



HAL
open science

Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs

Xavier Hinaut, Alexandre Variengien

► **To cite this version:**

Xavier Hinaut, Alexandre Variengien. Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs. SNL 2020 - 12th Annual Meeting of the Society for the Neurobiology of Language, Oct 2020, Virtual Edition, Canada. hal-03146558

HAL Id: hal-03146558

<https://inria.hal.science/hal-03146558v1>

Submitted on 19 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recurrent Neural Networks Models for Developmental Language Acquisition: Reservoirs Outperform LSTMs

Xavier Hinaut^{1,2,3*}, Alexandre Variengien^{1,2,3,4}

¹Inria, Bordeaux, France. ²LaBRI, UMR 5800, Talence, France. ³Neurodegeneratives Diseases Institute, UMR 5293, Bordeaux, France. ⁴Ecole Normale Supérieure de Lyon, France. *xavier.hinaut@inria.fr

INTRODUCTION

We previously developed cortico-striatal models for sentence comprehension (Hinaut & Dominey 2013) and sentence production (Hinaut et al. 2015). The sentence comprehension model is based on the reservoir computing principle (Jaeger & Haas 2004): a random recurrent neural network (a reservoir) provides a rich recombination of sequential word inputs (e.g. a piece of prefrontal cortex), and an output layer (e.g. striatum) learns to "reads-out" the roles of words in the sentence from the internal recurrent dynamics. The model has several interesting properties, like the ability to predict the semantic roles of words during online processing. Additionally, we demonstrated its robustness to various corpus complexities, in different languages (Hinaut & Twiefel 2019), and even its ability to work with bilingual inputs.

ABSTRACT

In this study, we propose to (1) use the model in a new task related to a developmental language acquisition (i.e. Cross-Situational Learning) (Juvén & Hinaut 2020; Dinh & Hinaut 2020), (2) provide a quantitative comparison with one of the best performing neural networks for sequential tasks (a LSTM), and (3) provide a qualitative analysis on the way reservoirs and LSTMs solve the task. This new Cross-Situational Task is as follows: for a given sentence, the target output provided often contains more detailed features than what is available in the sentence. Thus, the models have not only to learn how to parse sentences to extract useful information, but also to statistically infer which word is associated with which feature. While reservoir units are modelled as leaky average firing rate neurons, LSTM units (Hochreiter & Schmidhuber, 1997) are engineered to gate information using a costly and biologically implausible learning algorithm (Back-Propagation Through Time).

RESULTS

We found that both models were able to successfully learn the task: the LSTM reached slightly better performance for the basic corpus, but the reservoir was able to significantly outperform LSTMs on more challenging corpora with increasing vocabulary sizes (for a given set of hyperparameters). We analysed the hidden activations of internal units of both models. Despite the deep differences between both models (trained or fixed internal weights), we were able to uncover similar inner dynamics: the most useful units (with strongest weights to the output layer) seemed tuned to keep track of several specific words in the sentence. Because of its learning algorithm, it is predictable to see such behaviour in a LSTM but not in a reservoir; in fact, the LSTM contained more tuned-like units than the reservoir.

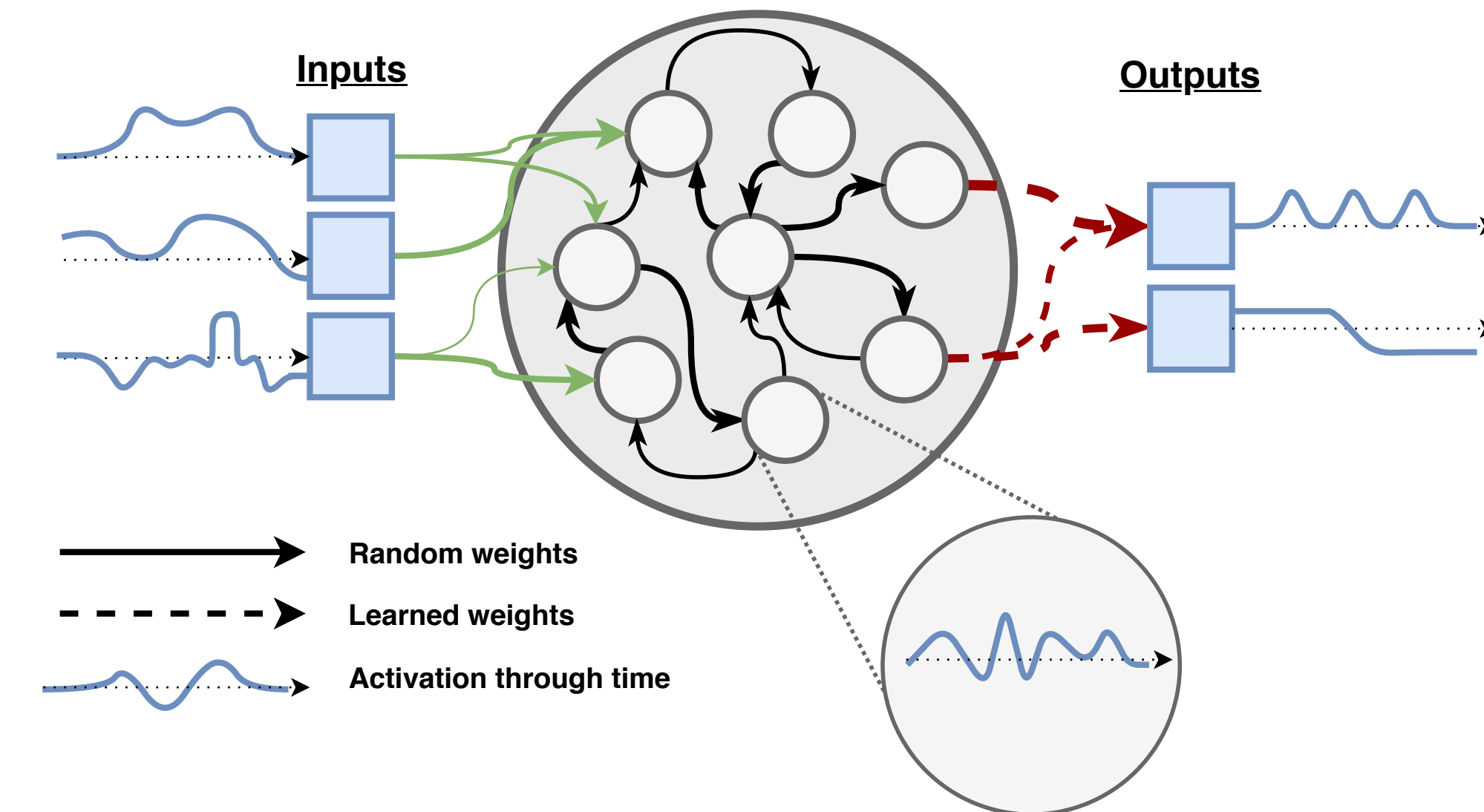
CONCLUSION

The qualitative and quantitative differences between LSTMs and reservoirs highlights the gap between classical Deep Learning approaches (based on back-propagation algorithm) and more plausible brain learning mechanisms. First, the reservoir is more efficient in terms of training time and cost (the LSTM needs several passes on the training data, while the reservoir uses only one). Secondly, only the reservoir model seems to scale to larger corpora without the need to adapt specifically the hyper-parameters of the model. Finally, the presence of more tuned units in the LSTM compared to the reservoir might be an explanation of why the LSTM overfits to the training data. More generally, a LSTM seems to have more limited generalization capabilities than a reservoir when available data is limited.

METHODOLOGY

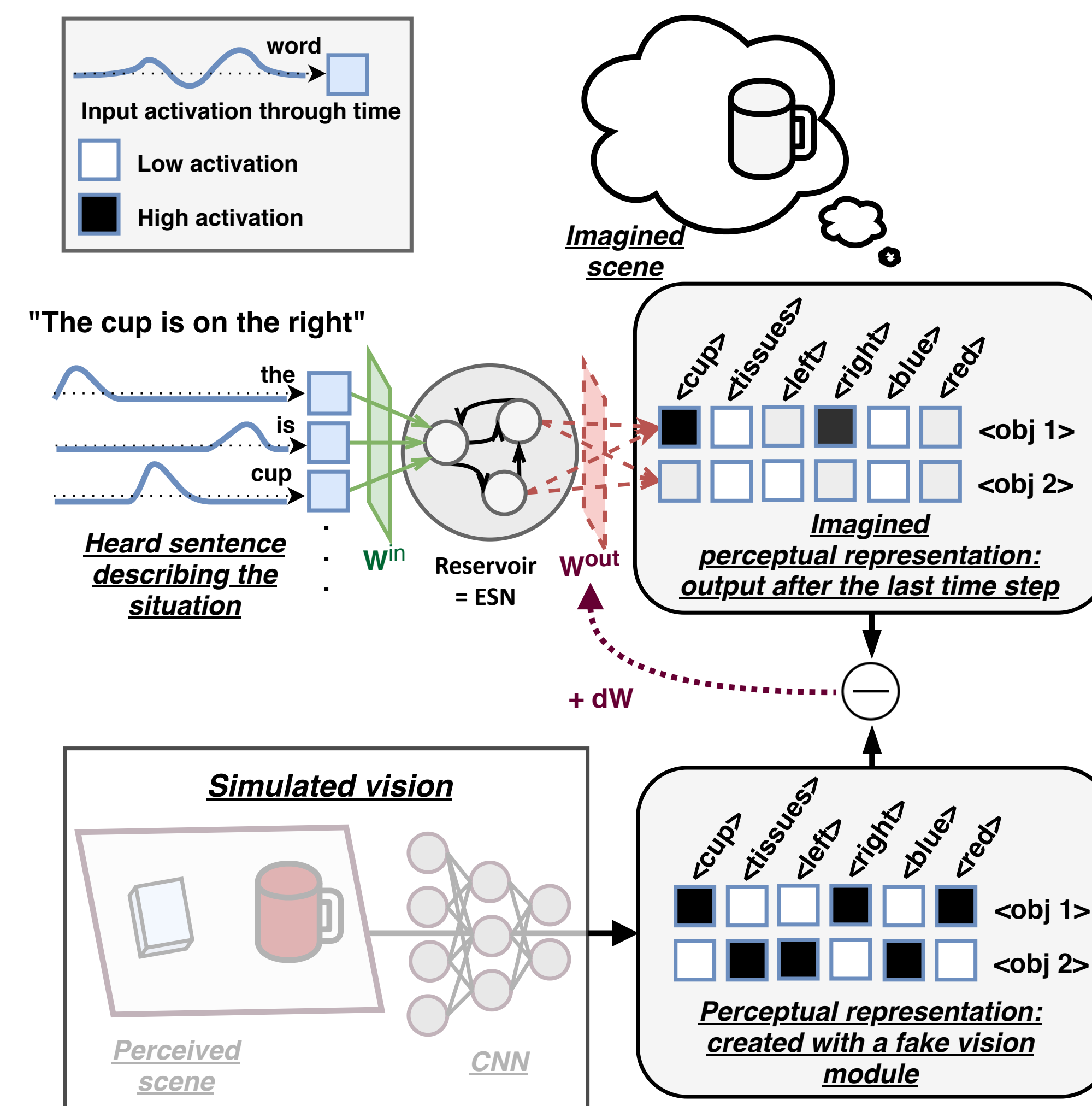
The Reservoir Computing principle as cortico-striatal model

Only the output connections (red dashed) are trained, the input (green) and recurrent connections (black) are randomly generated and kept fixed. Non-linear dynamics inside the reservoir create "a pool of inputs recombinations" from which interesting features could be "read out" from the outputs. The reservoir used here is an Echo State Network (ESN) [3].



Model and set-up

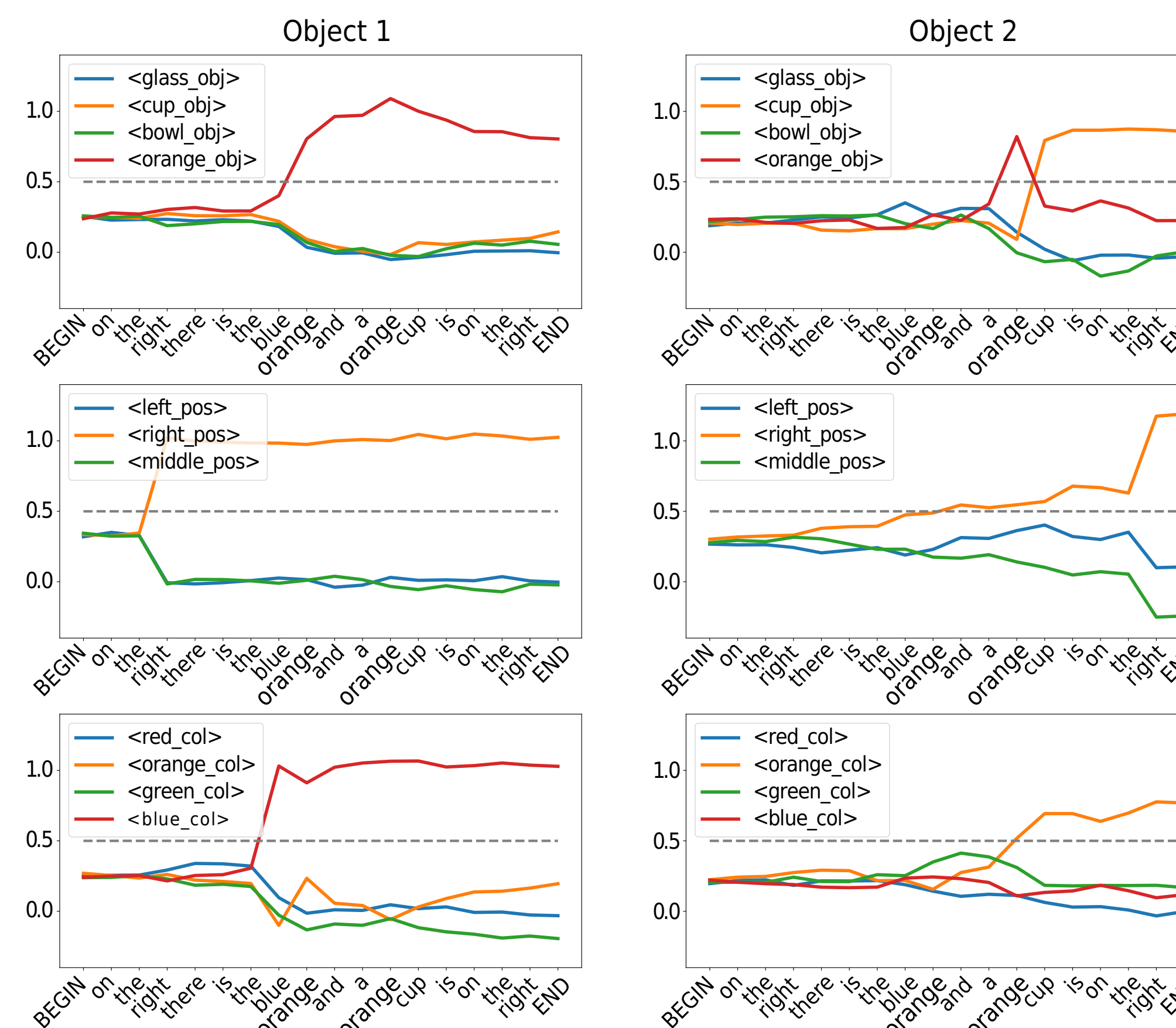
The model has to reconstruct an imagined scene from the sentence given word by word. The simulated vision creates a perceptual representation corresponding to the full description of objects in the scene. This representation is used as target outputs for the reservoir, even if the sentence only partially describes the objects in the scene, or if it describes only one object. Thus the model has to learn statistically which word is linked to which feature. This particular set-up creates cross-situational learning conditions similar to the ones children are facing. The set-up, input and target outputs were the same for the LSTM experiments.



RESULTS

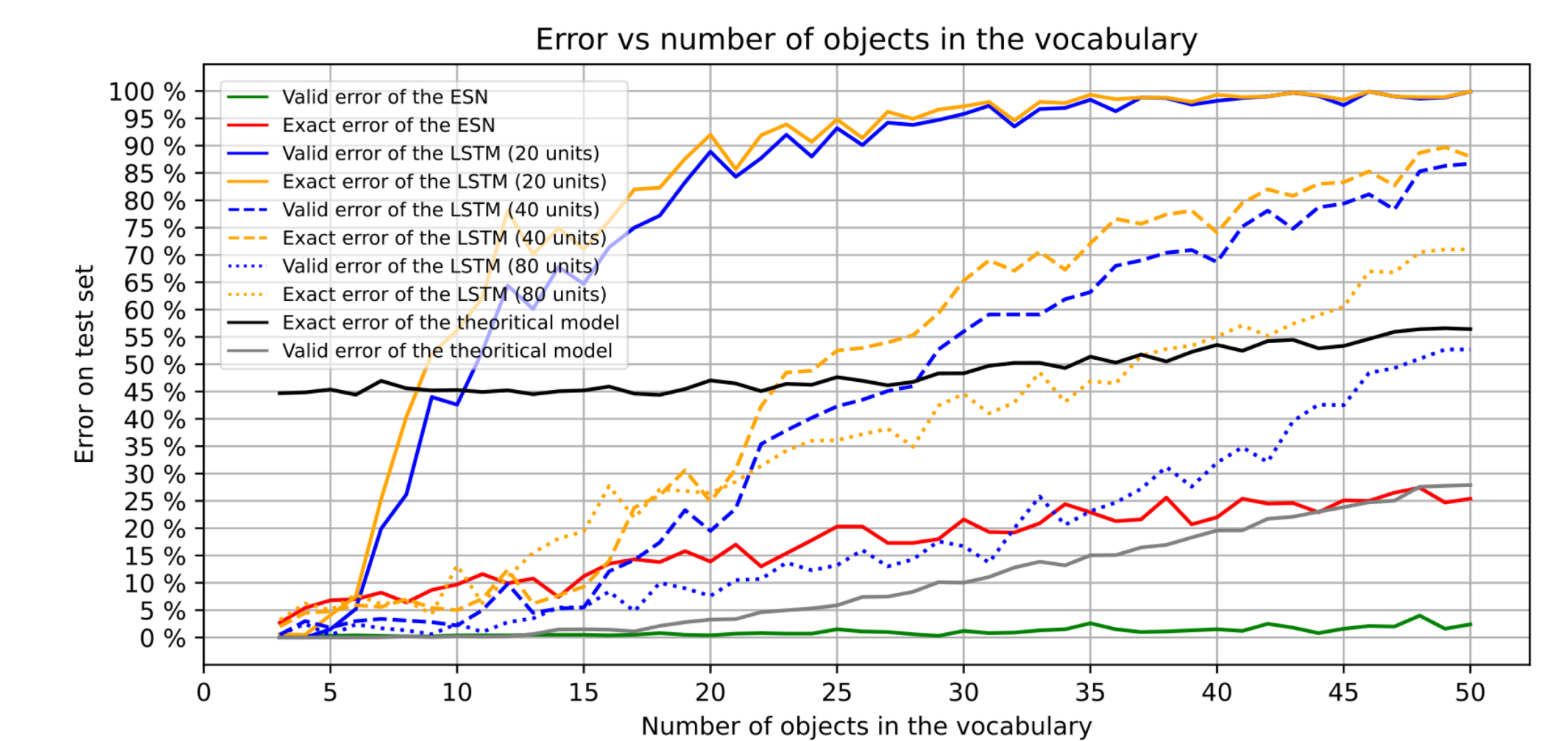
Outputs of the reservoir after training

x-axis is time, y-axis is the output neurons activities. We can see a jump in the related output after a keyword is seen. Which means that after each word the model updates the outputs. This is an interesting property of the model because it was only trained to give the output during the whole sentence, not when a specific word appeared. In the *Object 2* outputs (left), we see that "orange" is first recognized as an object (noun), but when "cup" arrives it is then recognized as a color (adjective).



RESULTS

Testing the generalisation capabilities of reservoirs and LSTMs. Averaged errors for test data set for the reservoir (=ESN) and different sizes of LSTMs (20/40/80 units). Hyperparameters were optimized for all models with 4 objects in the dataset.



"The cup is on the right!"	Valid ?	Exact ?
	X	X
	✓	X
	✓	✓

REFERENCES

- Hinaut & Dominey, Plos ONE, 2013.
- Hinaut et al., Brain & Language, 2015.
- Jaeger & Haas, Science, 2004.
- Hinaut & Twiefel, IEEE TCDS, 2019.
- Juven & Hinaut, IJCNN, 2020.
- Dinh & Hinaut, ICANN, 2020.
- Hochreiter & Schmidhuber, Neural Computation, 1997.

RESULTS

Detailed of internal neuron activations for a LSTM and a reservoir

(Left) LSTM: (top) Activity variation of all 20 neurons of a LSTM. (bottom) Detail of 1 neuron particularly sensitive to the word "and". (Right) Reservoir: (top) Activity variation of a selection of neurons. (bottom) Raw activity of a selection of neurons in the reservoir and sum of variations.

