Connectionist Modeling of Linguistic Quantifiers

Rohana K. Rajapakse[1], Angelo Cangelosi[1], Kenny R. Coventry[2], Steve Newstead[2], Alison Bacon[2]

[1] Adaptive Behaviour & Cognition Group, School of Computing Comms and Electronics,
[2] Centre for Thinking & Language, School of Psychology
University of Plymouth, Drake Circus,
PL4 8AA, Plymouth, UK
{rrajaakse, acangelosi kcoventry, snewstead, abacon}
@plymouth.ac.uk
http://www.tech.plym.ac.uk/soc/research/ABC

Abstract. This paper presents a new connectionist model of the grounding of linguistic quantifiers in perception that takes into consideration the contextual factors affecting the use of vague quantifiers. A preliminary validation of the model is presented through the training and testing of the model with experimental data on the rating of quantifiers. The model is able to perform the "psychological" counting of objects (fish) in visual scenes and to select the quantifier that best describes the scene, as in psychological experiments.

1 Introduction

The selection and use of vague linguistic quantifiers, such as *a few, few, several, many, lots of* is greatly influenced by the communicative aim of the speaker. For example, in the two sentences "A few people went to the cinema. They liked the movie" and "Few people went to the cinema. They preferred the restaurant", the selection of the quantifier *a few* vs. *few* indicates differences in the focus of attention signaled by the quantifier. *A few* is chosen to put emphasis on those that actually went to the cinema; *Few*, instead, shifts the attention to the people that did not go. Understanding the meaning of such terms is important as they are among the set of closed class terms which are generally regarded as having the role of acting as organizing structure for further conceptual material. Although some researchers have proposed that quantifiers can be mapped directly to numbers on a scale (e.g. [3]), there is compelling evidence that the comprehension and production of quantifiers can be affected by a range of factors which go beyond the number of objects present. These include contextual factors (e.g. the relative size of the objects involved in the scene, the expected frequency of those objects based on prior experience, the functionality present in the scene - e.g. [15]) and communicative factors (e.g. to control the pattern of inference as in the example above - see [14]).

The existence of these contextual and communicative effects highlights the fact that language cannot be treated as an abstract, self-referential symbolic system. On the contrary, the understanding of language strongly depends on its grounding in the individual's interaction with the world. The importance of grounding language in perception, action and cognition has recently received substantial support both from cognitive psychology experiments (e.g. [2,9,11]) and computational models (e.g. [4,12]). In particular, connectionist systems are being increasingly used as the basis for modeling grounding [6]. They permit a straightforward way to link perceptual stimuli to internal categorical representations, upon which semantic and linguistic representations are anchored. The use of connectionist components within embodied cognitive systems, such as agents and robots, also permit the link between language and sensorimotor stimuli.

Various connectionist models of quantification and number learning have been proposed, although none has directly focuses on linguistic quantifiers. Two main directions of research can be identified in the neural network literature. A first set of models focuses on learning number sequences. They are able to process the objects in the input scene sequentially to reproduce the sequences and/or compute distances between two numbers. For example, Rodriguez et al. [18] modeled the learning of sequences of letters. The identification of the correct number of presentations of a first letter permits the prediction of the presentation of a second letter. Ma and Hirai [13] simulated the production of the number word sequence as observed from children.

The second approach includes models that learn to identify the number of objects in the input visual scene. Dehaene & Changeux [10] developed a numerosity detection system comprised of three modules: an input retina, an intermediate topological map of object locations, and a map of numerosity detectors). This was able to replicate the distance effect in counting, by which performance improves with increasing numerical distance between two discriminated quantities. Peterson and Simon [16] presented a connectionist model of subitizing, the phenomenon by which subjects appear to produce immediate quantification judgments (normally up to 4 objects) without the need to do sequential counting. The simulation results suggested that subitizing emerges through experience, rather than being the result of a limited representational capacity of the architecture. Similar results were found in the model by Ahmad et al. [1], which uses Kohonen's SOM networks. This model also used a recurrent backpropagation network for articulating the numerosity of individual objects in the collection and a static backpropagation network for the next object pointing task. This model is distinct in the way it is trained to count by decomposing the counting task into that of number word update\storage from the next-object pointing task.

In this paper we present a new connectionist model of the grounding of linguistic quantifiers in perception that takes into consideration the contextual factors affecting the use of vague quantifiers. The model is able to perform both the "psychological" counting of objects (fish) in visual scenes and to select the quantifier that best describes the scene. A preliminary validation of the model will be presented through the training and testing of the model with experimental data on the rating of quantifiers.

2 Architecture of the Model

The computational model consists of a hybrid artificial vision-connectionist architecture (Figure 1). The model has four main modules: (1) Vision Module, (2) Compression Networks, (3) Quantification Network, and (4) Dual-Route Network. This architecture is partially based on a previous model on the grounding of spatial language [5,8]. The overall idea is to ground the connectionist and linguistic representation of quantification judgments directly in input visual stimuli.

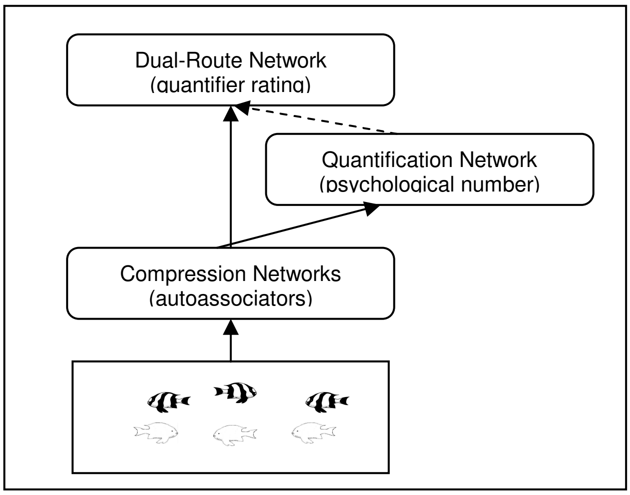

Fig. 1. Modular architecture of the model

The vision module processes visual scenes involving varying quantities of two types of objects: striped fish and white fish. It uses a series of Ullman-type vision routines to identify the constituent objects in the scene. The input to the Vision module consists of static images with the two kinds of fish. The system must pay attention to striped fish, whilst white fish are only used as distracters (or vice versa). The input images are processed at a variety of spatial scales and resolutions for object features yielding a visual buffer. The processing of each image results in two retinotopically organized arrays of 30x40 activations (one per fish type). The output of the vision module represents data of isotropic receptive fields.

The Compression Networks is needed to convert the output data from the vision module into compressed neural representation of the input scene. This is to reduce the complexity of the vision module output. Two separate auto-associative networks are used, respectively for each of the object types in the scene (stripy fish and white fish). Both networks have 1200 input and output units, and 30 hidden units. The activation values of the hidden units will be utilized by the following networks to make quantification and linguistic judgments. The compression network for each type of fish will learn to autoassociate all the stimili with varying number of fish.

The Quantification Network is a feedforward multi-layer perceptron trained to reproduce the quantification judgments of the number of fish made by subjects during experiments on psychological counting [7]. Previous simulations only focused on the counting of the striped fish, those the subjects are asked to consider when making quantification decisions [17]. In the current, updated version, the same network has to count both sets of fish. The network has 60 input units (30 per compressed fish type), 50 hidden nodes, and 2 output nodes. Each output node has a modified activation function that produce activation values in the range 0 to 20, to include the actual range of 0 to 18 stripy fish used in the stimulus set.

The fourth module consists of a Dual-Route neural network. This architecture combines visual and linguistic information for both linguistic production and comprehension tasks [6]. This is the core linguistic component of the model, as it integrates visual and linguistic knowledge to produce a description of the visual scene. The network receives, in input, information on the scene through the activation values of the compression (and quantification) networks' hidden units. It will then produce, in output, judgments regarding the appropriate ratings for the quantifier terms describing the visual scene. The activation values of the linguistic output nodes correspond to rating values given by subjects for the five quantifiers considered: a few, few, several, many, lots.

In the current version of the model, only the hidden representations of the compression networks are used. The dual-route network will require 60 input visual units and 5 input linguistic nodes, one for each quantifier. The linguistic units correspond to the 5 vague quantifiers *a few*, *few*, *several*, *many* and *lots*. The 60 visual nodes corresponded to the 30 hidden units of the two compression networks (of stripy fish and white fish). The output layer has the same number and type of units as those in the input layer. After training with data from psycholinguistic experiments, the network will be capable of producing two different outputs: (1) acceptability ratings for quantifiers given only the vision inputs (language production) and (2) imaginary output pictures, given only a description of the scene in terms of quantifiers (comprehension). Results of the simulation on the production route (predicted ratings for the quantifiers) will be compared to the actual ratings of experiments with human subjects.

3 Simulation Results

The model uses as input stimuli to the vision module 216 scenes used in quantification experiments with subjects [7]. In the experimental design, the number of fish (of both types) is varied from zero to 18 fish per scene, with incremental steps of 3. The fish are arranged in random locations, with equal spacing between them (two levels of inter-fish distances are used). In addition, two levels of grouping of fish from the same type (grouped or mixed) are used, with another factor regarding the two levels of the position of the grouped stimuli (top or bottom of the image).

The 216 scenes are first presented to the vision module. Its output is then used to train the autoassociative networks of the Compression module. For the training, 195 scenes are used as training stimuli and 21 as generalization test stimuli. The learning rate is 0.01 and momentum 0.8. The networks are trained for 2000 epochs. The autoassociative network is able to learn both training stimuli (average RMS error of 0.019 and 0.014 for stripy and fish data respectively) and novel generalisation stimuli (average RMS error of 0.080 and 0.070 for stripy and white fish data). This permits a significant reduction of complexity of the 1200 output values of the visual module into only 30 compressed hidden activation values.

Results with the Quantification network [17] have shown that networks are able to reproduce the production of “psychological numbers” produced by subjects [7] and that they use similar mechanisms in producing such judgements.

New simulations have focused on the training of the dual route network. The same stimulus set is used to collect data on the use of the five vague quantifiers *a few*, *few*, *several*, *many* and *lots*. In this psycholinguistic experiment, subjects are asked to rate the use of vague quantifiers by using a 9-point Likert scale for the appropriateness of sentences like “There are a few stripy fish”. The average rating data of the subjects are converted into presentation frequencies for the training of the dual route network, as in Cangelosi et al. [6]. For the training, 195 scenes are used as training stimuli and 21 as generalization test stimuli. The learning rate is 0.001 and momentum 0.8. The networks are trained for 1000 epochs. The autoassociative network is able to learn both training stimuli (average RMS error of 0.051) and novel generalisation stimuli (average RMS error of 0.084).

Simulation No.	**Learning rate**	**Random novel set**	**Hidden nodes**	**Training error**	**Novel error**
1	0.001	TrnTst1	30	0.055079	0.086295
2	0.001	TrnTst2	30	0.055002	0.070183
3	0.001	TrnTst3	30	0.043332	0.095066
			average	0.051138	0.083848

4 Conclusions

This paper reports some preliminary simulation results on a new artificial vision/connectionist model of the grounding of linguistic quantifiers in perception. Experimental data are used for the training and testing of the dual-route neural network that selects the linguistic quantifiers that best describe the scene.

Future simulation experiments will focus on the analyses of the specific effects that the various contextual factors manipulated in the experiment (e.g. number, spacing and grouping of fish) produce in the selection of vague quantifiers. In addition, future simulations will address the contribution of explicit numerical judgments, as in quantification experiments [7], in the use of linguistic quantifiers.

The model presented here proposes a new approach to the study of linguistic quantifiers, by grounding quantification judgments and vague quantifier directly in perception. From the semantic point of view, these terms have the virtue of relating in some way to visual scenes being described. Hence, it will be possible to offer more precise semantic definitions of these, as opposed to many other expressions, because the definitions can be grounded in perceptual representations.

Acknowledgements

This research was supported by the UK Engineering and Physical Research Sciences Council (EPSRC Grant GR/S26569)

References

1 Ahmad, K., Casey, M.C., Bale, T: (2002). Connectionist Simulation of Quantification Skills. Connection Science , 14 (2002), 1739-1754
2 Barsalou, L. W.: Perceptual symbol systems. Behavioral and Brain Sciences, 22 (1999), 577-660.
3 Bass, B.M, Cascio, W.F., O'Connor, E.J.: Magnitude estimations of expressions of frequency and amount. Journal of Applied Psychology, 59 (1974), 313-320.
4 Cangelosi, A., Bugmann, G., Borisyuk, R. (eds.): Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop. World Scientific, Singapore (2005)
5 Cangelosi, A., Coventry, K.R., Rajapakse, R., Bacon, A., Newstead S.N.: Grounding language into perception: A connectionist model of spatial terms and vague quantifiers. In: [4].
6 Cangelosi, A., Greco, A., Harnad S.: From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. Connection Science, 12 (2000) 143-162
7 Coventry, K.R., Cangelosi, A., Newstead, S.N., Bacon, A., Rajapakse R.: Vague quantifiers and visual attention: Grounding number in perception. XXVII Annual Meeting of the Cognitive Science Society (2005)
8 Coventry, K. R., Cangelosi, A., Rajapakse, R., Bacon, A., Newstead, S., Joyce, D., Richards, L. V:. Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. In C. Freksa et al. (eds.), Spatial Cognition, Volume IV. Reasoning, action and interaction, pp 98-110. Lecture notes in Computer Science. Springer (2005)
9 Coventry, K.R., Garrod, S.C.: Saying, Seeing and Acting. The Psychological Semantics of Spatial Prepositions. Essays in Cognitive Psychology Series. Psychology Press. Hove and New York (2004)
10 Dehaene, S., Changeux, J.P.: Development of Elementary Numerical Abilities: A Neuronal Model. Journal of Cognitive Neuroscience, 5 (1993), 390-407
11 Glenberg, A.M., Kaschak, M.: Grounding language in action. Psychonomic Bulletin and Review, 9 (2002) 558-565
12 Harnad S. (1990). The symbol grounding problem. Physica D, 42: 335-346
13 Ma, Q., Hirai, Y.: Modeling the Acquisition of Counting with an Associative Network. Biological Cybernetics, 61 (1989) 271-278
14 Moxey, L.M., Sanford, A.J.: Communicating Quantities. A Psychological Perspective. Lawrence Erlbaum Associates; Hove, East Sussex (1993)
15 Newstead, S.N., Coventry, K.R.: The role of expectancy and functionality in the interpretation of quantifiers. European Journal of Cognitive Psychology, 12(2) (2000), 243–259
16 Peterson, S.A., Simon, T.J.: Computational Evidence for the Subitizing Phenomenon as an Emergent Property of the Human Cognitive Architecture. Cognitive Science, 24 (2000) 93-122
17 Rajapakse, R.K., Cangelosi, A., Coventry, K., Newstead, S., Bacon A.: Grounding linguistic quantifiers in perception: Experiments on numerosity judgments. 2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics. April 21-23, 2005, Poznañ, Poland (submitted).
18 Rodriguez, P., Wiles, J., Elman, J.L.: A Recurrent Neural Network that Learns to Count. Connection Science, 11 (1999), 5-40