ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

Distributed Representation of Vocabularies in the RECONTRA Neural Translator *

G.A. Casañ, M.A. Castaño

Dpto. de Informática. Universidad Jaume I de Castellón. Spain.
e-mail: casan@tomba.uji.es castano@inf.uji.es
http://nuvol.uji.es/~castano

10.21437/Eurospeech.1999-532

Abstract

A simple neural translator called RECONTRA (REcurrent CONnectionist TRAnslator) has recently shown to successfully approach simple text-to-text limited-domain Machine Translation tasks. In this approach the vocabularies involved in the translations were represented according to (simple and clear) local codifications. However, in order to deal with large vocabularies local representations would led to networks with an excessive number of connections to be trained. Consequently, distributed representations of both source and target vocabularies are required. This paper studies appropriate types of distributed codifications to represent large vocabularies in the RECONTRA translator.

1. INTRODUCTION

In contrast to traditional *Knowledge-Based* Machine Translation (MT) systems, translation models can be automatically built from (large enough) training data, resulting in *Example-Based* (EB) approaches with lower development costs. This EB techniques have recently led to successful limited-domain applications, as it is shown in [1] [2] [11]. In this direction, *Neural Networks* (so-called *Connectionist Models*) can be considered as an encouraging approach to MT. However, only a few connectionist MT systems have been developed in the literature, such as those proposed in [7] and [12]. Nevertheless, the connectionist system in [7] employs static topologies which are not appropriate to approach a real MT task and in [12] the connectionist system separately approaches the syntactic and semantic features associated to a language, resulting in a complex translation model.

In contrast to these approaches, a simple and dynamic EB neural translator for text-to-text, limited-domain applications called RECONTRA (REcurrent CONnectionist TRAnslator) was recently introduced in [3]. It carries out directly the translation between both the language to be translated and the translated language (with no intermediate items) and consequently neither syntactic nor semantic parsing are required. This connectionist translator was tested in preliminary experiments [3] on a simple pseudo-natural task with small vocabularies [6]. Encouragingly accurate translations were obtained in these experiments employing local representations of both source and target vocabularies.

However, the size of the neural nets required for dealing with large vocabularies in the translations (and consequently, the learning time) can be prohibitive when local codifications are employed. To this end, more compact (that is, distributed) representations of the input and output vocabularies are required, which are explored in this paper.

The paper is organized as follows: Section 2 briefly describes the basic connectionist architecture of the RECONTRA translator, as well as the main features of the distributed codifications employed for the vocabularies. Section 3 presents the MT task adopted for the experimentation. The performances obtained by the RECONTRA are later reported in Section 4. Finally, Section 5 discusses the conclusions of the experimental process.

2. THE RECONTRA TRANSLATOR

2.1. Network architecture

The neural architecture adopted for the RECONTRA translator was a network with simple recurrences which was proposed by Elman [5]. In addition, the preceding and the following contexts of the input signal were also presented to the net. In this way, the information about past and future input events can be reinforced and, consequently, the performance of the model can be increased. Figure 1 illustrates the resulting neural topology.

* Work partially funded by the Spanish CICYT (project TIC-97-0745-CO2)

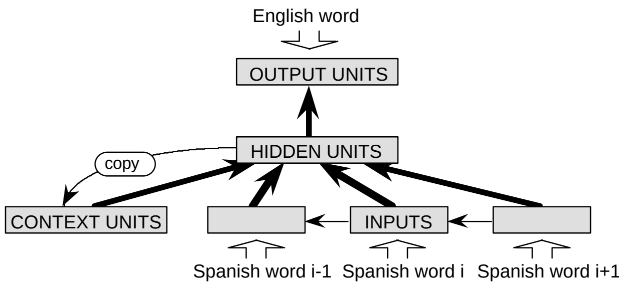

Figure 1. The RECONTRA translator.

With regard to the running of the connectionist architecture, the words of the sentence to be translated were sequentially presented to the input of the network, while the net had to provide the successive words of the corresponding translated sentence. An additional output word was included to mark the end of the translated sentence. It should be noted that the context of the delayed input words should be wide enough for RECONTRA to have enough information at the input, in order to translate the appropriate output word.

2.2. Codifications of the source and target lexicons

The information to be processed or managed by a connectionist model can be locally or distributedly represented (coded) [10]. In *local codifications* every unit (neuron) is dedicated to the representation of one of the possible items to be represented (words in a MT task). On the other hand, *distributed codifications* assume that each of these items is represented by an activation pattern which is distributed along all (or some of) the units. Two main types of distributed representations can be considered: *symbolic distributed codifications*, in which every unit corresponds to a microfeature associated to a syntactic or semantic feature (with higher level of abstraction than the set of concepts to be represented); and *subsymbolic distributed codifications*, in which not every unit can be semantically interpreted on its own.

Local representations are quite clear and simple. However, they require a large number of units to code large sets of items and, consequently, a large number of connections to be trained. On the contrary, distributed representations are not so clear and are more complicated to implement, although they consume low resources per information unit. Consequently, they allow more compact representations and networks with lower number of trainable connections.

Therefore, in order to tackle MT between languages which involve large vocabularies using the RECONTRA translator, a distributed representation of both source and target vocabularies is required. To this end, different symbolic and subsymbolic distributed representations of the lexicons (described in sections 4.3 and 4.4) were considered in the experiments presented in this paper.

2.3. Training procedure

The RECONTRA translator described above was trained using an on-line version of the *Backward-Error Propagation* algorithm [9]. The forward and backward steps of the algorithm were continuously repeated for every input window of the source sentence until the target value of the output neurons identified the end of the translated sentence. A sigmoid function (0,1) was assumed as the non-linear function and, context activations were initialized to 0.5 at the beginning of every input-output pair. The choice of the learning rate and momentum was carried out inside the unitary bidimensional space which they defined, by analyzing the residual mean squared error of a network trained for 10 random presentations of the complete learning corpus (10 epochs). Training continued for the learning rate and momentum which led to the lowest mean squared error. And the learning process stopped when a certain established criterion was verified.

With regard to the translated message provided by the RECONTRA, the net continuously generated (at each time cycle) output activations. These activations can be interpreted by assuming that the net supplied the output word for which the pre-established codification in the target lexicon was nearest to the corresponding output activations.

3. THE EXPERIMENTAL MACHINE TRANSLATION TASK

The task chosen in this paper to study appropriate distributed representations for the RECONTRA translator was a simple pseudo-natural task called *Miniature Language Acquisition Machine Translation* (MLA-MT). It had been originally introduced in [6] as an interdisciplinary task and was adequately reformulated later as a MT Task [4]. This task consisted in translating (from Spanish into English and vice versa) descriptions of simple visual scenes as well as the addition and removal of objects to or from a scene. Since many of the sentences are worded by using the passive voice, the degree of asynchrony between the Spanish and the corresponding English sentence is substantial.

The sizes of the Spanish and English vocabularies involved were 29 and 26 words, respectively. In order to (slightly) increase the sizes of these vocabularies, an extension of the MLA-MT task with 50 Spanish words and 38 English words was also considered in the experiments presented in the paper. Some examples of this extension of the task are shown in Figure 2.

4. EXPERIMENTAL RESULTS

First, the MLA-MT task with the smaller lexicons was approached in order to analyze different global features of the distributed codifications. Later, different symbolic and subsymbolic distributed representations of the larger vocabularies were studied. In both experiments, only Spanish-to-English translations were taken into account. All connectionist experiments presented in the paper were trained and tested using the SNNS neural simulator [13].

4.1. Training and recognition data

The corpus adopted in the considered task were sets of text-to-text pairs which consisted in a sentence in the source language and the corresponding translation in the target language. For both versions of the MLA-MT task, a learning sample of 3,000 pairs was adopted to train the RECONTRA translators. The learnt models were evaluated later on a different test set, which consisted of 2,000 sentences (for each of the tasks) that were generated independently of those employed for training.

Spanish: un cuadrado diminuto y claro y un círculo claro tocan a una pirámide y un número pequeño
English: a tiny light square and a light circle touch a pyramid and a little number

Spanish: se elimina el círculo grande que está encima de la línea mediana y oscura y del triángulo negro
English: the large circle which is above the medium dark line and the black triangle is removed

Figure 2. Spanish-English pairs of sentences from the MLA-MT task.

4.2. Criterion assessing correct translations

A *source test sentence* supplied to a connectionist architecture was considered to be correctly translated if the output provided by the model exactly coincided with the expected translation for this source sentence. In order to determine *word accuracy*, the obtained and expected translations corresponding to every source sentence in the test sample were compared using a conventional Edit-Distance (Dynamic Programming) procedure [8]. In this way, the number of insertions, deletions and substitutions errors were obtained. The word accuracies reported here correspond to the ratio of the total number of non-errors with respect to the total number of edit (total error+correct) operations.

4.3. Experiment on global features of the codifications

In a first experiment, different *global features of the distributed codifications* were analyzed. To this end, both the Spanish and the English vocabularies of the MLA-MT task with the smaller lexicons were coded on 15 distributed boolean units. Random or pseudo-random codifications were assigned taking into account the two following issues whose effect on the translation rates was studied later:

- Employing similar codifications for those words in the vocabulary which appeared in similar syntactic contexts (for example, *square* and *triangle*).
- Employing the same codification for the source word to be translated and the corresponding translated target word (as far as possible).

According to these codifications, different RECONTRA translators with 15 input units, 15 outputs, 14 delayed input words (with 6 and 7 words for the left and the right context, respectively) and 140 hidden neurons were trained. Learning stopped after 500 random presentations of the training set. The learnt neural translators were then evaluated on the (same) test set of 2,000 sentences. Table 1 shows the sentence and word translation accuracies achieved. The performances which had been previously obtained [3] on a RECONTRA translator with analogous features (for the hidden layer and the delayed input window), but using local codifications of the lexicons are also reported in Table 1.

The translation performances shown in this Table suggested the employment of similar (distributed) representations with those words in the vocabulary which appeared in similar syntactic contexts. These experiments also showed that the learning convergence was increased when the same codification for the source word to be translated and the corresponding translated target word was adopted. On the other hand, the translation rates obtained using these distributed representations of the lexicons were quite similar to those achieved using local representations.

FEATURES OF CODIFICATIONS			TRANSLATION RATES	
Local	**Distributed. Similar for similar contexts**	**Distributed. Identical for translations**	**SAR**	**WAR**
No	No	No	98.6%	87.2%
No	Yes	No	99.5%	95.5%
No	Yes	Yes	99.8%	97.8%
Yes	----	-----	99.9%	98.8%

Table 1. Sentence and word accuracy translation rates (SAR and WAR) for the MLA-MT task with the smaller lexicons, using different features for their codifications. Distributed representations were coded in 15 units.

4.4. Experiment on symbolic and subsymbolic distributed representations

Taking the best global features obtained in the previous section using distributed representations into account, different symbolic and subsymbolic representations of the vocabularies were studied in a second experiment. With regard to symbolic representations, specific units of the codifications were dedicated to distinguish the gender and/or number and/or syntactic functions (verb, noun or adjective) of the words. In order to study better the effect of the codifications on the translations, the more complex MLA-MT task (that with the larger lexicons) was approached at this time. In addition, more compact distributed representations of the lexicons (coded in 10 boolean units) were adopted.

According to these codifications, RECONTRA translators with 10 input units, 10 output units, 14 (6+1+7) delayed input words and 140 hidden neurons were used in this experiment. Learning was carried out for 500 random presentations of the training set. The learnt neural translators were then evaluated on the test set of 2,000 sentences. The translation rates achieved are shown in Table 2.

The results reported in the table revealed that symbolic codifications led to better translation accuracies than pure subsymbolic representations (the first experiment of Table 2), as it was expected. In addition, the experiment showed that performances were increased by employing symbolic codifications to differentiate the greatest number of morphological features and syntactic functions of the words as possible.

REPRESENTATIONS INCLUDING			**TRANSLATION RATES**	
Number	**Gender**	**Syntactic function**	**SAR**	**WAR**
No	No	No	80.5%	21.6%
Yes	No	No	83.0%	26.7%
Yes	Yes	No	92.7%	61.2%
Yes	Yes	Yes	96.7%	79.0%

Table 2. Sentence and word accuracy rates (SAR and WAR) for the MLA-MT task with the larger lexicons, using different subsymbolic and symbolic representations of the lexicons (coded in 10 units).

5. CONCLUSIONS AND FUTURE WORK

A preliminary study of different distributed codifications of the vocabularies involved in a Machine Translation task approached by the RECONTRA translator [3] has been presented in this paper. The results obtained revealed that symbolic representations led to more accurate translations than purely subsymbolic representations. In addition, performances were increased by employing simple a priori knowledge of the syntax, lexicons and translations.

Based on the performances presented in this paper, more studies on more compact distributed codifications are still required to approach translation tasks which involve large lexicons. The grouping of some words and words sequences into categories (which would be automatically learnt) could also help to decrease the size of the RECONTRA translators. Future work dealing with more complex limited domain translations should also be considered.

6. REFERENCES

[1] J.C. Amengual, J.M. Benedí, K. Beulen, F. Casacuberta, M.A. Castaño, A. Castellanos, D. Llorens, A. Marzal, H. Ney, F. Prat, E. Vidal, J.M. Vilar. *Speech Translation Based on Automatically Trainable Finite-State Models.* Procs. of the 5th European Conference on Speech Communication and Technology (EUROSPEECH-97), vol. 1, pp. 91–91, Rhodes, Greece. 1997.

[2] P.F. Brown, S.A. Della Prieta, R.L. Mercer. *The Mathematics of Statistical Machine Translation: Parameter Estimation.* Computational Linguistics, vol. 19, no. 2, pp. 263–311. 1993.

[3] M.A. Castaño, F. Casacuberta. *A Connectionist Approach to Machine Translation.* Procs. of the 5th European Conference on Speech Communications and Technology (EUROSPEECH-97), vol. 1, pp. 91–94, Rhodes, Greece. 1997.

[4] A. Castellanos, I. Galiano, E. Vidal. *Applications of OSTIA to Machine Translation Tasks*. In "Lecture notes in Computer Science–Lecture Notes in Artificial Intelligence: Grammatical Inference and Applications", vol. 862, pp. 93–105, R.C. Carrasco and J. Oncina (Eds.), Springer-Verlag. 1994.

[5] J.L. Elman. *Finding Structure in Time.* Cognitive Science, vol. 2, no. 4, pp. 279–311. 1990.

[6] J.A. Feldman, G. Lakoff, A. Stolcke, S.H. Weber. *Miniature Language Acquisition: A Touchstone for Cognitive Science.* Technical Report no. TR-90-0009, Int. Computer Science Institute, Berkeley, California. 1990.

[7] N. Koncar, G. Guthire. *A Natural Language Translation Neural Network.* Procs. of the Int. Conf. on New Methods in Language Processing, pp. 71–77, Manchester, UK. 1994.

[8] A. Marzal, E. Vidal. *Computation of Normalized Edit Distance and Applications.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, num. 9. 1993.

[9] D.E. Rumelhart, G. Hinton, R. Williams. *Learning Sequential Structure in Simple Recurrent Networks.* In “Parallel distributed processing: Experiments in the microstructure of cognition”, vol. 1. Rumelhart D.E., McClelland J.L. and the PDP Research Group (Eds.), MIT Press. Cambridge. 1986.

[10] N.E. Sharkey. *Connectionist Representations for Natural Language: Old and New*. Procs. of the VI SEPLN, Donostia, Spain. 1990.

[11] S. Vogel, H. Ney, C. Tillmann. *HMN-Based Word Alignment in Statistical Translation.* Procs. Of the Int. Conference on Computational Linguistics, pp. 836–841, Copenhagen, Denmark. 1996.

[12] A. Waibel, A.J. Jain, A.E. McNair, H. Saito, A.G. Hauptmann, J. Tebelskis. *JANUS: A Speech-to-Speech Translation System using Connectionist and Symbolic Processing Strategies*. Procs. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP-91), vol. 2, pp. 793–796. 1991

[13] A. Zell et al. *SNNS: Stuttgart Neural Network Simulator. User manual, Version 4.1.* Technical Report no. 6195, Institute for Parallel and Distributed High Performance Systems, University of Stuttgart. 1995.