Recurrent Neural Network for Syntax Learning with Flexible Representations

Xavier Hinaut

Recurrent Neural Network for Syntax Learning with Flexible Representations

Xavier Hinaut[1]

***Abstract*— We present a Recurrent Neural Network (RNN), namely an Echo State Network (ESN), that performs sentence comprehension and can be used for Human-Robot Interaction (HRI). The RNN is trained to map sentence structures to meanings (e.g. predicates). We have previously shown that this ESN is able to generalize to unknown sentence structures in English and French. The meaning representations it can learn to produce are flexible: it enables one to use any kind of "series of slots" (or more generally a vector representation) and are not limited to predicates. Moreover, preliminary work has shown that the model could be trained fully incrementally. Thus, it enables the exploration of language acquisition in a developmental approach. Furthermore, an "inverse" version of the model has been also studied, which enables to produce sentence structure from meaning representations. Therefore, if these two models are combined in a same agent, one can investigate language (and in particular syntax) emergence through agent-based simulations. This model has been encapsulated in a ROS module which enables one to use it in a cognitive robotic architecture, or in a distributed agent simulation.**

I. INTRODUCTION

How do children learn language? In particular, how do they associate the structure of a sentence to its meaning? This question is linked to the more general issue: how does the brain associate sequences of symbols to internal symbolic or sub-symbolic representations? We propose a framework to understand how language is acquired based on a simple and generic neural architecture (Echo State Networks) [1] which is not hand-crafted for a particular task.

First, we present the general ESN architecture. Then, we detail the reservoir sentence processing model. Finally we give some perspectives.

II. METHODS & RESULTS

A. Echo State Networks

The language module is based on an ESN [1] with leaky neurons: inputs are projected to a random recurrent layer and a linear output layer (called "read-out") is modified by learning (which can also be done in an online fashion). The units of the recurrent neural network have a *leak rate* (α) hyper-parameter which corresponds to the inverse of a time constant. These equations define the update of the ESN:

$$\mathbf{x}(t+1) = (1-\alpha)\mathbf{x}(t) + \alpha f(W^{in}\mathbf{u}(t+1) + W\mathbf{x}(t)) \quad (1)$$

$$\mathbf{y}(t) = W^{out}\mathbf{x}(t) \quad (2)$$

with $\mathbf{x}(t)$, $\mathbf{u}(t)$ and $\mathbf{y}(t)$ the reservoir (i.e. hidden) state, the input and the read-out (i.e. output) states respectively at time t, α the *leak rate*, W, W^{in} and W^{out} the reservoir, the input and the output matrices respectively and f the *tanh* activation function. After the collection of all reservoir states the following equation defines how the read-out (i.e. output) weights are trained:

$$W^{out} = Y^d[1; X]^+ \quad (3)$$

with Y^d the concatenation of the desired outputs, X the concatenation of the reservoir states (over all time steps for all train sentences) and M^+ the Moore-Penrose pseudo-inverse of matrix M. Hyper-parameters that can be used for this task are the following: *spectral radius*: 1, *input scaling*: 0.75, *leak rate*: 0.17, *number of reservoir units*: 100.

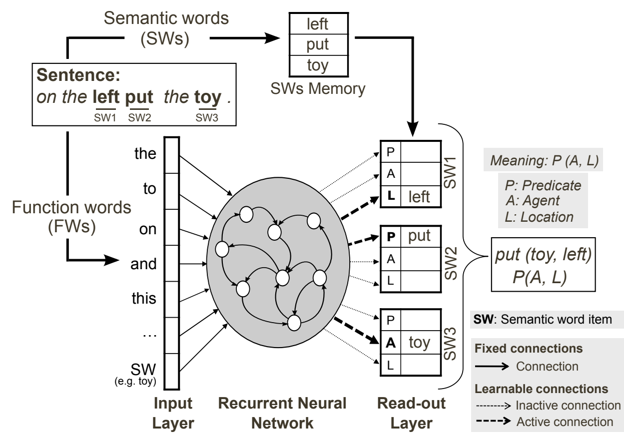

Fig. 1. Sentences are converted to a sentence structure by replacing semantic words by a SW marker. The ESN is given the sentence structure word by word. Each word activates a different input unit. During training, the connections to the readout layer are modified to learn the mapping between the sentence structure and the arguments of the predicates. When a sentence is tested, the most active units are bound with the SW kept in the SWs memory to form the resulting predicate. (Adapt. from [2].)

B. Reservoir Sentence Processing Model

The reservoir sentence processing model has been adapted from previous experiments on a neuro-inspired model for sentence comprehension using ESN [3] and its application to HRI [2]. The model learns the mapping of the semantic words (SW; e.g. nouns, verbs) of a sentence onto the different slots (the thematic roles: e.g. action, location) of a basic event structure (e.g. *action(object, location)*). As depicted in Fig. 1, the system processes a sentence as input and generates corresponding predicates. Before being fed to the

[1]X. Hinaut is with Inria Bordeaux Sud-Ouest, Talence, France; LaBRI, UMR 5800, CNRS, Bordeaux INP, Université de Bordeaux, Talence, France; and Institut des Maladies Neurodégénératives, UMR 5293, CNRS, Université de Bordeaux, Bordeaux, France xavier.hinaut@inria.fr

ESN, sentences are transformed into a sentence structure (or *grammatical construction*) semantic words (SW), i.e. nouns, verbs and adjectives that have to be assigned a thematic role, are replaced by the *SW* item. The processing of the grammatical construction is sequential (one word at a time) and the final estimation of the thematic roles for each SW is read-out at the end of the sentence.

By processing *constructions* [4] and not sentences *per se*, the model is able to bind a virtually unlimited number of sentences to these sentence structures. Based only on a small training corpus (a few tens of sentences), this enables the model to process future sentences with currently unknown semantic words if the sentence structures are similar. One major advantage of this neural network language module is that no parsing grammar has to be defined a priori: the system learns only from the examples given in the training corpus. Here are some input/output transformations that the language model performs:

- "Please give me the mug" $\rightarrow$ *give(mug, me)*
- "Could you clean the table with the sponge?" $\rightarrow$ *clean(table, sponge)*
- "Find the chocolate and bring it to me" $\rightarrow$ *find(chocolate), bring(chocolate, me)*

As shown, the system can robustly transform different types of sentences. Recently, we have explored how flexible our system is: we found that it can handle various kinds of representations at the same time. For instance, it allows to use nouns as main elements of a predicate, and use its arguments to fill in adjectives:

- "Search for the small pink object" $\rightarrow$ *search(object), object(small, pink)*

It also allows to process various kinds of complex sentences:

- "Bring me the newspaper which is on the table in the kitchen" $\rightarrow$ *bring(newspaper, me), newspaper(on, table), table(in, kitchen)*

C. Perspectives

In Hinaut et al. [5], it has been shown that the model can learn to process sentences with out-of-vocabulary words. Moreover, we demonstrated that it can generalize to unknown constructions in both French and English at the same time. To illustrate how the robot interaction works, a video can be seen at `youtu.be/FpYDco3ZgkU` [6][7]. The source code, implemented as a ROS module, is available at `github.com/neuronalX/EchoRob`. With an ESN of 100 units, the training of 200 sentences takes about one second on a laptop computer. Testing a sentence is of the order of 10 ms.

This ROS module could be employed to process various hypotheses generated by a speech recognition system (like in [7]), then returning the retrieved predicates for each hypothesis, thus, enabling a semantic analyser or world simulator to choose the predicates with the highest likelihood. Preliminary work has shown that the model could be trained fully incrementally (i.e. changing weights at each time step) [8].Moreover, an "inverse" version of the model has been also studied, which enables to learn mapping from meaning representations to sentence structure [2]. The representations may not be limited to "series of slots": neural networks are able to deal with hierarchical structures when they are embedded in a vector representation (e.g. Fluid Construction Grammar using Holographic Reduced Representations like in [9]. With these properties, one can investigate language syntax emergence through agent-based simulations.

In a nutshell, the objectives of this model are to improve HRI and provide models of language acquisition. From the HRI point of view, the aim of using this neural network-based model is (1) to gain adaptability because the system is trained on corpus examples (no need to predefine a parser for each language), (2) to be able to process natural language sentences instead of stereotypical sentences (i.e. "put cup left"), and (3) to be able to generalize to unknown sentence structures (not in the training data set). Moreover, this model is quite flexible when changing the output predicate representations, as we have shown here. From the computational neuroscience and developmental robotics point of view, the aim of this architecture is to model and test hypotheses about child learning processes of language acquisition [10].

ACKNOWLEDGMENT

I thank William Schueller for interesting discussions and ideas.

REFERENCES

[1] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148:34, 2001.

[2] X. Hinaut, M. Petit, G. Pointeau, and P. F. Dominey. Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks. *Frontiers in Neurorobotics*, 8, 2014.

[3] X. Hinaut and P.F. Dominey. Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing. *PloS one*, 8(2):e52946, 2013.

[4] A.E. Goldberg. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, 1995.

[5] X. Hinaut, J. Twiefel, M. Petit, P. Dominey, and S. Wermter. A recurrent neural network for multiple language acquisition: Starting with english and french. In *NIPS 2015 Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches*, 2015.

[6] X. Hinaut, J. Twiefel, M. Borghetti Soares, P. Barros, L. Mici, and S. Wermter. Humanoidly speaking – learning about the world and language with a humanoid friendly robot. In *IJCAI Video competition, Buenos Aires, Argentina*. https://youtu.be/FpYDco3ZgkU, 2015.

[7] J. Twiefel, X. Hinaut, M. Borghetti, E. Strahl, and S. Wermter. Using Natural Language Feedback in a Neuro-inspired Integrated Multimodal Robotic Architecture. In *Proc. of RO-MAN*, New York City, USA, 2016.

[8] X. Hinaut and S. Wermter. An incremental approach to language acquisition: Thematic role assignment with echo state networks. In *Proc. of ICANN 2014*, pages 33–40, 2014.

[9] Y. Knight, M. Spranger, and L. Steels. A vector representation of Fluid Construction Grammar using Holographic Reduced Representations. 2015.

[10] M. Tomasello. *Constructing a language: A usage based approach to language acquisition*. Cambridge, MA: Harvard University Press, 2003.