ABSTRACT

This paper emphasizes the importance of the question of word-order variation in connectionist language modeling. More precisely, it investigates whether recurrent networks can integrate various linguistic constraints to process variable word-order languages. This paper reports three experiments on the understanding of spoken French that suggest that recurrent architectures could apply to the understanding of variable spoken languages.

Keywords: neural networks, language modeling, word-order, speech understanding.

1. INTRODUCTION: THE PROBLEM OF WORD ORDER VARIATION

The use of connectionist methods for speech technologies has so far concerned essentially speech recognition, where the hybridization of Hidden Markov Models (HMM) with neural networks has led to significant results [1]. In spite of some encouraging studies [2,3], connectionist language modeling for spoken dialog systems remains unfortunately widely unsolved.

It was claimed for a long time that connectionist architectures, which are efficient for static pattern classification, were not able to deal with symbol sequences. Now, experimental results as well as theoretical studies [5] have demonstrated the ability of recurrent networks [6] to process word sequences. Actually, two main problems seem to prevent still the use of connectionist models:

- their time and data-consuming nature during the learning phase,
- their relative difficulties to model complex linguistic structures.

Beside these generic problems, this paper questions another possible limitation of neural networks that is too often forgotten in human language technologies (HLT): the question of the processing of variable word-order languages.

Indeed, most HLT research concerns the English language. Now, English presents a specific structure which is distinguished by a very rigid word order that compensates for a poor morphology. This feature has led to the predominance of stochastic language models — among which N-grams — that strongly emphasize word order to describe linguistic structure. English is however quite atypical. The understanding of most languages implies indeed a variety of morphological, syntactic or semantic cues that thereby allow a more flexible syntax. Variable word-order languages (Russian, Finnish, Japanese, German, Spanish...) differ thus from non-flexible ones (English, Chinese). It is therefore not surprising that English-inspired approaches apply hardly to variable word-order languages [8,9].

Considering this problem, this paper investigates whether recurrent networks can integrate various linguistic constraints to process variable word-order languages. It focuses on the understanding of spontaneous spoken French, which should be considered as a variable word-order language to a certain extent [10]. This paper reports three experiments that intends to show that recurrent networks can apply to the understanding of variable spoken languages.

2. EXPERIMENT

2.1. Task: speech understanding

Generally speaking, information speech dialog systems are based on the same architecture (figure 1):

![Figure 1. Architecture of speech understanding systems](image-url)
- First, the speech recognizer processes the speech signal and provides a sequence of words that should correspond to the spoken sentence.
- Then, the understanding process provides a semantic representation that corresponds to the literal meaning of the sentence.
- Finally, the dialog manager translates the semantic structure into a SQL query. The answer of the database is then provided to the user (speech synthesis, visual display,...).

Because of the ungrammatical nature (hesitations, repetitions, repairs,...) of spontaneous speech, the semantic representation cannot be obtained through a complete parsing of the sentence. On the contrary, the understanding process considers only some concepts in the sentence. Concepts are the smallest units of meaning that are relevant to the task [11]. Consider for instance:

(1) I'd like to book a seat on a direct flight from Nantes-Atlantique airport to Budapest.

The conceptual segmentation for this sentence should be:

- **request_category**: book
- **seat-num**: a (seat)
- **itinerary-stop**: direct
- **itinerary-departure**: (from) Nantes
- **itinerary-arrival**: (to) Budapest

Each concept will then correspond to a specific semantic case in the semantic representation [4]: Category, Num-seat, Stop, From, To:

- `<category = book>`
- `<num-seat = 1>`
- `<stop = NO>`
- `<from = Nantes>`
- `<to = Budapest>`

By opposition with ATIS systems, our study is not domain specific. Our experiments assess the assignation of three general semantic cases, as originally defined by Fillmore [12]: Agent, Action and Patient. Consider the following French sentence:

(2) Le chateau Marie le regarde

*The castle Mary is watching it*

The expected semantic structure of this sample is:

- `<Agent = Marie>`
- `<Action = regarder>`
- `<Patient = chateau>`

Given simple French sentences with a flexible word-order structure, the task of the network is to assign the concepts of the sentence to the correct semantic cases.

### 2.2. Computer model: simple recurrent net

Standard multilayer perceptrons [13] have proven their ability to solve multiple constraints for various kinds of problems. These connectionist architectures tend however to treat inputs as global patterns. They present consequentially difficulties to process input sequences. By opposition, recurrent networks deal naturally with such sequences. Because recurrent connections endow networks with internal dynamics, recurrent nets can be used to model sequential domains such as natural language processing [2,6,14].

![Three-layer architecture of the understanding networks used](image)

The experiments are carried out on recurrent nets with the same three-layer architecture: input, hidden and output, with recurrent links on the hidden layer (figure 2). The networks used differ only in the size of the hidden layer. This architecture corresponds to Elman's Simple Recurrent Network (SRN [6]) and uses ordinarily backpropagation as learning procedure.

**Input** — The sentence is provided sequentially, e.g. word by word. The coding is local: each word of the lexicon is represented by a specific neuron of the input layer.

**Output** — The output layer consists actually of three sub-layers that corresponds to the three semantic cases. The coding is also local, each lexical neuron in a sub-layer standing for a possible filler of the semantic case.

**Context** — The context layer is fed by a copy of the hidden layer activations at the preceding time step.

The experiments are carried out on recurrent nets with the same three-layer architecture: input, hidden and output, with recurrent links on the hidden layer (figure 2). The networks used differ only in the size of the hidden layer. This architecture corresponds to Elman's Simple Recurrent Network (SRN [6]) and uses ordinarily backpropagation as learning procedure.

**Input** — The sentence is provided sequentially, e.g. word by word. The coding is local: each word of the lexicon is represented by a specific neuron of the input layer.

**Output** — The output layer consists actually of three sub-layers that corresponds to the three semantic cases. The coding is also local, each lexical neuron in a sub-layer standing for a possible filler of the semantic case.

**Context** — The context layer is fed by a copy of the hidden layer activations at the preceding time step.

At the end of the sequential processing of the sentence, the output layer shows the probable cases of the words (or conceptual segments).

All neurons use a sigmoid function, initial link weights are random values in the range [-0.2; +0.2]. Several training rates has been used during the experiments. When not mentioned, the training rate is assigned a default value of 0.1.

### 2.3. Corpora and experiments

The aim of this research is only to assess the ability of connectionist architectures to deal with variable word-order languages. This is why these experiments have been carried out on rather small corpora. Although a
larger study would be necessary to represent really the problem space, it is not the aim of this preliminary experiment. Yet, our artificial corpora presents enough word-order variation to answer our question. Every corpus is designed with a lexicon of 18 words:

- 5 transitive verbs (V),
- 10 nouns: respectively 5 humans (H) and 5 objects (O),
- 2 pronouns (French clitics: *le* and *il*)
- 1 determiner (French definite article *le*)

We designed three corpora which correspond to a progressive increase in complexity. All sentences are syntactically and semantically well-formed and present word-order variations which are typical features of spoken French.

**Experiment 1 (HVO corpus)** — In this corpus, each word fulfills always the same semantic case (human H = Agent; object O = Patient, verb V = Action) regardless of word order. Consider the following two examples:

(3) Marie mange le pain
    Agent Action Patient
    [Mary *is eating the bread*]

(4) Le pain elle le mange Marie
    Patient Action Patient
    *[The bread she is eating it Mary]*

Here, the case assignment depends only on the semantic affinities of each lexical concept (Marie, mange, pain) whatever their order is. The understanding process requires therefore only lexical-semantic constraints: no consideration of the linguistic context is needed [16].

**Experiment 2 (HVH corpus)** — Each sentence contains here two human concepts, which fill respectively the Agent and Patient cases. For instance:

(5) Pierre regarde Jean
    Agent Action Patient
    [Peter *is watching John*]

(6) Jean regarde Pierre
    Agent Action Patient
    [John *is watching Peter*]

Since the same concept can fill two different cases in two different sentences, the consideration of lexical-semantic constraints is useless in this experiment. On the contrary, the system must consider the linguistic context to succeed [16]. More precisely, word-order is here the most relevant cue: in the active voice¹, the first human concept fills preferentially the Agent case.

**Experiment 3 (HVH+HVO)** — This last experiment is achieved on the union of the two previous corpora: human may fill Agent or Patient cases whereas the objects should only correspond to Patient roles. As a result, the system must associate both lexical-semantic and word-order constraints to succeed.

It is important to note that word-order is variable in this experiment too: the network must learn word-order constraints in a robust way in spite of this flexibility.

In these experiments, determiners are not associated with a semantic case since they are ignored by conceptual segmentation. Likewise, pronouns are not considered in the output semantic representation. In these fairly simple corpora, the referent of the pronoun is indeed clearly expressed in the sentence and plays the same role as its pronominal clitic.

### 3. RESULTS

#### 2.1. Error measures

The corpora are classically divided into training and test corpora. Although a classical RMS error rate is used for training, this global measure is not very informative for the assessment of semantic case assignation. We have thus defined two other error rates:

- **Selectivity** — it measures the rate (%) of words that are assigned to a correct semantic case and present an output activation $S$ which is inside a tolerance range of 0.2 (e.g. the activation of the other words are all lower than $S - 0.2$).

- **Accuracy** — it measures the rate (%) of words that are correctly assigned, without considering any tolerance range.

#### 2.2. Experimental results

Table 1 summarizes the results of the three experiments. Several networks with different training rates and different hidden layer size have been tested every time. This table reports the results that were obtained with the best networks with a training rate of 0.1.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>hidden</th>
<th>epochs</th>
<th>accuracy</th>
<th>selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>10</td>
<td>5</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>(2)</td>
<td>12</td>
<td>155</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>110</td>
<td>96%</td>
<td>94%</td>
</tr>
<tr>
<td>(3)</td>
<td>16</td>
<td>74</td>
<td>97%</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>121</td>
<td>94%</td>
<td>93%</td>
</tr>
</tbody>
</table>

¹ The passive voice requires the consideration of both lexical-semantic and contextual constraints. This situation is investigated in the third experiment.

In the first experiment, a network with 10 hidden neurons met a complete success on the test corpus. This result is not surprising, since the HVO corpus only requires the consideration of lexical-semantic constraints: the network behaves here as a simple
classifier (one word = one case).

On the contrary, the network must consider the linguistic context (word-order constraints) to understand the HVH sentences. A recurrent network with 12 hidden neurons succeeded completely on this second experiment. Once again, this result is not surprising: recurrent networks can parse word sequences with a rigid syntax [5].

The most interesting result is provided by the last experiment: it shows that recurrent nets can associate various linguistic constraints to understand flexible word-order sentences. Two nets, with respectively 15 and 16 hidden neurons, present indeed an interesting accuracy and selectivity (error rates lower than 7%). Besides, a perfect behavior can be obtained easily by decreasing the training rate to a value of 0.03 (Table 2).

Table 2: experimental results (hidden = size of the hidden layer; epochs = number of training cycles); training rates = 0.03.

<table>
<thead>
<tr>
<th>Exp</th>
<th>hidden</th>
<th>epochs</th>
<th>accuracy</th>
<th>selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2)</td>
<td>12</td>
<td>210</td>
<td>99%</td>
<td>98%</td>
</tr>
<tr>
<td>(3)</td>
<td>15</td>
<td>175</td>
<td>97%</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>88</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

4. DISCUSSION AND CONCLUSION

In spite of the limited extent of this experiment, these results suggest that recurrent architectures offer a good technique to learn a diversity of linguistic constraints, including word-order, that enable the understanding of a variable word-order language such as spoken French. Besides, these results support our previous experiments on Japanese [15].

Nevertheless, it should be stressed that the increase of the task complexity from the first to the last experiment led to an inescapable growth of the hidden layer (10 to 16 hidden neurons) and learning time (5 to several hundreds epochs). Now, the sentences of the third experiment are still very simple! The question of how connectionist techniques will scale up to more complex problems (large lexicon, complex embedded syntactic structures, irregular structures of spontaneous speech, ...), in order to apply to spoken dialog systems, remains therefore widely open. The achievement of some complex modular neural architectures [7], as well hybrid approaches, are certainly promising answers to this problem.

In conclusion, this work tried to emphasize the important question of word-order variation in language modeling. Our results show that a fairly simple recurrent net trained with common backpropagation is able to overcome this problem on a restricted task. This preliminary result is, in our opinion, a good indication of the suitability of a — more complex — connectionist modeling of flexible languages.

4. REFERENCES

[8] Covington M. (1990), A dependancy parser for variable-order languages, research Report AI-1990-01, University of Georgia, USA.
[16] St John M.F. (1990), Learning and applying contextual constraints in sentence interpretation, Artificial Intelligence, 46 (1-2).