

## Grounding Linguistic Quantifiers in Perception: Experiments on Numerosity Judgments

Rajapakse<sup>1</sup>, R.K., Cangelosi<sup>1</sup>, A., Coventry<sup>2</sup> K., Newstead<sup>2</sup> S. and Bacon<sup>2</sup>, A.

<sup>1</sup>School of Computing, Communications and Electronics

<sup>2</sup>School of Psychology

University of Plymouth

{rrajapakse, acangelosi, kcoventry, snewstead, abacon}@plymouth.ac.uk

### Abstract

The paper presents a new computational model for the grounding of numerosity judgments and of the use of linguistic quantifiers. The model consists of a hybrid, artificial vision-connectionist architecture. Preliminary simulation experiments show that the part of the model trained to judge “psychological number” uses some of the same factors known to play a major role in the production of quantification judgments in human subjects. This supports the ongoing development of a psychologically-plausible model of linguistic quantifiers which uses the contextual factors such as object properties and their functionality.

### Introduction

Talking about numbers of objects in a visual scene often involves the use of descriptions of quantity which are vague. Furthermore, quantifiers, whether they be of number (e.g., many), amount (e.g., much), or time/frequency (e.g., often) pervade natural language, and therefore constitute an essential part of the lexicon for the child to acquire, and consequently for integration into NL systems. An understanding of quantifiers is often largely couched in terms of the notion that quantifiers refer to points on a scale. In its most extreme form, the temptation is to treat quantifiers in terms of a quantifier-to-number mapping (e.g., Bass, Cascio & O’Connor, 1984; Reyna, 1981). In computational terms, a scene can be parsed for the number of entities present, and the mapping between the number and the quantifier associated with the appropriate point on the scale can be easily achieved. However, 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, including 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, and the need to control the pattern of inference of those involved in the communication (see Moxey & Sanford, 1993 for a review). Given the existence of these context effects, some researchers have argued that the number of objects has a minimal impact on the comprehension of quantifiers (e.g., Moxey & Sanford, 1993).

In the computational model we are developing for quantifiers, we draw a distinction between the actual number of objects in a scene presented for description, and “psychological number”. In experimental work reported elsewhere (Coventry et al., under review), we have shown that judgements about the appropriateness of quantifiers to describe given pictures (e.g., of a number of white fish, with other striped fish in the scene see Figure 1) correlate directly with judgements of the number of objects people think are in the pictures when asked to respond under time pressure. In other words, the number of objects people think is present on a quick scan is predictive of how appropriate a quantifier is to describe the number of objects in a scene. Put simply, quantifier judgements are grounded in perception, consonant with recent theories of grounding language in perception (e.g., Glenberg & Kaschak, 2002; Barsalou, 1999; Coventry & Garrod, 2004). In this paper we provide an overview of our computational model for natural language quantifiers – focusing on how psychological number is generated in the model, mapping onto the number of objects participants think is present in a scene to be described.

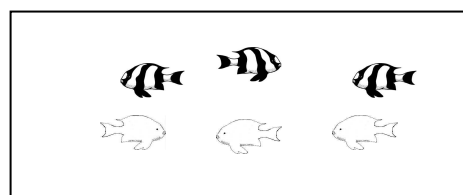


Figure 1 – An example of a scene used

### Visual Object Enumeration

An extensive body of psychological literature has been reported on the human knowledge of numbers, number systems and numerosity judgements. These studies have revealed at least three strategies used by the brain: a fast and accurate processing of small groups of four or fewer items in almost constant response time (often called subitizing), a slow process of serial counting of more than five (less than 9) objects, and a more error prone estimation process for larger groups of objects

In press, 2<sup>nd</sup> Language & Technology Conference, Poland 2005

(>9) (e.g. Trick & Pylyshyn 1993, Mandler & Shebo 1982). Based on these findings, the focus of the past research has been mainly on the distinction between subitizing and counting phenomena, using fewer than 10 objects, while a few publications has reported experiments on a larger range (~20). An on-going debate among those who have proposed many theories to explain this dichotomy is whether the brain uses completely different cognitive processes or this is a result of a single process but due to different levels of difficulty along a continuum of difficulty in processing (Piazza et al. 2002).

### Connectionist Models of Enumeration

Connectionist models, based on the use of artificial neural networks for cognitive modelling, have been employed to study quantification and enumeration. These models primarily focus on the understanding of numerical processing in the human brain, such as the different aspects of visual object enumeration tasks including subitizing, counting and seriation. For example, some models have been directly inspired by the psychophysical findings which have identified the involvement of various brain areas for the representation of different forms of abstract cardinal numbers (written, auditory, visual) such as in the triple code model of Dehaene (1997).

Two main directions of research in connectionist counting modelling can be identified. The first modelling approach focuses on the tasks of learning number sequences and sequential series. These use networks that process the input stimuli sequentially, using a short term memory for keeping intermediate states of counting. They are capable of reproducing a learned sequence and/or computing distance between two numbers. For example, Rodriguez et al. (1999) employed a simple recurrent network to learn strings consisting of 'a's followed by the same number of 'b's. The idea was to train the network to count the number of 'a's presented, in order to predict the number of 'b's. This work demonstrates the concept of subitizing as a form of preverbal counting, as counting in this model is not based upon a number word sequence, instead on an abstract understanding of the number of objects presented sequentially.

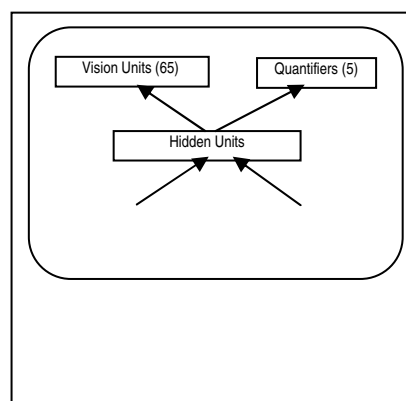
The second approach has instead focused on the tasks of counting of the number of objects in input visual scenes. For example, Dehaene & Changeux's numerosity detection system (1993) uses a model based on three modules: an input retina, an intermediate topological map of object locations, and a map of

numerosity detectors. Simulations on numerosity discrimination showed the distance effect (performance improved with increasing numerical distance between two discriminated numerosities) as well as Fechner's law for numbers (for equal distance discriminating numerosities, performance is better with small numerosities than large ones). Ahmad et al. (2002) employed a sequential multi-net system (SSUBSYST) for object enumeration. Their subitizing system (SSUBSYST) used two major subsystems, (1) a magnitude representation based on Kohonen's SOM, bidirectionally connected to a verbal SOM and (2) a mapping subsystem to process a visual scene to map the information in a scale and translation invariant manner onto the magnitude representation SOM. This model has been further extended for counting by including a recurrent backpropagation network for articulating the numerosity of individual objects and a static backpropagation network for the task of pointing at the next object.

### The Computational Model

The computational model we are developing aims at grounding numerosity judgments and the use of linguistic quantifiers in perception. It consists of a hybrid, artificial vision-connectionist architecture (Figure 2) with three main modules: (1) Vision Module, (2) Compression and Quantification Networks and (3) Dual-Route Network. This architecture is broadly based on a previous model on the grounding of spatial language (Cangelosi et al., in press; Coventry et al., in press).

The Vision module uses a series of Ullman-type vision processing routines (Joyce et al. 2002) to identify the constituent objects of a visual scene. The input to the Vision module consists of static images with two kinds of fish: stripy and white (see Figure 2). The network must count stripy 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 arrays of 30x40 activations, representing retinotopically organised and isotropic receptive fields for each type of fish.



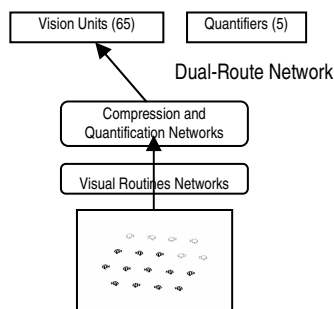


Figure 2 – Dual-route network with input from Visual Routines and the Compression Networks modules.

The Compression and Quantification Networks module utilises the output information from the vision module to produce a compressed neural representation of the input scene. In addition, this module performs quantification judgments. For the compression task, 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 25 hidden units. The quantification network is trained to reproduce the quantification judgments of the number of stripy fish made by subjects during experiments on psychological counting (Coventry et al., submitted). This feedforward network has 1200 input units, 15 hidden nodes, and 1 output node. The output node has a modified activation function that allows activation values in the range 0 to 20 (to include the actual range of 0 to 18 stripy fish).

The third module consists of a Dual-Route neural network. This architecture combines visual and linguistic information for both linguistic production and comprehension tasks (Plunkett et al., 1992; Cangelosi et al., 2000). 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, a judgment 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).

The dual-route network will require 65 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 65 visual nodes corresponded to the 25 hidden units of the two compression networks (of stripy fish and

white fish) and the 15 hidden nodes of the quantification network. 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 picture, 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.

### Experimental Data for Network Training

The training of a psychologically-plausible model of quantification, able to reproduce the performance of human subjects, requires data from two psychological experiments. The first experiment asks participants to make numerosity judgements on the number of stripy fish presented in each visual scene. Estimated numbers will be used to train the numerosity judgment network. The second set of experiments asked subjects to rate the acceptability level of linguistic quantifiers. The data are then used to train the dual route network.

Experiments on numerosity judgements require subjects to look at visual scenes of stripy and white fish and to respond with a number corresponding to the subjective assessment on the number of stripy fish present. Stimuli are presented in a computer screen only for a short time (500 milliseconds), to avoid exact counting, and the response time is also recorded (onset of typing of number). In the experimental design, the number of fish (of both types) is varied (six levels) from zero to 18 fish per scene, with incremental steps of 3 fish (the zero fish condition only applies to the white fish distracters). The fish are arranged in random locations, but with equal spacing between them. Two levels of inter-fish distances (spacing) 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). This constitutes an experimental design of 228 conditions/scenes in total (i.e. experimental design of 6x6x2x3 plus an additional 12 scenes with striped fish alone).

The same stimuli are used to collect data on the use of the 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-

In press, 2<sup>nd</sup> Language & Technology Conference, Poland 2005

point Likert scale for the appropriateness of sentences like “There are *a few* stripy fish”. Rating data will be converted into presentation frequencies for the training of the dual route network.

### Simulation Experiments

Two modules of the model have already been developed and tested. The vision module has been fully implemented and the retinotopical encodings of all images have been created. The Compression and Quantification module has also been implemented. Here we report data on the compression autoassociative networks (Table 1) and the quantification experiments (Table 2).

Results with the autoassociative networks show that the model is able to learn both training stimuli (average RSM error of 0.04) and novel generalisation stimuli (average RSM error of 0.081). This permits a significant reduction of complexity of the 1200 output values of the visual module into only 25 compressed hidden activation values. Results with the quantification networks also are good. Networks have an average training error of 0.042 (= 0.02% considering the single output node with activation range 0-20) and generalisation error of 1.56 (=8%).

Sim. No.	Learn. Rate	Hidden Nodes	Trn/Tst patterns	Train. Error	Gen. Error	Train Epocs
1	0.1	15	204/24	0.08	0.089	4000
2	0.1	25	204/24	0.032	0.083	4000
3	0.1	30	204/24	0.028	0.077	4000
4	0.1	35	204/24	0.019	0.072	1350

Table 1: Details of training and testing of compression autoassociative network for stripy fish scenes.

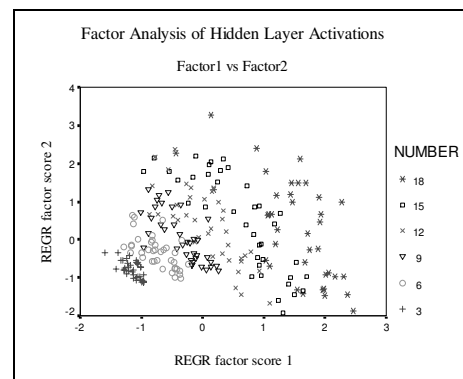
Sim. No.	Lrn. Rate	Hidden Nodes	Trn/Tst patterns	Trn. Error	Gen. Error	Train. Epocs
1	0.01	15	206/22	0.119	1.681	3000
2	0.001	15	206/22	0.017	1.555	3000
3	0.0005	15	206/22	0.0001	1.422	3000
4	0.01	15	206/22	0.115	2.143	3000
5	0.001	15	206/22	0.0004	1.356	3000
6	0.0005	15	206/22	0.0001	1.201	3000

Table 2: Details of training and testing of numerosity judgment networks. Please note that the error range refers to a single output node with activation range of 0 to 20.

The training of the quantification network provides the data for the final, dual-route network training on the production and rating of linguistic quantifiers. This part of the research is still at a preliminary stage, due to the

ongoing collection of the quantifier rating data. However, the quantification network already provides useful insights for the identification of important factors in psychological quantification. For example, the analysis of the hidden activation of the quantifier network highlights the factors that play a major role in the production of quantification judgments, and compares these to data from psychological experiments.

Preliminary analysis of the hidden layer activations of the network, conducted with principal components factor analysis (PCA), indicate that the networks use both the information on the number of fish and the different spacing between fish (Figure 3). The first factor, that explains 59.43% of variance, clearly groups the hidden representation of scenes by the number of fish in it. The second factor groups stimuli by the two sizes of inter-fish distances, explaining 14.41% of the variance. The relevance of these two mechanisms is consistent with the results of psychological experiments, where both the number and the spacing factors are statistically significant. However, in contrast to the empirical data, the network does not seem to use the information on the grouping of fish (separate groups vs. mixed stripy/white fish). The PCA only shows a marginal effect on the placement of the stripy fish when these are separated by the top/bottom position. A reason for the lack of the effects of grouping in the network hidden data could be explained by the fact that the network only processes the stripy data. As a result, “grouping” does not mean much to the network, except for the cases in which the grouping causes the placement of all the stripy fish in the top or the bottom part of the scene. Further investigation of the grouping effects will therefore require the integration of both stripy and white fish in the same quantification network.



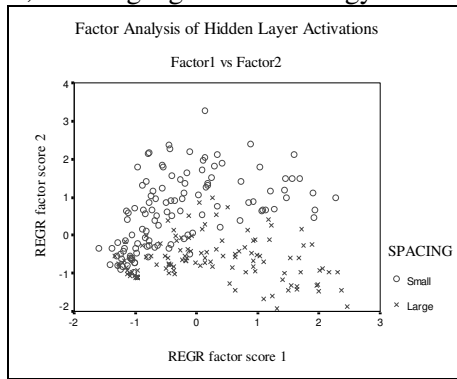


Figure 3: Principal component analyses hidden node activations for the first two factors: NUMBER (top figure) and SPACING (bottom figure)

The data on the implementation and testing of the Vision module and of the Compression and the Quantification networks now permit the further training of the dual-route network and the focus on the model's production of linguistic quantifiers. This is part of the ongoing research, in parallel with further psychological investigations.

Overall, the results of the preliminary simulations support this development of a psychologically-plausible model of linguistic quantifiers which uses contextual factors such as object properties (e.g. spacing, grouping) and their functionality for the generation of quantification judgments and the use of linguistic quantifiers.

In press, 2<sup>nd</sup> Language & Technology Conference, Poland 2005

## References

- Ahmad, K., Casey, M.C. & Bale, T. (2002). Connectionist Simulation of Quantification Skills. *Connection Science* (vol. 14(7), pp. 1739-1754).
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.
- Bass, B.M., Cascio, W.F. & O'Connor, E.J. (1974) Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, 59, 313-320.
- Cangelosi A., Coventry K.R., Rajapakse R., Bacon A. & Newstead S.N. (in press), Grounding language into perception: A connectionist model of spatial terms and vague quantifiers. In A. Cangelosi et al. (eds.), *Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific.
- Cangelosi A., Greco A. & Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2), 143-162.
- Coventry K.R., Cangelosi A., Newstead S.N., Bacon A. & Rajapakse R. (submitted). Vague quantifiers and visual attention. Grounding number in perception.
- Coventry, K. R. & Garrod, S. C. (2004). *Saying, Seeing and Acting. The Psychological Semantics of Spatial Prepositions*. Essays in Cognitive Psychology Series. Psychology Press. Hove and New York.
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. London: Allen Lane, The Penguin Press.
- Dehaene, S. & Changeux, J.P. (1993). Development of Elementary Numerical Abilities: A Neuronal Model. *Journal of Cognitive Neuroscience* (vol. 5(4), pp. 390-407).
- Glenberg, A. M., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9(3), 558-565.
- Joyce D., Richards L., Cangelosi A., Coventry K.R. (2002), Object representation-by-fragments in the visual system: A neurocomputational model. In L. Wang et al. (Eds), *Proceedings of the 9th International Conference on Neural Information Processing (ICONP02)* IEEE Press.
- Mandler, G. & Shebo, B.J. (1982). Subitizing: An Analysis of its Component Processes. *Journal of Experimental Psychology: General* (vol. 111, pp. 1-22).
- Moxey, L. M., & Sanford, A. J. (1993). *Communicating Quantities. A Psychological Perspective*. Lawrence Erlbaum Associates; Hove, East Sussex.
- Piazza, M., Mechelli, A., Butterworth, B. and Price, C.J. (2002). Are Subitizing and Counting Implemented as Separate or Functionally Overlapping Processes?. *NeuroImage* (Vol. 15, pp. 435-446).
- Plunkett, K., Sinha, C., Moller, M.F & Strandsry, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4(3-4), 293-312.
- Rodriguez, P., Wiles, J., Elman, J.L. (1999). A Recurrent Neural Network that Learns to Count. *Connection Science* (vol. 11(1), pp. 5-40).
- Reyna, V. F. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory and Cognition*, 9, 642-650.
- Trick, L.M. & Pylyshyn, Z. (1993). What enumeration studies can show us about spatial attention: Evidence for preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance* (vol. 19, pp. 331-351).