A Recurrent Neural Network for Multiple Language Acquisition: Starting with English and French

Xavier Hinaut, Johannes Twiefel, Maxime Petit, Peter Dominey, Stefan Wermter

A Recurrent Neural Network for Multiple Language Acquisition: Starting with English and French

Xavier Hinaut*
Dept. of Informatics, Uni. of Hamburg
Hamburg, Germany
hinaut@informatik.uni-hamburg.de

Johannes Twiefel
Dept. of Informatics, Uni. of Hamburg
Hamburg, Germany
twiefel@informatik.uni-hamburg.de

Maxime Petit
SBRI, INSERM 846
Bron, France
m.petit@imperial.ac.uk

Peter Dominey
SBRI, INSERM 846
Bron, France
peter.dominey@inserm.fr

Stefan Wermter
Dept. of Informatics, Uni. of Hamburg
Hamburg, Germany
wermter@informatik.uni-hamburg.de

Abstract

How humans acquire language, and in particular two or more different languages with the same neural computing substrate, is still an open issue. To address this issue we suggest to build models that are able to process any language from the very beginning. Here we propose a developmental and neuro-inspired approach that processes sentences word by word with no prior knowledge of the semantics of the words. Our model has no "pre-wired" structure but only random and learned connections: it is based on Reservoir Computing. Our previous model has been implemented in the context of robotic platforms where users could teach basics of the English language to instruct a robot to perform actions. In this paper, we add the ability to process infrequent words, so we could keep our vocabulary size very small while processing natural language sentences. Moreover, we extend this approach to the French language and demonstrate that the network can learn both languages at the same time. Even with small corpora the model is able to learn and generalize in monolingual and bilingual conditions. This approach promises to be a more practical alternative for small corpora of different languages than other supervised learning methods relying on big data sets or more hand-crafted parsers requiring more manual encoding effort.

1 Introduction

How do children learn language? In particular, how do they link the structure of a sentence to its meaning? This question is linked to the more general issue: How does the brain link 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 [1, 2] which is not hand-crafted for a particular task, but on the contrary can be used for a broad range of applications (see [3] for a review). This idea of "canonical" neural circuits has been coined by several authors:

*www.informatik.uni-hamburg.de/~hinaut ; source code available soon at github.com/neuronalX

it is an aim in computational neuroscience to model generic pieces of cortex [4, 5]. As such simplified canonical circuits we use Echo State Networks (ESN) [1] which are neural networks with a random recurrent layer and a single linear output layer (called "read-out") modified by online or offline learning.

Much research has been done on language processing with neural networks [6, 7, 8] and more recently also with Echo State Networks (ESN) [9, 10]. The tasks used were diverse, from predicting the next symbol (i.e. word) in a sentence to thematic role assignment. In this paper, the task we perform is the latter. Previously the kind of inputs sequence used was mainly based on one-level symbols, i.e. symbols that belong to the same level of abstraction (e.g. only words). However language, like many other cognitive tasks, contains several levels of abstraction, which could be represented hierarchically from phonemes to discourse. Hierarchical processing is a strategy of the brain, from raw perception layers to abstract processing, and even inside the higher-level cognitive computations performed by the prefrontal cortex there exists a hierarchy of processes [11].

Some recent results with end-to-end word recognition from raw audio data with RNN are impressive [12]. This could give some insights on what kinds of features are extracted by the brain during speech processing. However, to uncover language acquisition mechanisms other modelling methods are needed. It is likely that the brain builds hierarchical representations in a more incremental and less supervised way: each newly built abstraction enables the formation of the next higher-order abstraction, instead of abstracting all at once. We hypothesize that the brain canonical circuits, such as the simple version proposed, can deal with different levels of abstraction mixed at the same time. In this paper, we present an initial model version which is able to deal with three kind of symbols: Function Words (FWs), Semantic Words (SWs) and Infrequent function Words (IWs). This model processes IWs and SWs as categories of words, thus they are on a different level of abstraction than FWs, which are processed as words (see Figure 1). Therefore, we have two levels of abstraction. Moreover, we show for the first time that this model is able to learn and generalize over a French corpus, and additionally over two languages at the same time, namely English and French.

2 Reservoir computing and grammatical construction approach

2.1 Echo State Networks (ESN)

The model is based on an Echo State Network (ESN) with leaky neurons. 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:

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

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

with $x(t)$, $u(t)$ and $y(t)$ the reservoir state, the input vector and the read-out vector 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 is the *tanh* activation function. After collection of all reservoir states the following equation defines how the read-out weights are computed:

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

with Y^d the concatenation of the desired outputs, X the concatenation of the reservoir states over all time steps and M^+ the Moore-Penrose pseudo-inverse of matrix M.

2.2 The sentence comprehension model with grammatical constructions approach

The proposed model processes sequences of symbols as input (namely sequences of words) and generates a dynamic probabilistic estimation of static symbols (namely thematic roles), see Figures 1 and 2. This work is based on a previous approach modelling human language understanding [13, 14], human-robot interaction [15, 16], and language acquisition in a developmental perspective (with incremental and online learning) [17]. Recently an "inverse" version of the model was able to produce sentences depending on the words of focus [18]. The general aim of having an architecture working with robots is to use them to model and test hypotheses about child learning processes of language acquisition [19]. It is also interesting to enhance the Human-Robot Interactions and it has

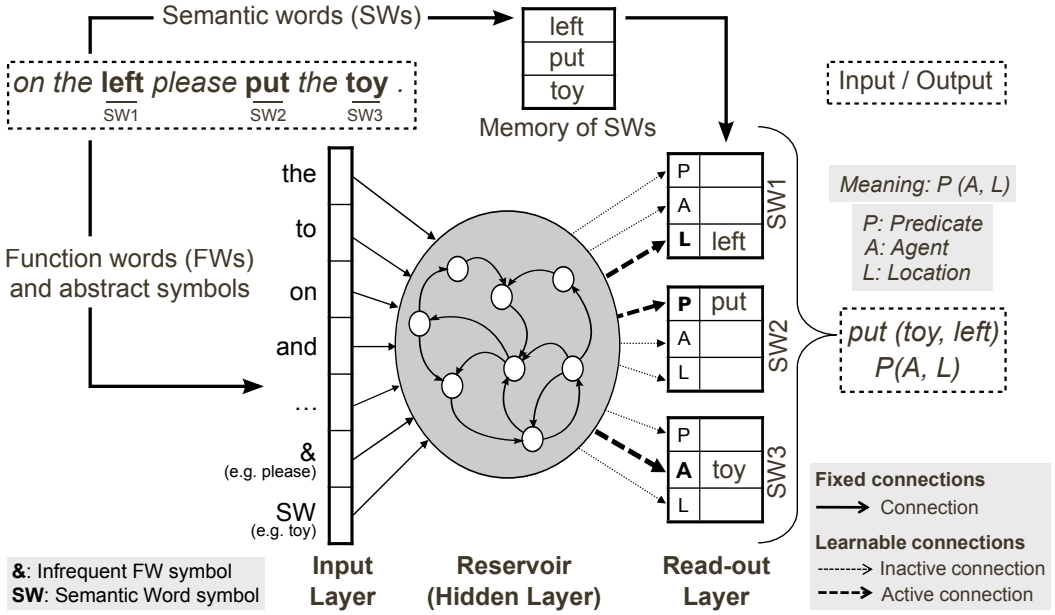

Figure 1: The sentence comprehension model enhanced with IW replacement.

already been implemented in humanoid robotic architectures (iCub and Nao) [15, 20] to enable users to use natural language instead of stereotyped vocal commands during interactions. To illustrate how the system works a video is available at [20].

Mapping the surface form (sequence of words) onto the deep structure (meaning) of a sentence is not an easy task since making word associations is not sufficient. For instance, a system relying only on the order of semantic words (cat, scratch, dog) to assign thematic roles is not enough for the following simple sentences, because even if *cat* and *dog* appear in the same order they have different roles: *"The **cat scratched** the **dog**"* and *"The **cat** was **scratched** by the **dog**"*. It was shown that infants are able to quickly extract and generalize abstract rules [21], and we want to take advantage of this core ability. Thus, to be able to learn such a mapping (from sequence of words to meaning) with a good generalization ability it seems natural to rely on abstractions of sentences rather than sentences themselves. In order to teach the model to extract the meaning of a sentence, we base our approach on the notion of grammatical construction: the mapping between a sentence's form and its meaning [22]. Constructions are an intermediate level of meaning between the smaller constituents of a sentence and the full sentence itself. Based on the cue competition hypothesis of Bates et al. [23] we make the assumption that 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 morphemes [14]. In our model (see Figure 1), this mapping corresponds to filling in the Semantic Words (SWs) of a sentence into the different slots (the thematic roles) of a basic event structure that could be expressed in a predicate form like *action(object, location)*. This predicate representation enables us to integrate it into the representation of actions in a robotic architecture.

As can be seen in Figure 1, the system processes inputs as follows: from a sentence (i.e. sequence of words) as input (left) the model outputs (right) a command that can be performed by a robot (i.e. thematic roles for each Semantic Word). Before entering the neural network, words are preprocessed, transforming the sentence into a grammatical construction: Semantic Words, i.e. nouns and verbs that have to be assigned a thematic role, are replaced by the *SW* symbol; Infrequent function words (IWs) are replaced by the *&* symbol. The processing of the grammatical construction is sequential: words are given one at a time, and the final thematic roles for each SW is read-out at the end of the sentence. Only the necessary outputs are shown in the figure for this example. In contrast to previous recurrent neural models [7, 9, 6, 10], the proposed model processes grammatical constructions, not sentences, thus permitting to bind a virtually unlimited number of sentences based only on a small training corpus, and enabling to process future sentences with currently unknown semantic words. Therefore, it is suited for modelling developmental language acquisition.

2.3 What to do with unknown symbols?

Children should be able to (at least partially) understand sentences with some unknown words and to extract the meaning of these new words from their environment [19]. Such a capability enables children to understand that in the sentence "The cat shombles the mouse" *shombles* is probably a verb and to potentially maps its meaning from the context. Even if some words are useless to understand the core meaning of a sentence, for instance "Could you & put some water in the cup?", with *&* symbolizing an unknown word (*e.g.* "please"), the child could still understand what is asked and perform the corresponding action. This ability to (partially) understand a sentence with unknown words is probably crucial (1) for the ability of children to bootstrap the language understanding process, and (2) to quickly learn new words and infer their meaning from the context. On the application side, when interacting with a robot through language, speech recognition will be more robust when the number of words that must be recognized is reduced [24]. That is why we propose to include a new kind of input symbol (&) to deal with infrequent words (see subsection 3.3).

3 Methods and experiments

3.1 Bilingual experiment

Our goal here is to see whether a neural network with no imposed structure (a random reservoir) could learn to process both English and French sentences and to provide the corresponding action commands that could be performed by a robot. For the experiments we use the same set of parameters in order to demonstrate that it is not necessary to tune the parameters for each language.

3.2 Natural language material

Corpora were obtained by asking naive users (agnostic about how the system works) to watch several actions in a video and give the commands corresponding to the motor actions performed, as if they wanted a robot to perform the same action. The video used is available online with the first experiments we did with robots [15]. Five users were recruited for each language, each user provided 38 commands: for each language there is a total of 190 sentences. The English corpus is a subset of the one used in [15]. A selection of sentences is provided in Table 1. For instance, for the "Action order" sentences, one can see that the order of actions to be performed does not necessarily correspond to the semantic word order in the sentence. The particular function of the FW "after" is difficult to get for the model because it appears in the middle of both sentences even if the actions to be performed are reversed. Note also that some sentences provided by users are grammatically incorrect (see Table 1). Each corpus is made of grammatical constructions, not sentences: this means that all the SWs, nouns and verbs, that should be attributed a role have been replaced by a common Semantic Word symbol "SW" in the corpus. In this way, we prevent the network to learn semantic information form nouns and verbs. Several sentences may then be recoded with the same grammatical constructions. The ratios of unique grammatical constructions in the corpora are: 0.410 (78/190) for French (FR), 0.616 (117/190) for English (EN) and 0.513 (195/380) for bilingual (FR+EN) corpora.

3.3 Infrequent symbols category

As mentioned in subsection 2.3 it is important, for a child or a robot, to be able to deal with unknown words. Out-of-vocabulary (OOV) words are a general problem in Natural Language Parsing [25]. In comparison to the previous approach developed [15] we implemented an additional method that replaces most infrequent words in the corpus. We used a threshold θ ($\theta = 7$, see subsection 3.5) that defines the limit under which a Function Word (FW) is considered infrequent and replaced by the Infrequent Word (IW) symbol "&". The preprocessing was performed on the whole dataset before performing the simulations. This new method enables us to process unknown words, which is a desirable property for online interaction when the model is implemented in a robotic platform. However, there is no a priori insight that would state whether this infrequent word replacement will enhance or decrease the generalization performances of the neural network model.

Table 1: Some sentence examples from the noisy English corpus.

TYPE	SENTENCE EXAMPLE
Sequence of actions	touch the circle **after** having pushed the cross to the left
	put the cross on the left side and **after** grasp the circle
Implicit reference to verb	**move** the circle to the left **then the cross to the middle**
Implicit reference to verb and object	**put** first the triangle on the middle **and after on the left**
"Crossed reference"	***push*** **the triangle** and *the circle* ***on the middle***
Repeated action	hit **twice** the blue circle
	grasp the circle **two times**
Unlikely action	put the cross to the right and **do a u-turn**
Particular FW	put **both** the circle and the cross to the right

3.4 Implementation details

We use one shot offline learning to get the optimal output weights in order to make generalization performance comparisons between the bilingual and monolingual corpora. The teacher signals for training the read-out layer are provided during the whole sentence: the rationale for that is that a child or a robot has just performed an action and the caregiver (the teacher) describes the actions that have just been performed. Thus the teacher signal is already available when the sentence is provided to the system. As shown previously [13], this provides the nice property of having the read-out units predicting the thematic roles during the sentence; see Figure 2.

The input W_{in} and recurrent W matrices are randomized following these distributions: values are taken with equiprobability in the set $\{-1, 1\}$ for W_{in}, and with a normal distribution with 0 mean and 1 variance for W. Both matrices have a sparsity of 0.1, i.e. only 10% of the connections are non-zero. After random initialization the input matrix W_{in} is scaled with a scalar value called *input scaling* (IS), and the absolute maximum eigenvalue of the recurrent matrix W is scaled by the *spectral radius* (SR) value. All hyper-parameters are described in section 3.5. As can be seen in Figure 1, the inputs consist of a localist representation of the Function Words (different for each language) and in addition the final dot, the IW symbol "&" and the "*SW*" symbol. Thus, the input dimensions for the different corpus are: 31 (28 + 3) for the French (FR) one, 32 (29 + 3) for the English (EN) one and 60 (57 + 3) for the bilingual (FR+EN) one. The total output dimension is 48 (8 SW * 3 roles * 2 actions): we set the maximum number of SW to 8 in this experiment.

3.5 Hyper-parameters

In the experiment we use a reservoir of 500 units in order to keep the simulation to be trained in a few seconds on a basic machine (without GPU computations), and running (after training) in less than a second: thus if used within a robotic platform it enables real-time interaction with the user. A few hyper-parameters were optimized using the *hyperopt* python toolbox [26], namely the *spectral radius* (SR), the *input scaling* (IS), the *leak rate* α and the threshold θ under which Infrequent Words are replaced. A set of rounded parameters were then chosen from the parameter space region leading to good performance: SR=1, IS=0.03, α=0.2 and θ=7. Actually θ could have been set to any value between 7 and 10 because there was no Function Word (FW) with this number of occurrences: this threshold appears to be a natural limit in the density distribution of FW occurrences. Note that for the *spectral radius* we disregarded the upper limit "advised" by the Echo State Property [1], namely 1, when we performed the hyper-parameter search, because as we are using leaky neurons the effective *spectral radius* is different from the "static" one [27].

3.6 Evaluation

In order to evaluate the performance, we record the activity of the read-out layer at the last time step, which is when the final dot is given as input. We first discard the activations that are below a threshold of 0.5. For sentences that do not contain the maximal number of SW (*i.e.* 8) we discard the remaining outputs because no SW in the input sentence could be linked to them: e.g. if there is only four SWs in the sentence, we discard outputs concerning SWs 5 to 8. Unit activations of

discarded SW outputs represent predictions about SWs that will never occur (see Figure 2). Finally, if there is several possible roles for a particular SW we do a winner-take-all and keep the role unit with the highest activation.

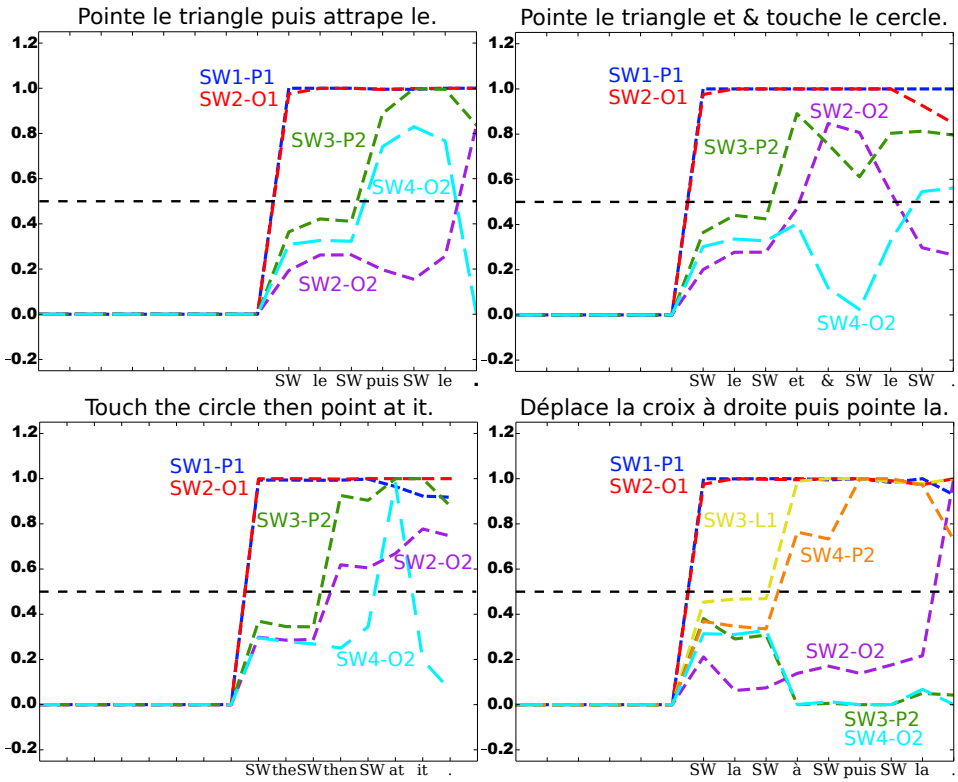

Figure 2: Examples of read-out units activations for different sentences.

4 Results

We start by providing a quantitative analysis (generalization capabilities) of the model for the different corpora, and then we perform a qualitative analysis by examining the output activity of the model for particular sentences.

4.1 Quantitative analysis

First we analyse the generalization errors and standard deviation for a 10-fold cross validation averaged over 50 different network instances. Since there are several thematic roles to be recognized in each sentence, the full meaning of a sentence is correct only if all roles are correct; if one role or more is incorrect the sentence is regarded as recognized incorrectly (*i.e.* sentence error of 1). We provide here the means and standard deviations of the *sentence errors*: 0.158 (+/-0.012) for the FR corpus, 0.214 (+/-0.022) for the EN corpus, and 0.206 (+/-0.013) for the FR+EN corpus. The EN corpus seems to incorporate some slightly more complex sentences than the FR corpus and has less redundant grammatical constructions than the French one: this probably explains the higher generalization error for the EN corpus. Even if results are not directly comparable, the new result for the English corpus outperforms the previous performance of 24.2% obtained[1] in [15].

It is remarkable that the performance for the FR+EN corpus (0.206) is not very impaired compared to the average error of the corpora processed separately (0.186).

[1]Results in [15] were obtained by taking the best of an ensemble of ten reservoirs with twice the number of units (1000 instead of 500) and leave-one-out CV (instead of 10-fold CV). Moreover the training corpus was two times larger.

4.2 Qualitative analysis

In this subsection we use the same instances for the input and reservoir weight matrices, only the read-out weight matrices are different (due to learning on different corpora). We analysed the read-out layer activity for the French and English training corpora and selected some interesting sentences (see Figure 2). For clarity and to limit the number of curves per plot only relevant output units have been kept, others have been discarded. The dotted line indicates the decision boundary threshold for the final extraction of thematic roles. In Figure 2 read-out units activations (i.e. outputs of the model) can be seen for four sentences: three in French[2] and one in English. Sentences before preprocessing are shown on top of each plot; corresponding grammatical constructions processed by the reservoir are shown on the x-axis. The top left plot shows activations for a grammatical construction that was both in the training and testing set of the given cross-validation. For trained grammatical constructions the output activations show online probabilistic estimations for the different roles based on the statistics of the training set. The three remaining plots were taken from grammatical constructions that were not in training set but which generalized correctly. For instance, we can see in the top right plot of Figure 2 that the model is able to generalize to the correct roles even in the presence of an infrequent word (IW symbolized by &). In all plots of the figure, two output roles units are active near the maximum value (*i.e.* 1) since the beginning of the sentence: *SW1-Predicate-1st_action* (SW1-P1; blue curve) unit and *SW2-Predicate-1st_action* (SW2-O1; red curve) unit. This is because for most sentences, the first two SWs are the *predicate* and *object* of the first action; *i.e.* the order of actions is not reverse by words like *before* or *after*.

We choose to focus more on the French language since this is the first time we demonstrate generalization capabilities with this language, and also because it has two interesting properties that English does not have: the words *le* or *la* (*the* in English) are gender specific, and they could be determiners or pronouns. We can see both functions in this sentence: *"Pointe **le** triangle puis attrape **le**."* ("*Point the triangle then catch it.*"): the first *le* is a determiner, and the second one is a pronoun referring to the *triangle*. As we can see in the top left plot of Figure 2, at the time step after the second occurrence of *le* (i.e. at the final dot) there is a particular "re-analysis" of the sentence because this occurrence of *le* is a pronoun, which implies that the object of the second action (O2) is not a potential semantic word (SW4) that could have followed *le*, the O2 is rather the same one as the first action, namely the SW in position 2 (SW2) in the sentence. That is why the activity of the output unit *SW4-Object-2nd_action* (SW4-O2, the unit that binds O2 with SW4; cyan curve), goes down to zero and the activity of the output unit *SW2-Object-2nd_action* (SW2-O2; purple curve) goes up to one. On the contrary, in the top right of Figure 2 the input of the last SW (SW4) confirms the determiner function of *le*, thus the activity of the unit SW4-O2 (cyan curve) increases above the threshold, and the activity of SW2-O2 (purple curve) goes down. The input of the infrequent word symbol "&" seems to "perturb" the on-going predictions compared to the top left plot: these activities may not reflect the statistics of the training corpus since the occurrence of "&" at this precise point in the sequence makes this sequence unique and produces a reservoir state that was not used during training. The bottom right plot of Figure 2 shows similar outputs as the top left plot, but for a sentence containing a location for the first action (SW3-L1; yellow curve). One can see how the following output activities of units SW3-P2 and SW4-O2 are modified in consequence. In the bottom left of Figure 2 is the readout activity for an equivalent English sentence of the French sentence in top left plot. One can see that the unit activations are similar until the last word in the sentence: *it* and *le* respectively. Some differences of units activation could be explained by the fact that the English sentence was not in the training set. This means that we have a model that is able to represent on-going sequences of words in two different languages with the same predictions of roles.

5 Discussion

A neuro-inspired model based on Reservoir Computing that processes sentences word by word with no prior semantic knowledge was used. The only assumptions are that the system is able to distinguish semantic words (SW) from function words (FW), because SWs are related to objects or actions the child or the robot already knows. Nouns and verbs were not distinguished in this SW

[2]Translation of sentences: (top left) "*Point the triangle and catch it.*"; (top right) "*Point the triangle and & touch the circle.*"; (bottom right) "*Move the cross to the left then point at it.*".

abstract symbolic category. The model processed grammatical constructions instead of sentences [22, 19] based on noisy natural language corpora produced by human users.

For the first time we showed that our architecture could process three different kinds of symbolic inputs: function words, semantic words and infrequent words. Moreover we showed that this architecture is able to process and generalize over the French language newly provided. Furthermore, we outperformed previous results obtained in [15] with the English corpus. Generalization performance on these noisy corpora, produced by users, is interesting considering the small corpus we used, about 200 sentences for each language: 84.2% for the French, 78.6% for the English and 79.4% for the bilingual corpora respectively. These figures indicate the percentage of sentences that have all their roles fully recognized, which means that the thematic role performance is higher.

When used in a robotic or other platforms, if a sentence is recognized partially (*i.e.* few roles in the sentence are incorrect), the system may recover based on further contextual and semantic post-processing, thereby reducing the number of sentences not recognized. In fact, these good performances could be enhanced by post-processing because as it has been shown in [13] that most of the sentences not fully recognized have only one or few erroneous roles. Moreover we showed here for the first time that the system was able to understand grammatical constructions that had infrequent unknown function words. This is not only interesting from the point of view of language acquisition [19] but also from the application side because it provides a natural way of dealing with the out-of-vocabulary (OOV) words problem [25]. In further work we will explore distributional encoding of semantic words based only on the context available to the system. For instance we could use *word2vec* representation [28] which is based on huge corpora. In this language acquisition perspective, an issue would be to create such representations with small corpora where not much context information is available, and moreover where this context information is available incrementally.

This bilingual experiment shows that the chosen architecture has interesting properties for multilingual processing. The network was able to learn and generalize without an important drop-off in performance. What is surprising is to have such a high performance for a fixed reservoir size even with an input dimension that doubled in size for the bilingual corpus, compared to the monolingual experiments: the bilingual corpus contained twice more function words than the French one or the English one. Moreover, the reservoir state spaces explored for each of the corpora are quite different, due to the different sets of inputs, nevertheless the linear regression performed by the read-out layer is still able to combine state spaces produced by very different inputs towards the same output roles. Deeper analysis of the reservoir states and read-out weights may provide some more explanation why the bilingual model is working better than one would expect. It is possible that the model benefits from the regularities of the syntax similarities between French and English. Further work is needed to compare this bilingual neural model to other models [29] and to analyse which insights it can give on bilingual language acquisition and second language acquisition [30] (if using an incremental learning). For instance, would a bilingual model that builds its own self-organized input word representations (shared by the two languages) be able to benefit from both languages and generalize better than a monolingual model? It would be also interesting to evaluate the ability of the current model to process grammatical constructions that have parts in French and parts in English.

Acknowledgments

This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme: EchoRob project (PIEF-GA-2013-627156). Authors are grateful to Cornelius Weber and Dennis Hamester for their very useful and interesting feedback.

References

[1] Jaeger, H. (2001) 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.

[2] Dominey, P. F. (1995) Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning. Biological Cybernetics, 73, pp. 265–274.

[3] Lukosevicius, M., & Jaeger, H. (2009) Reservoir computing approaches to recurrent neural network training. Computer Science Review 3: 127–149.

[4] Rigotti, M. et al. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451), 585–590.

[5] Maass W., Natschlger T., & Makram H. (2003) A Model for Real-Time Computation in Generic Neural Microcircuits. In Proc. of NIPS 2003, 213–220.

[6] Elman J (1990) Finding structure in time. Cognitive Science 14: 179–211.

[7] Miikkulainen, R. (1996) Subsymbolic case-role analysis of sentences with embedded clauses. Cognitive Sci 20: 47–73.

[8] Wermter, S., Arevian, G., & Panchev, C. (2000) Meaning Spotting and Robustness of Recurrent Networks. In Proc. of IJCNN, pp. III-433-438. Como, Italy.

[9] Tong, M. H. et al. (2007) Learning grammatical structure with Echo State Networks. Neural networks 20: 424–432.

[10] Frank, S. L. (2006). Strong systematicity in sentence processing by an Echo State Network. In Proc. of ICANN 2006, pp. 505–514.

[11] Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. Neuron, 50(6), 963-974.

[12] Hannun, A. et al. (2014) Deep Speech: Scaling up end-to-end speech recognition. arXiv:1412.5567

[13] Hinaut, X., & Dominey, P. F. (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.

[14] Dominey, P. F., Hoen, M., & Inui, T. (2006) A neurolinguistic model of grammatical construction processing. Journal of Cognitive Neuroscience, 18(12):2088–2107.

[15] Hinaut, X. et al. (2014) Exploring the Acquisition and Production of Grammatical Constructions Through Human-Robot Interaction. Frontiers in NeuroRobotics 8:16.

[16] Dominey, P. F., & Boucher, J. D. (2005). Developmental stages of perception and language acquisition in a perceptually grounded robot. Cognitive Systems Research, 6(3), 243-259.

[17] Hinaut, X., & Wermter, S. (2014) An Incremental Approach to Language Acquisition: Thematic Role Assignment with Echo State Networks. In Wermter, S. et al., Proc. of ICANN 2014, pp. 33-40.

[18] Hinaut, X. et al. (2015) Cortico-Striatal Response Selection in Sentence Production: Insights from neural network simulation with Reservoir Computing. Brain and Language, vol. 150, Nov. 2015, pp. 54–68.

[19] Tomasello M (2003) Constructing a language: A usage based approach to language acquisition. Cambridge, MA: Harvard University Press. 388 p.

[20] Hinaut, X. et al. (2015) Humanoidly Speaking – How the Nao humanoid robot can learn the name of objects and interact with them through common speech. Video, IJCAI 2015. http://bit.ly/humanoidly-speaking

[21] Marcus, G. F. et al. (1999). Rule learning by seven-month-old infants. Science 283, 77–80.

[22] Goldberg, A. (1995) Constructions: A Construction Grammar Approach to Argument Structure; Fauconnier G, Lakoff G, Sweetser E, editors. Chicago: University of Chicago Press. 265 p.

[23] Bates, E. et al. (1982) Functional constraints on sentence processing: a cross-linguistic study. Cognition 11: 245–299.

[24] Twiefel, J. et al. (2014) Improving Domain-independent Cloud-based Speech Recognition with Domain-dependent Phonetic Post-processing. In Brodley C.E. et al. (eds.). Proc. of AAAI 2014, pp. 1529-1535.

[25] Jurafsky, D., and Martin, J. H. (2009) Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson International, 2nd edition.

[26] Bergstra, J., Yamins, D., & Cox., D. D. (2013) Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. In SciPy13.

[27] Jaeger, H. et al. (2007). Optimization and applications of echo state networks with leaky-integrator neurons. Neural Networks, 20(3), 335-352.

[28] Mikolov, T. et al. (2013) Distributed Representations of Words and Phrases and their Compositionality. In Proc. of NIPS 2013.

[29] Frank, S. L. (2014). Modelling reading times in bilingual sentence comprehension. In P. Bello et al. (Eds.), Proc. of CogSci 2014, pp. 1860–1861.

[30] Berens, M. S.; Kovelman, I. & Petitto, L.-A. (2013) Should Bilingual Children Learn Reading in Two Languages at the Same Time or in Sequence? Bilingual Research Journal, 36, pp. 35–60.