Grounding Natural Language Quantifiers in Visual Attention.

Kenny R. Coventry[1] (kcoventry@plymouth.ac.uk), Angelo Cangelosi[2] (acangelosi@plymouth.ac.uk),
Stephen Newstead[1] (snewstead@plymouth.ac.uk), Alison Bacon[1] (ambacon@plymouth.ac.uk),
and Rohanna Rajapakse[2] (rrajapakse@plymouth.ac.uk).
[1]Centre for Thinking and Language, School of Psychology, & [2]Adaptive Behaviour and Cognitive Group, School of Computing, Communications and Electronics,
University of Plymouth
Devon, PL4 8AA England (UK)

Abstract

The literature on vague quantifiers in English (words like "some", "many", etc.) is replete with demonstrations of context effects. Yet little attention has been paid to the issue of where such effects come from. We explore the possibility that they emanate from a visual attentional bottleneck which limits the accuracy of judgments of number in visual scenes under conditions of time pressure. We present the results of 3 experiments which reveal a range of new context effects on the acceptability of vague quantifiers to describe a number of objects in a visual scene, and show corresponding effects on judgments of number using the same visual scenes under speeded conditions.

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 is 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 (e.g., Hormann, 1983, Newstead & Coventry, 2000), the expected frequency of those objects based on prior experience (e.g., Moxey & Sanford, 1993), the functionality present in the scene (Newstead & Coventry, 2000), and the need to control the pattern of inference of those involved in the communication (e.g., Moxey, Sanford & Dawydiak, 2001). To give an example of one of these context effects (expected frequency in this case), *a few people outside a cinema* is associated with more people than *a few people outside a fire station* (Moxey & Sanford, 1993).

The Origins of Context Effects

Given these myriad context effects, while some have argued that number is critical for quantifier comprehension, others have proposed that quantifiers are not about number at all (Moxey & Sanford, 1993). Yet, clearly as the number of objects increases, the appropriateness of high magnitude quantifiers increases. There are two key issues emerging from this debate. First, the existence of context effects is important in its own right, but discussion of the issue of where such effects come from is noticeably absent from this literature. Why is it that set size, expected frequency, relative size etc. matter? The second issue is intricately bound up with the answer to this question. It has been assumed in the quantifier literature that the actual number of objects in a scene being described is the number used to map onto comprehension of vague quantifiers. Yet there is much evidence that, under time pressure conditions, people do not give very accurate judgments about the number of objects there are in a scene. When a small number of objects is presented in a visual scene, we are able to know the number of objects present almost instantly (an effect called *subitizing* in the number literature). However, when the number of objects increases, we are much less accurate in our judgments of how many objects there are in a visual scene. Hence, we will argue that in the quantifier literature the lack of success of scalar approaches to quantifiers, and the corresponding abandonment of this approach, has been a result of a conflation of number with the actual number of objects being referred to. We will show that the "psychological" numbers returned from visual attentional constraints under time pressure allow number to predict the appropriateness of natural language quantifiers much more successfully. Specifically, in three experiments we uncover several new context effects for vague quantifiers and show that the same effects also occur for number judgments. However, we first briefly overview the literature on object enumeration with a view to predicting these new context effects.

Visual Object Enumeration

There is much evidence that participants do not give very accurate numerosity judgements under time pressure. 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 (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 (>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 have reported experiments on a larger range (~20). An ongoing debate among those who have proposed 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). However, from our point of view it is clear that it takes time to count objects, and therefore it is likely that participants, when producing or comprehending quantifiers, are likely to use estimates of numbers of objects in the scene rather than the actual number of objects present. If this is the case, then one might expect that the estimated number would impact in their judgements about the appropriateness of quantifers rather than the actual number.

If this argument is correct, then we can make some predictions regarding the parameters in a visual scene which may affect both estimates of number and consequently ratings of vague quantifiers to describe the scenes. For the first two experiments reported below, participants were instructed to rate how appropriate sentences were to describe visual scenes. We presented two types of objects in the scene being described: striped fish and white fish. The number of striped fish and white fish is systematically varied, from 3-18 for the fish in focus (fish quantified in the sentence describing the scene), increased in increments of 3, and 0-18 fish for the other fish in the scene (fish not quantified in the sentence describing the scene). We predicted that the number of other fish present in the scene would affect the acceptability of quantifiers to describe the scenes. We also manipulated two other parameters. First, in Experiment 1 we manipulated grouping. Either the fish of a given type were grouped together or mixed together (see Figure 1). We expected that grouping would affect judgments as grouping the fish together makes them easier to count, while mixing them up does not. Second, given existing effects of relative size (Newstead & Coventry, 2000) we also expected that the spacing between objects might affect judgments of quantifiers as spacing effects the overall area the fish cover. This is tested in Experiment 2. Finally in Experiment 3 we presented the scenes involving the same manipulations as those used in Experiments 1 and 2, but as a numerosity judgment task. Participants saw the scenes for 500msec and had to give an estimate for the number of fish present. We show that the results of Experiment 3 mirror the results of Experiments 1 and 2, and therefore that there is a correspondence between number estimates in the scenes being described and context effects for vague quantifiers. In other words, quantifier judgments are grounded in perception, consonant with recent theories of grounding language in perception (e.g., Glenberg, 1997;

Glenberg & Kaschak, 2002; Barsalou, 1999; Coventry & Garrod, 2004; Zwaan, 2004).

Experiment 1

Participants
Twenty undergraduate students participated for payment.

Design and Materials
The task involved participants rating how appropriate sentences of the form *There are [QUANTIFIER] striped/white fish* were to describe given pictures of fish. The quantifiers used were *a few, few, several, many,* and *lots of.* Therefore the terms used included two low and two high magnitude quantifiers, together with a mid-range magnitude quantifier. The scenes used varied the number of striped and white fish present. The fish mentioned in the sentence to be rated will hereafter be calls the *focus* fish, and the other fish not mentioned will be called the *other* fish, but we varied whether the white fish or striped fish were the focus fish. The number of focus fish varied from 3-18 in increments of three, and the number of other fish varied from 0-18, again in increments of three. In this experiment we manipulated a second variable: whether the focus fish were grouped together or mixed in with the other fish. When scenes contained both striped and white fish either the fish were arranged in groups, with the focus fish all at the top or bottom of the picture (but with equal spacing between fish), or alternatively the fish were randomly mixed together (see Figure 1).
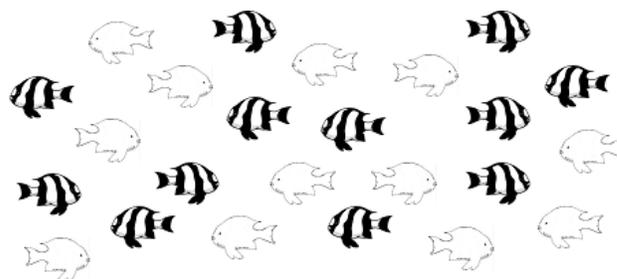

Figure 1. Example of a mixed scene used in Experiment 1.

Procedure
Participants were instructed that they had to rate how appropriate each sentence is to describe each picture. Sentences were presented together under each picture, always in the same order (from low to high magnitude quantifiers). Participants were asked to circle a number on the seven-point rating scale beside each sentence, where 1 = totally inappropriate and 7 = totally appropriate. The order of pictures was randomized.

Results and discussion
In order to analyze the data, we used a four-way within-subjects analyses of variance. The variables were grouping (grouped or mixed) x number of focus objects (3, 6, 9, 12, 15, 18) x number of other object (3, 6, 9, 12, 15, 18) x quantifier (*a few, few, several, many, lots of*). To begin with,

as expected there were main effects of number of focus objects, $F(5, 95) = 11.89$, $p < 0.0001$, MSe = 6.2, of quantifier, $F(4, 76) = 36.66$, $p < 0.0001$, MSe = 22.6, and there was a significant two-way interaction between number of focus objects and quantifier, $F(20, 380) = 106.6$, $p < 0.0001$, MSe = 7.4. Low magnitude quantifiers (*a few, few*) were rated as being more appropriate to describe scenes with small numbers of focus objects than high magnitude quantifiers (*many, lots of*), and conversely high magnitude quantifiers were rated as being more appropriate to describe scenes with large numbers of focus objects than low magnitude quantifiers. There results are consistent with those found in previous studies (e.g., Newstead & Coventry, 2000).

Of most interest were significant interactions involving number of other objects and grouping. First, there were significant two-way interactions between number of focus objects and number of other objects, $F(25, 475) = 1.90$, $p < 0.01$, MSe = 1.1, between number of other objects and quantifier, $F(20, 380) = 22.55$, $p < 0.0001$, MSe = 1.2, and the three-way interaction between number of focus objects, number of other objects and quantifier was also significant, $F(100, 1900) = 4.19$, $p < 0.0001$, MSe = 1.0. This interaction is displayed in Figure 2.
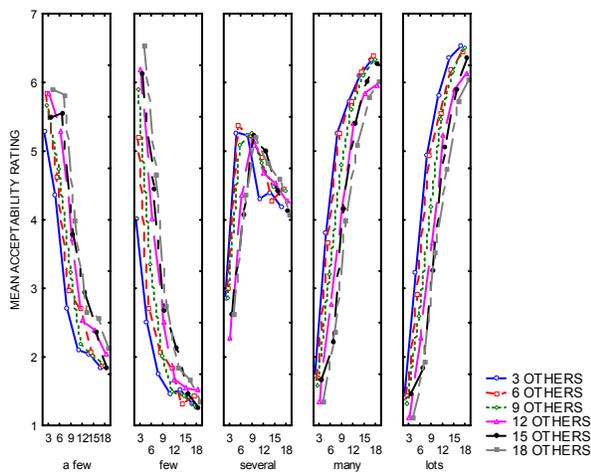

Figure 2 – Three-way interaction between number of focus objects, number of other objects and quantifier in Experiment 1.

These results show that the number of other objects present affects appropriateness of quantifiers to describe scenes. For low magnitude quantifiers, the larger the number of other objects present, the more acceptable the quantifier to describe the number of focus objects. Conversely, with high magnitude quantifiers, ratings are higher when the number of other objects is low. Hence, the other objects appear to be being used as a contrast set for judgments about the appropriateness of quantifiers used to describe the focus set.

Second, there was a significant two-way interaction between grouping and quantifier, $F(4, 76) = 8.38$, $p < 0.0001$, MSe = 1.2, and the three-way interaction between grouping, number of focus objects and quantifier

was also significant, $F(20, 380) = 1.64$, $p < 0.05$, MSe = 0.9. This interaction is displayed in Figure 3.
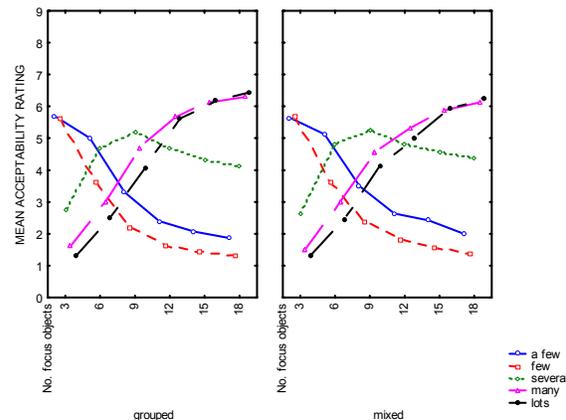

Figure 3 - Three-way interaction between number of focus objects, number of other objects and quantifier in Experiment 1.

For low magnitude quantifiers ratings are higher for the mixed scenes when there are 9 fish or more, and conversely for high magnitude quantifiers ratings are lower for the mixed scenes when there are 9 fish or more.

So, the results of the first experiment reveal two new "context" effects for quantifiers. Both grouping and number of other fish present affect judgments of the appropriateness of quantifiers to describe visual scenes. The second experiment focuses on the spacing between fish to investigate whether this similarly affects sentence comprehension with quantifiers.

Experiment 2
This experiment, like Experiment 1, manipulated the number of focus fish and number of other fish. The design was identical to Experiment 1, but this time instead of varying grouping, spacing between fish was varied instead. Either fish were spaced far apart (5cm between the centre of mass of each object; each fish had a diameter of 1.5 cm) or close together (2 cm between the centre of mass of each object).

Participants
Twenty undergraduate students participated for payment.

Results
The results were analysed using a four-way within-subjects analyses of variance. The variables were spacing (large spacing or small spacing) x number of focus objects (3, 6, 9, 12, 15, 18) x number of other object (3, 6, 9, 12, 15, 18) x quantifier (*a few, few, several, many, lots of*).

To begin with, as expected there were main effects of number of focus objects, $F(5, 95) = 13.10$, $p < 0.0001$, MSe = 5.0, of quantifier, $F(4, 76) = 14.67$, $p < 0.0001$, MSe = 32.1, and there was a significant two-way interaction between number of focus objects and quantifier, $F(20, 380)$

= 80.23, $p < 0.0001$, MSe = 8.0. Consistent with the results of Experiment 1, low magnitude quantifiers (*a few, few*) were rated as being more appropriate to describe scenes with small numbers of focus objects than high magnitude quantifiers (*many, lots of*), and conversely high magnitude quantifiers were rated as being more appropriate to describe scenes with large numbers of focus objects than low magnitude quantifiers.

Of most interest were significant interactions involving number of other objects and spacing. First, there were significant two-way interactions between number of focus objects and number of other objects, $F(25, 475) = 2.66$, $p < 0.0001$, MSe = 1.0, between number of other objects and quantifier, $F(20, 380) = 20.35$, $p < 0.0001$, MSe = 2.6, and the three-way interaction between number of focus objects, number of other objects and quantifier was also significant, $F(100, 1900) = 3.37$, $p < 0.0001$, MSe = 1.5. These interactions mirrored those found in Experiment 1 (see Figure 2). For low magnitude quantifiers, the larger the number of other objects present, the more acceptable the quantifier to describe the number of focus objects. Conversely, with high magnitude quantifiers, ratings are higher when the number of other objects is low. So, again the other objects appear to be being used as a contrast set for judgments about the appropriateness of quantifiers used to describe the focus set.

Second, there was a significant three-way interaction between spacing, number of focus objects and quantifier, $F(20, 380) = 2.01$, $p < 0.001$, MSe = 1.2. This interaction is displayed in Figure 4.
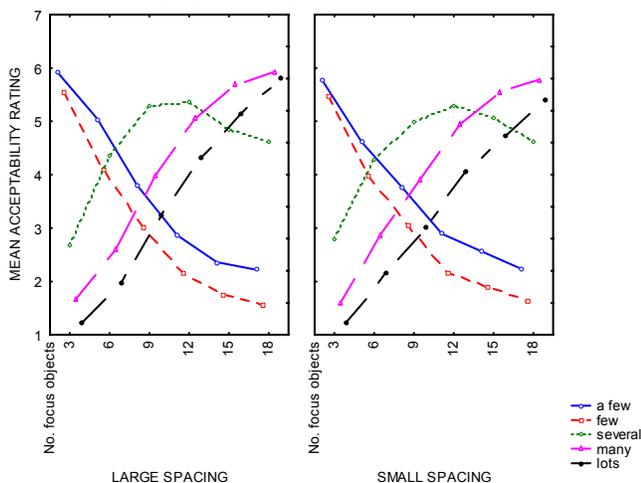

Figure 4 - Three-way interaction between spacing, number of focus objects and quantifier in Experiment 2.

For high magnitude quantifiers ratings are higher for the scenes with large spacing than the scenes with small spacing when there are 9 fish or more. For low magnitude quantifiers the effects of spacing are less marked.

There were also significant interactions between spacing and number of other objects, $F(5, 95) = 4.34$, $p < 0.01$, MSe = 0.7, between spacing, number of other objects

and quantifier, $F(100, 1900) = 2.70$, $p < 0.0001$, MSe = 1.4, and the four-way interaction between spacing, number of focus objects, number of other objects and quantifier was also significant, $F(100, 1900) = 1.51$, $P < 0.01$, MSe =- 1.0. These interactions show that the effects of spacing and number of other fish depend on the quantifier and number of focus objects present.

Overall, the results of the second experiment replicate the effect of number of other objects on quantifier comprehension and introduce another new "context" effect: spacing and number of other fish present both affect judgments of the appropriateness of quantifiers to describe visual scenes. The third experiment examines whether number of other objects, spacing and grouping impact on judgments about the number of objects present in a visual scene.

Experiment 3
Given the results of Experiments 1 and 2, we were interested in establishing whether the effects found may be due to estimates of the number of objects present in the scenes to be described. Consistent with the number literature reviewed above, when the number of objects increases beyond a small number, we expected that the estimated numbers given by participants would deviate from the actual number of objects present. Furthermore, we were also interesting to establish whether spacing and grouping also affect number estimates. Grouping objects together makes them easier to count, and spacing affects the total area/amount of clutter present in the scene, so we expected that estimates would be affected by both variables.

Participants
Twenty undergraduate students participated for payment.

Design and Materials
Participants had the task of estimating how many fish were shown in scenes presented for 500msec. Scenes were presented in blocks, and at the start of each block participants were told to estimate the number of either striped fish or white fish. Scenes were randomized within blocks.

Procedure
Participants were instructed to estimate as accurately as they could how many fish of a given type were present in scenes. At the start of a block participants were instructed to respond to a particular type of fish (white or striped). Practice trials were given at the start of each block to ensure participants were estimating the right type of fish, and a reminder prompt was given at the beginning of test trials. For each trial, a fixation cross was presented in the middle of the screen for 500msec followed by the scene for 500msec followed by a mask (a chequered board) presented with a space in it in which to type estimates. Participants responded by typing in their estimated numbers using the computer keyboard.

Results

We were interested in establishing whether number of other fish, grouping and spacing affected judgments about the estimated number of fish present. We initially looked at the data, and it became clear that participants were subitizing when there were only three focus objects as there was no variance in estimates for these scenes. Therefore, scenes with 3 focus fish were dropped from the analyses. The remaining data were analyzed using a four-way within-subjects analyses of variance. The variables were spacing (large spacing or small spacing) x grouping (grouped or mixed) x number of focus objects (3, 6, 9, 12, 15, 18) x number of other objects (3, 6, 9, 12, 15, 18). To begin with, there was a main effect of number of focus objects, $F(4, 76) = 242.85$, $p < 0.0001$, MSe = 16.5. As the number of fish in focus increased, so did the estimates. However, consistent with the number literature, when the number increased above 6 fish participants underestimated the number of fish (mean estimates of 8.11, 10.39, 12.19 and 13.88 for 9, 12, 15, 18 fish respectively). There were also main effects of spacing $F(1, 19) = 6.94$, $p < 0.05$, MSe = 12.9, and of grouping, $F(1, 19) = 14.1$, $p < 0.0001$, MSe = 14.1. Overall estimates were higher when the fish were spaced close together (M = 10.31) than when they were spaced far apart (M = 9.89), and when they were grouped (M = 10.62) than when they were mixed (M = 9.58).

There was a significant two-way interaction between spacing and number of focus objects, $F(4, 76) = 3.90$, $p < 0.01$,l MSe = 5.1. This is displayed in Figure 5. An effect of spacing was only found for higher numbers of fish (12, 15, 15, 18).
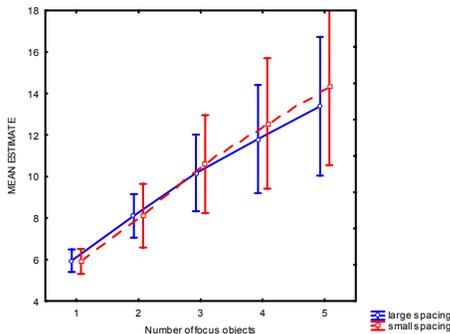

Figure 5. Interaction between spacing and number of focus objects in Experiment 3.

There were also significant two-way interactions between grouping and number of focus objects, $F(4, 76) = 31.93$, $p < 0.0001$, MSe = 3.6, between grouping and number of other objects, $F(4, 76) = 6.69$, $p < 0.001$, MSe = 2.1, and between number of focus objects and number of other objects, $F(16, 304) = 3.51$, $p < 0.0001$, MSe = 2.2. The three-way interaction between grouping, number of focus objects and number of other objects was also significant, $F(16, 304) = 3.25$, $p < 0.0001$, MSe = 2.2. This interaction is displayed in Figure 6. As can be seen in this Figure, as the number of

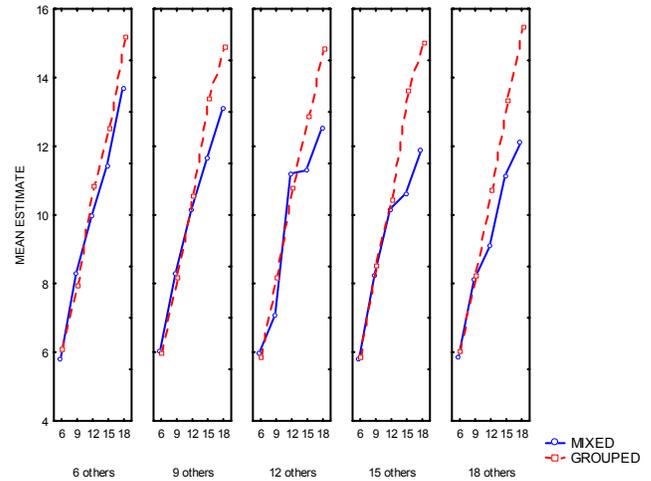other objects becomes greater, the effect of grouping becomes more extreme.

Figure 6 - Three-way interaction between grouping, number of focus objects and number of other objects in Experiment 3.

These results show that grouping, spacing and number of other objects all affect estimates of the number of objects present in a visual scene when the scene is presented for a short period of time. Critically, the effects of grouping and number of other objects only occur when there are larger numbers of objects present. This is consistent with the notion that for small numbers participants are able to subitize, and hence to return accurate number estimates (e.g., Trick & Pylyshyn, 1993).

General Discussion

When people make judgments about the appropriateness of quantifiers to describe scenes involving a number of objects, we suspected that they were likely to make judgments based on estimates of the number of objects present rather than counting the number of objects present. The results of the experiments indeed support the claim that there is a correspondence between estimations of number and the ratings of the appropriateness of quantifiers to describe the same scenes.

Three new effects on both quantifier rating and number judgments have been uncovered. Most strikingly, when judging number of objects (Experiment 3), both spacing and grouping affect judgments, but only when the number of objects in the scene rises beyond the point where participants are able to subitize. Similarly, both these variables affect quantifier ratings, but again only when the number of focus objects rises above the subitizing region. Notably, no context effects are found for quantifiers when only 3 objects are present in scenes, and in the number judgment experiment participants were so accurate at estimating the number there was no variance in the data for these scenes.

These results indeed suggest a correspondence between estimates of numbers of objects in a scene and context effects for quantifiers. Counting large numbers of objects in scenes is effortful and time consuming, and would get in the way of communication. Similarly, when judging a number of objects, it is to be expected that we would evolve mechanisms of estimating number without counting. Furthermore, it could be that other context effects noted in the quantifier literature emerge from this relation. For example, knowing how many objects are in a set affects the likelihood that a certain number of objects from that set is present. So one might predict expected frequency effects when estimating number, just as one gets the effect with quantifiers. We are currently examining whether this, and other context effects indeed affect number judgments.

More generally, the argument that language comprehension maps onto perceptual processing is consistent with a growing literature showing relations between language comprehension and perception (e.g., Barsalou, 1999; Glenberg & Kaschak, 2000; Coventry & Garrod, 2004; Zwaan, 2004). We are currently developing a computational model which grounds the meaning of quantifiers directly in perceptual representations of the scene being described. Preliminary simulation experiments show that the part of the model trained to do "psychological counting" uses some of the same factors found to be important in the present experiments (Rajapakse et al., under review). Furthermore this model is a development of a model we have previously developed for spatial prepositions (Coventry et al., in press). Such an approach affords the possibility of linking context effects across syntactic categories, which may ultimately lead to a set of common perceptual parameters important across a wide range of syntactic categories. As a start, it has been noted that prepositions and quantifiers do indeed share some of the same perceptual determinants affecting their situation specific meaning (Newstead & Coventry, 2000; Coventry & Garrod, 2004; Coventry et al., in press).

References

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. (1984). Magnitude estimations of frequency and amount. *Journal of Applied Psychology, 53*, 313-320.

Coventry, K. R., Cangelosi, A., Rajapakse, R., Bacon, A., Newstead, S., Joyce, D. & Richards, L. V. (in press). Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. In C. Freksa, B. Nebel, M. Knauff & B. Krieg-Bruckner (Eds.), *Spatial Cognition, Volume IV.* Lecture notes in Artificial Intelligence. Springer Verlag.

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.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20(1),* 1-55.

Glenberg, A. M., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9(3),* 558-565.

Hormann, H. (1983). Then calculating listener, or how many are einige, mehrere and ein paar (some, several and a few). In R. Bauerle, C. Schwarze, & A, von Stechow (Eds.), *Meaning, use and interpretation of language.* Berlin; De Gruyter.

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.

Moxey, L. M., Sanford, A. J., & Dawydiak, E. J. (2001). Denials as controllers of negative quantifier focus. *Journal of Memory and Language, 44,* 427-442.

Newstead, S. E., & Coventry, K. R. (2000). The role of context and functionality in the interpretation of quantifiers. *European Journal of Cognitive Psychology, 12(2),* 243-259.

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).

Rajapakse, R., Cangelosi, A., Coventry, K., Newstead, S., & Bacon, A. (under review). *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

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).

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation,* Vol. 44. New York: Academic Press.