huang, F.; Chen, J.; Zhang, Z.; and Wang, Y. 2020. Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network. *Sensors*.