In: Proceedings of the 24th International Conference on Artificial Neural Networks (ICANN 2014),
LNCS 8681, pp. 33-40, Springer Heidelberg. Hamburg, DE, September 15-19, 2014.

An Incremental Approach to Language Acquisition: Thematic Role Assignment with Echo State Networks

Xavier Hinaut & Stefan Wermter

University of Hamburg, Department Informatics, Knowledge Technology, WTM
D - 22527 Hamburg, Germany
hinaut, wermter: @informatik.uni-hamburg.de

Abstract.
In previous research a model for thematic role assignment (θRARes) was proposed, using the Reservoir Computing paradigm. This language comprehension model consisted of a recurrent neural network (RNN) with fixed random connections which models distributed processing in the prefrontal cortex, and an output layer which models the striatum. In contrast to this previous batch learning method, in this paper we explored a more biological learning mechanism. A new version of the model (i-θRARes) was developed that permitted incremental learning, at each time step. Learning was based on a stochastic gradient descent method. We report here results showing that this incremental version was successfully able to learn a corpus of complex grammatical constructions, reinforcing the neurocognitive plausibility of the model from a language acquisition perspective.

Keywords: reservoir computing, recurrent neural network, language acquisition, incremental learning, anytime processing, grammar acquisition.

1 Introduction

How do humans link the form of a sentence and its meaning? Here we propose a general approach to understand how language can be acquired based on a simple and generic neural architecture, namely recurrent neural networks. This approach provides a robust and scalable way of language processing for robotic architectures (Hinaut et al., 2014). The proposed model is not embodied, but it has many requirements to be included in a more global embodied architecture as it uses a generic architecture that is not hand-crafted for a particular task, but can be used for a broad range of applications (see Lukoševičius et al. 2009 for a review).

Mapping the surface form onto the meaning (or deep structure) of a sentence is not an easy task since simply associating words to specific actions or objects is not sufficient to take into account the expressive content of sentences in language. For instance given the two sentences “The cat scratched the dog” and “The dog was scratched by the cat” which have the same meaning but a different focus or point of view, how could a purely word-based system extract the exact meaning of the sentence? How could an infant determine who is doing the action (the *agent*) and who endures the action (the *object* or *patient*)? As simple as this example is, relying only on the semantic (i.e. content) words, and their order in the sentence, will not permit to reliably distinguish the *agent* from the *object*.

To begin to answer this question, we consider the notion of grammatical construction as the mapping between a sentence's form and its meaning (Goldberg 2003). Goldberg defines constructions as "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns". Constructions are an intermediate level of meaning between the smaller constituents of a sentence (grammatical markers or words) and the full sentence itself.

How could one use grammatical constructions to solve this thematic role task for different surface forms? According to the cue competition hypothesis of Bates et al. (1982) the identification of distinct grammatical structures is based on a combination of cues including grammatical (i.e. function) words, grammatical morphemes, word order and prosody. Thus the mapping between a given sentence and its meaning can rely on the order of words, and particularly on the pattern of function words and markers (Dominey et al. 2003). As we will see in section 3, this is the assumption made in the model in order to bind the sentence surface to its meaning.

Typical grammatical constructions could be used to achieve thematic role assignment, that is answering the question "Who did what to whom". This corresponds to filling in the different slots, the roles, of a basic event structure that could be expressed in a predicate form like *predicate(agent, direct object, indirect object* or *recipient).* Different recurrent neural models have been used to process sentences, namely Recursive SOM (Farkas et al. 2008) and ESN (Tong et al. 2007). In contrast, the θRARes model processes grammatical constructions, not sentences, thus it permits putting the emphasis on the exploration of complex sentence structures.

2 Previous work

Previously, it was demonstrated that a recurrent neural network, based on the reservoir computing approach, could learn grammatical constructions by mapping a sentence structure to thematic roles (Dominey et al. 2006; Hinaut et al. 2013). This sentence structure could be thought of as a "sentence category" where content words – or semantic words – are replaced by "slots" that could be filled by any noun or verb according to the position in the sentence structure.

The model processed categories (i.e. abstractions) of sentences called "grammatical constructions". In (Hinaut et al. 2013), the Thematic Role Assignment Reservoir (θRARes) model was able to (1) process the majority of the given grammatical constructions in English correctly, even if not learned before, demonstrating generalization capabilities, and (2) to make online predictions while processing a grammatical construction. Moreover, the model provided insights on how language could be processed in the brain. Hinaut et al. were able to show that for less frequent inputs an important shift in output predictions occurred. It was proposed that this could be a potential explanation for human electrophysiological data, like event-related potentials (e.g. P600) that occur when processing unusual sentences (Hinaut et al. 2013).

This language model was successfully used in the iCub humanoid robot platform. Naïve users could teach the robot basic language capabilities through interaction with the robot (Hinaut et al., 2014).

A very similar model of prefrontal cortex was used to model abstract motor sequence processing, and was able to represent categorical information inside the reservoir (Hinaut et al. 2011), thus reinforcing the plausibility of a common generic area used for both abstract motor sequence processing and syntactic processing.

3 Material and Methods

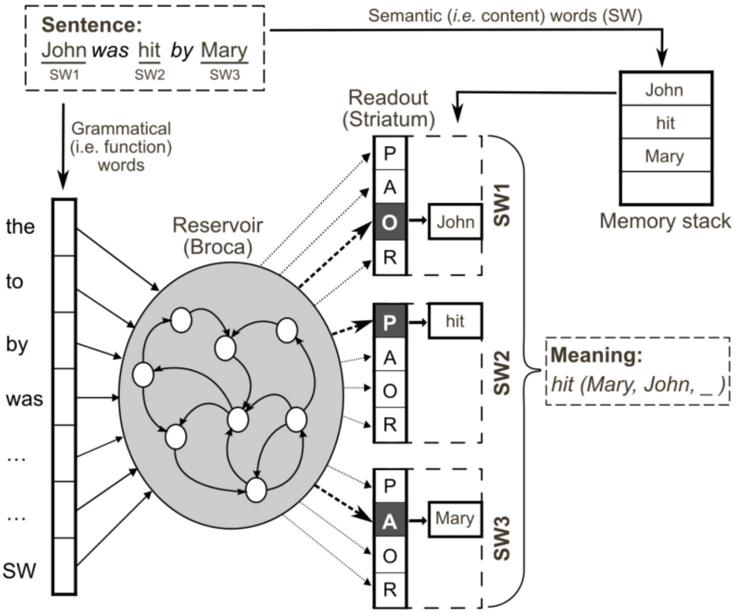

Fig. 1. Thematic Role Assignment Reservoir model (θRARes). All the process from the input sentence (top left) to output predicate-form meaning (middle right) is displayed. Semantic words (SWs) are extracted from the input sentence, in order to transform the sentence into grammatical construction before giving it as input to the reservoir (one word at a time). The output meaning is expressed as the following: predicate(agent, object, recipient). Dashed connections (between reservoir and readout) are the only ones modified during learning. The schema is the same for the incremental version (i-θRARes). Inspired from (Hinaut et al. 2013).

3.1 The neural language model

The core part of the θRARes model is an Echo State Network (Jaeger et al. 2004). It is a recurrent neural network with sparse, random and fixed connectivity: in this paper we will refer to this core part as the "reservoir". Only the weights from the reservoir to the readout (thematic role layer) are learned. Input and reservoir weights are generated randomly. In the previous model, weights were set by linear regression in batch form. Please refer to (Hinaut et al. 2013) for a more detailed description.

3.2 Input and output coding

The localist input at each time step t is coded as a symbolic vector, with all values except one set to zero. The only non-zero input (i.e. equal to 1) is the current word. At each time step a word is presented (*i.e.* activation time=1). The input vector dimension is 13. This corresponds to 10 for grammatical words or markers '-ed', '-ing', '-s', 'by', 'is', 'it', 'that', 'the', 'to', 'was', 1 for SW, 1 for the comma and 1 for the period. One can see that the verb inflexions (such as '-ed', '-s', '-ing') are part of the grammatical markers. The maximum length of an input is 20 time steps, because constructions has at maximum 19 words and a final period. All nouns and verbs are coded by a common input marker 'SW' that just indicates the presence of a Semantic Word, independently whether it is a noun or a verb. Thus one could not distinguish between nouns and verbs based on the raw input. There are maximum 6 semantic words per grammatical construction (i.e. sentence abstraction).

The output is coded as follows: for each noun (content word), there are 4 readout units for the different thematic roles: agent, predicate, object, recipient. As there could be at maximum 2 clauses in a construction (main and relative), there are 8 output units for each semantic word (SW) in total, because each SW could have a role in each clause. Thus the output vector dimension is 4*2*6=48. The target output is as the following: all output units corresponding to the correct thematic roles are activated, at 1, since the beginning of the sentence, and all other units remain at zero. Note that in the corpus used here the relative clauses do not have semantic words with recipient (R) role, so the total number of effective outputs is 42.

One interesting property of the output is that it is sparse. In fact semantic word (SW) could be associated to only one slot for each predicate-form. If the output is represented as a table with each row being a given SW and each column is an atomic role (e.g. A-2: agent of predicate-form 2), we can see that each column has at maximum one unit activated because one atomic role corresponds to none or to only one SW.

We used the same error measures as in Hinaut et al. (2013): the *meaning* error is the average number of correct thematic roles, averaged over all constructions; and the *sentence* error is the average of sentences/constructions fully recognized (*i.e.* the percentage of sentences that are fully understood). The latter is a very strict measure.

3.3 Incremental learning

We want to extend the neurocognitive plausibility of this reservoir language model approach by using a simple incremental learning algorithm based on stochastic gradient descent, namely the Least Mean Square (LMS) algorithm. We used only the information available at a given time step t in order to modify the weights. During the learning phase, the error between the model output and the desired output is computed. The error is used to modify the weights (from the reservoir to the read-out layer) for the last 2 time steps of each construction (i.e. during the last word and the final period). Thus the output weights are modified only at the end of the sentence, when enough information has been processed to establish the thematic roles. The error made by the output units is defined as follows:

$$\mathrm{err}(t+1) = \mathbf{W}_{\mathrm{out}}(t) \,.\, \mathbf{v}(t+1)' - \mathbf{d}(t+1) \quad (1)$$

with v(t) the reservoir state at time t, d(t) the desired activity at time t, $W_{out}(t)$ the output weight matrix at time t. **x**' indicates the transpose of vector **x**. During training the weights are updated as follows:

$$W_{\mathrm{out}}(t+1) = \mathbf{W}_{\mathrm{out}}(t) - \mathrm{eta} * \mathrm{err}(t+1).\mathbf{v}(t+1)' \quad (2)$$

with eta the learning rate, and with Wout(0)=0. We found that a learning rate of 10-3 permitted to obtain reasonable results. Note that the output activity at time step t+1 is computed with the previous error value. The output activity of the readout units is as follows:

$$\mathrm{out}(t+1) = W_{\mathrm{out}}(t).v(t+1)' \quad (3)$$

3.4 Corpus and reservoir parameters

The corpus was first used in Hinaut et al. (2012). The full corpus can be found in the supplementary materials of Hinaut et al. (2013). It contains 462 English grammatical constructions with their corresponding predicate-meaning. Constructions could have zero or one relative clause (e.g. "The boy *that hit the cat* ate the apple."), which corresponds to one level of nesting. Each construction thus could have 1 or 2 verbs, and a total of 6 semantic words, of which 2 belong to the core part of the relative clause, and 1 belongs to both the main and relative clauses. This corpus was generated in order to have all the possible constructions combining 6 semantic words in all the possible orders. Given the following notation A: agent, P: predicate, O: object, R: recipient, and m or r denotes the main or relative resp., here are examples of semantic word order with their corresponding constructions:

- m(AP), r(): "The giraffe walk -s ."
- m(AOP), r(): "By the beaver the fish was cut -ed ."
- m(APRO), r(): "The dog give -s to the mouse the cat ."
- m(APO), r(AP): "The beaver that think -s cut -s the fish ."
- m(PA), r(AP): "Walk -ing was the giraffe that think -s ."
- m(PARO), r(OAP): "Give -ed by the dog to the mouse that the guy kiss -s was the cat ."

Semantic Word (SW) markers were replaced by words so that human readers could understand them, there are shown in italic. Note that this corpus includes unlikely word orders, so is more difficult to learn than other generated corpora for a given number of semantic words. On the contrary, the sentence structures generated in (Miikkulainen, 1996) are much more regular.

The parameters of the reservoir were the following: units in the ESN: 500 ; spectral radius: 6; input scaling: 2.75; time constant: 6; input connectivity: 10%; reservoir connectivity: 20%; activation time (i.e. number of time steps during which each word was presented): 1; activation function of ESN units: hyperbolic tangent (tanh).

4 Results

The error measures obtained by 10-fold cross-validation (CV) for 4 reservoir instances of 500 units each during 1000 epochs of training are the following: 17.6% (±0.8) for the meaning error, and 86.5% (±1.5) for the sentence error. Better results could be obtained by increasing the number of units; for 4 instances with 1000 units during 500 epochs: 15.6% (±0.8) for the meaning error, and 81.9% (±2.9) for the sentence error. With 1000 units, even better results could be obtained with more epochs (<70% for the sentence error for 5000 epochs for instance).

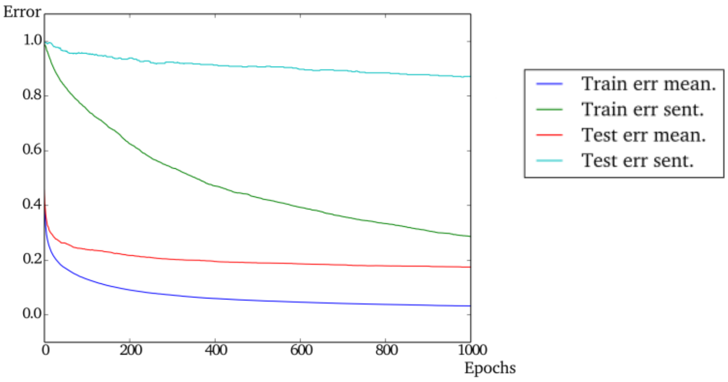

Fig. 2. Meaning and sentence errors over 1000 epochs of training for the i-θRARes model, obtained with a 10-fold CV for a reservoir instance. Number of neurons: 500. The minimum error values are: error meaning: train 3.3%, test 17.5%; error sentence: train 28.7%, test 87.0%.

In Figure 2 we can see an example of the decay of both meaning and sentence errors across training epochs. In Figure 3 a subset of output (readout) units for the construction 244 can be seen. Even if the weights are modified only at the end of the sentence during learning, spontaneous representation emerges after learning.

5 Discussion

It was shown in Hinaut et al. (2013) that forcing the network to decide (about a thematic role) as soon as possible, by considering the activity during the whole sentence to learn the weights, enables the output units to represent the on-going pseudo-probabilities of the current parse of the grammatical construction. This means that even before the network has enough input information to assign the role for each semantic word, the target output is already used to compute the current error.

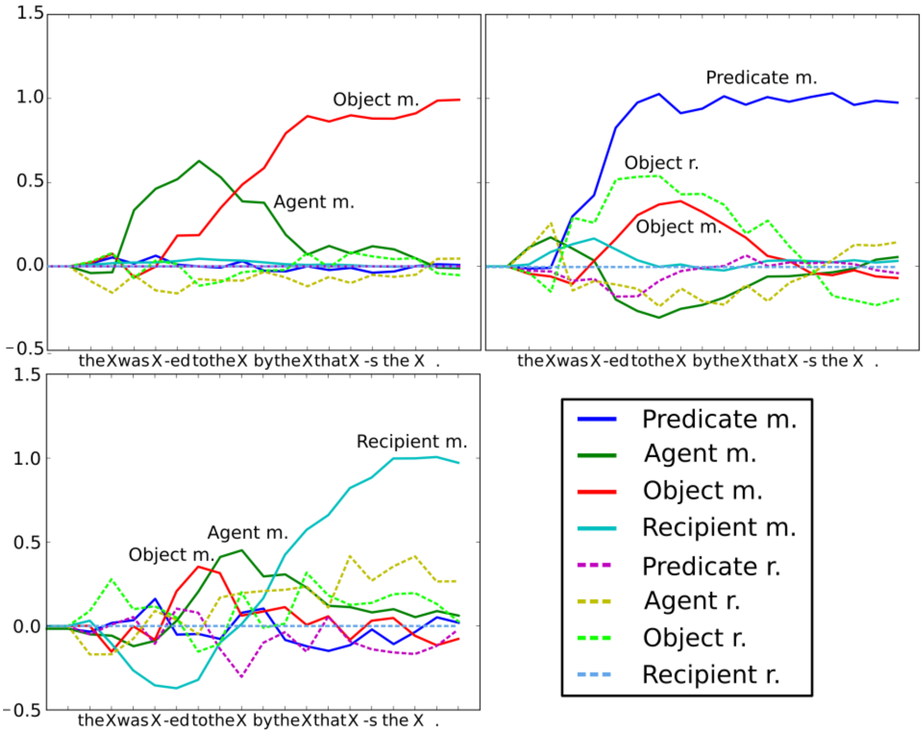

Fig. 3. Subset of output units for the construction 244: "The *cat* was *give* -ed to the *mouse* by the *dog* that *kiss* -s the *girl* ." after only 500 epochs of training. Output units for the first three semantic words (SW) are shown (top left, top right, bottom left resp.). On the x-axis we can see the corresponding input words at each time step. The output activity becomes active since the beginning of the construction. All possible thematic roles for each SW are shown for both possible clauses: main (m.) and relative (r.).

Here, only the last two time steps are used to learn the weights. Nevertheless, a spontaneous output representation seems to emerges during the sentence processing. This spontaneous activity seems different for different reservoir instances. It remains to be explored what kind of information is embedded in this representation, and if an average of the activities over several reservoir instances might produce pseudo-probabilities even without target forcing since the beginning of the sentence. This spontaneous output representation is an interesting property to understand how such on-going information on the current parse might be represented in the human brain.

Direct comparison with performances obtained in Hinaut et al. (2013) is not possible here due to the important number of epochs and reservoir units needed. But it is already very interesting that generalization works with a simple incremental algorithm (LMS) using a complex corpus. Performance could also be improved using an activation time (i.e. number of time steps during which a word is presented) greater than 1 and by using a corpus with redundancies (e.g. small variations of grammatical constructions by adding adjectives; see Hinaut et al. (2013)). This would enable the model to learn from more diverse data points and provide a more robust learning. It also remains to be investigated if the addition of feedback connections from the read-out to the reservoir may enhance the performance. In this way the output "representations" obtained during the processing of a grammatical construction may be of use to constrain the choice of possible thematic roles. In fact, adding feedback should reduce the dimensionality of the reservoir state (Hoerzer, 2012), providing attractors easier to learn on.

Acknowlegment.
This research has been partly supported by the EU 235065 ROBOT-DOC from the 7th Framework Programme (FP7), Marie Curie Action ITN.

References.

1. Bates E, McNew S, MacWhinney B, Devescovi A, Smith S (1982) Functional constraints on sentence processing: a cross-linguistic study. Cognition 11: 245–299.
2. Dominey PF, Voegtlin T (2003) Learning word meaning and grammatical constructions from narrated video events. Proc HLT-NAACL.
3. Dominey PF, Hoen M, Inui T (2006) A neurolinguistic model of grammatical construction processing. J Cogn Neurosci 18: 2088–2107.
4. Farkas I, Crocker MW (2008) Syntactic systematicity in sentence processing with a recurrent self-organizing network. Neurocomputing 71: 1172–1179.
5. Goldberg AE (2003) Constructions: a new theoretical approach to language. Trends Cogn Sci 7: 219–224.
6. Hinaut X, Dominey PF (2011) A three-layered model of primate prefrontal cortex encodes identity and abstract categorical structure of behavioral sequences. J Physiol - Paris 105: 16–24.
7. Hinaut, X, Dominey PF (2012) On-Line Processing of Grammatical Structure Using Reservoir Computing. In Artificial Neural Networks and Machine Learning – ICANN 2012, LNCS vol. 7552, 2012, pp 596-603.
8. Hinaut, X, Dominey PF (2013) 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.
9. Hinaut X, Petit M, Pointeau G Dominey PF (2014) Exploring the Acquisition and Production of Grammatical Constructions Through Human-Robot Interaction with Echo State Networks. Front. Neurorobot. 8:16.
10. Hoerzer GM, Legenstein R, Maass W (2012) Emergence of complex computational structures from chaotic neural networks through reward-modulated Hebbian learning. Cereb Cortex, Advance online publication. Retrieved November 11, 2012.
11. Jaeger H, Haas H (2004) Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science, 304, 78–80.
12. Lukoševičius M, Jaeger H (2009) Reservoir computing approaches to recurrent neural network training. Comput Sci Rev 3: 127–149.
13. Miikkulainen R (1996) Subsymbolic case-role analysis of sentences with embedded clauses. Cognitive Sci 20: 47–73.
14. Tong MH, Bickett AD, Christiansen EM, Cottrell GW (2007) Learning grammatical structure with Echo State Networks. Neural networks 20: 424–432.