HAYDEN WALLES

Psychobabble: Grounding Language in the Brain

1. INTRODUCTION

We tend to take language for granted. The ease with which children learn to speak – and our fluency as adults – fools us into thinking it is simple. Even when some find reading and writing difficult to learn, the literate among us still assume that some simple problem stands in the way of picking up these apparently easy skills.

But language is not simple. In fact Morten H Christiansen and Simon Kirby have gone so far as to call it the hardest problem in science.[1] Despite centuries of expertise with the formalities of language, and fifty years of modern scientific linguistic research, nobody really knows how human language works. Noam Chomsky[2], the instigator of modern approaches to language (which are generally inspired by him or reactions against him) has yet to completely describe the so-called *Universal Grammar* which his theory posits to exist inside every human brain. Nor has anyone else come up with a coherent theory that comprehensively explains the data.

There is plenty of data, though. Linguists have collected details on languages from all over the world. Patterns have emerged, but the linguistic database remains fragmented.

I am applying a relatively new method for the study of language that has appeared only in the last few years. This method assumes that at least some parts of language can be explained in terms of other aspects of our psychology. Formal linguists, such as the Chomskyans, tend to base their theories solely on *linguistic* data. This leads very easily to a belief that language is self-contained, its rules essentially arbitrary.

Cognitive linguists like Ronald W Langacker[3] and George Lakoff[4] reacted against Chomsky. They believe that linguistic structure is entirely due to psychological phenomena that in the past were not considered pertinent to the study of language. Cognitive linguists argue that the meaning of language is tied to human mental processes, *not* the world viewed in some objective fashion. According to this view a wide range of psychological phenomena, from perception to "higher" cognition like metaphorical reasoning, constrain language and leave their stamp in its design. In this case it makes sense to look to psychology for insights into language.

I look particularly to the sensorimotor system, the nuts-and-bolts part of the brain that controls perception and action at a concrete level. I believe there is good reason to bring this basic system into the study of language. This is a bold claim, given the apparent disparity between these two fields, and in the remainder of this paper I hope to convince you of its validity. First I will briefly cover some general arguments for such a position. Then I will present a specific hypothesis linking grammatical number (singular, plural, etc.) to perceptual grouping processes in the visual system. I am developing a computational model to explore this hypothesis, and will present some interim results that I believe bolster my claim.

2. LANGUAGE AND THE SENSORIMOTOR SYSTEM

Language is conventionally considered as one of several independent *faculties*, or *modules*, within the brain. Jerry Fodor is associated with an extreme version of this[5], in which language in particular is an isolated component of the mind, communicating with other components through strictly controlled channels. By talking of language as a module here, I mean that all of the representations and processing required for language are partitioned off from other parts of the brain. The language module can then be viewed as a black box, which takes in sensory input and produces abstract 'ideas', or takes 'ideas' and converts them into speech or gestures. This conventional view is illustrated in Figure 1(a). Ray Jackendoff has noted that though this is a common view among linguists and others, it is unnecessary. Chomsky himself identified this strict view early on and rejected it.[6] He distinguished between functions and the way they were implemented in the brain. This allows one to talk about the language faculty as a separate *functional* entity, as shown in Figure 1(a), even though it might share parts of the brain with other mental modules.

Ray Jackendoff, a former student of Chomsky, presents an alternative view of this isolation. He also views language as a distinct module, but emphasises the interaction of all cognitive modules – in particular the way interacting modules constrain one another.[7] This interaction allows the language module to be affected by other cognitive modules far more than the conventional view allows. While the language module may be autonomous, it is constrained by information fed to it by other modules. This approach assumes the same *structure* as the strictly modular view, but also considers the interactions between modules.

An elegant example of how the sensory system can influence language this way is the cross-linguistic research conducted by Brent Berlin and Paul Kay on colour names.[8] They found that languages with only two basic colour names used them to label black and white (or, more correctly, dark and light). Languages with three colour names labelled these plus red. In languages with more colour names those with the same number of basic names used them to label the same colours, with little variation. What's more, the actual colours (measured objectively based on the frequency of light) were the same when labelled in different languages. Kay and Chad K McDaniel later found that these key frequencies corresponded to the maximum sensitivity of the cells in the retina that detect colour.[9] Nobody contends that the eye is part of an autonomous language module. But the visual cortex perceives colour as interpreted by the retina, which constrains how other parts of the brain represent colour. Thus in a very basic way language can be influenced by our physiology – even aspects of physiology usually considered remote from language.

The alternative is for the language module to be less than autonomous in the brain. Figure 1(b) illustrates how this could work, with the language faculty sharing some of its components with other faculties, such as the sensorimotor system. Why should one believe such a thing possible or likely, especially since language and perception or motor action seem so unrelated? There are, I think, some good reasons. Natural selection is a very good recycler.[10] New functionality evolves by modification of existing structures. The wings of a bat are modified legs, for example. Such changes can lead to great complexity when combined with duplication. Consider insects, whose bodies consist of a series of segments, each performing a different function (antennae, eyes, legs, and so on), but all derived from the same basic pattern. It makes sense that functionality in the brain would emerge in a similar way, so that the neural machinery for language would emerge from a part of the brain structured for some other job. This might occur if an older structure was duplicated and freed from its original purpose. Language might even end up sharing parts of the brain with the older function. In either case we would expect that the original function might leave its stamp on the language faculty.

So from an evolutionary perspective I think it quite sensible to look to other aspects of cognition to help throw light on language, and the sensorimotor system is a good candidate. After all, much of what we use language to talk about are the objects and relationships we perceive around us, and the actions we take in the world. Langacker's early work in cognitive linguistics[11] was aimed at motivating the distinction between nouns and verbs as a distinction between regions (of real or abstract spaces) and processes, aping the way humans perceive the world.

3. BACKGROUND TO MY HYPOTHESIS

I am looking for a causal link between an aspect of sensorimotor processing and a corresponding aspect of the structure of language. In particular, I am interested in how knowledge about the way the visual system attends to and categorises objects might be brought to bear on the structure of *noun phrases*. Noun phrases are phrases built around nouns that can serve as grammatical subjects and objects. For example, "a dog", "some trees" and "John" are simple noun phrases. They are the linguistic structures employed to describe actual objects in the world (among other things).

In the remainder of this paper I am going to discuss a psychological model of the way that groups of objects are classified in the visual system, which I believe can also be used to explain the syntax of noun phrases – at least the aspects relating to *number*. But before I get into the details, I need to give a bit of background, on the psychological and linguistic fronts, and define my terms. What exactly do I mean by the visual system; and what exactly is this concept of number in language? I will deal with the second point first.
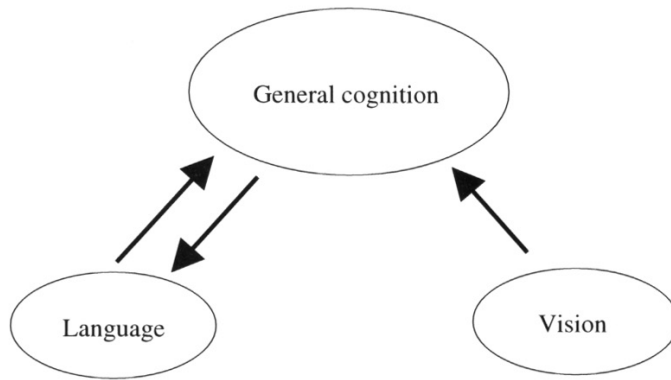
Figure 1(a): the traditional view of the language faculty in relation to other mental faculties. Language exists as a separate module within the brain.
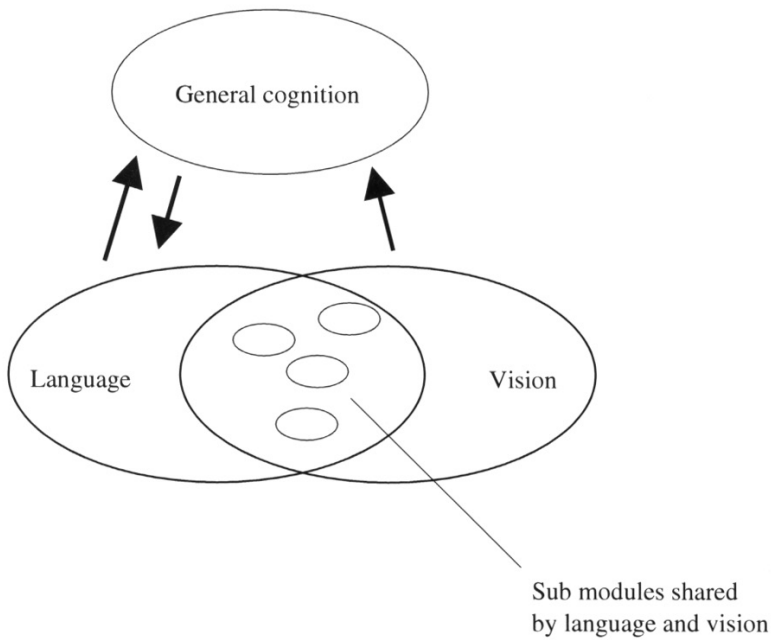

Sub modules shared by language and vision

Figure 1(b): the view of language that I am suggesting. Some subsystems are shared between language and other mental faculties.

Figure 1

3.1   Linguistic Number

Number comes up in a few places in language – obviously in the case of words like one, two, three, and more complicated phrases like four million, two hundred and six. It also comes up in quantifiers like some, lots and many. But the manifestation that I am interested in is the so-called *number feature* that attaches itself to nouns and verbs. In English, for example, there is a difference between singular and plural noun phrases: "the dog" versus "the dogs". Furthermore, the subject of a sentence and its verb must agree in the number feature: "The dog runs away" versus "*The dogs runs away".[12] This distinction suggests that there are at least two separate components to nouns: the noun itself indicating a class, and the number feature indicating whether one or more members of the class are intended. In an introductory text on the subject, Haspelmath[13] notes that the regularity of number marking is best represented as a rule for combining two separate elements. The linking of the number feature with verbs also suggests it has a life independent of its noun, as it were.

Other languages exhibit even more variations of the number feature; features that indicate the speaker is referring to precisely two things, or three things. But across languages there are some generalisations that can be made which allow English to serve as a reasonable model for language generally. In those languages which do divide noun phrases up by number, the distinction between singular and plural seems to be most basic.[14] What causes singular and plural noun phrases to be marked differently?

According to one view, this distinction is entirely arbitrary and reflects arbitrary organisation of the language faculty. This is the default Chomskyan position. But even the Chomskyans are willing to accept the possibility that the distinction is driven by organisation of other brain mechanisms – such as the visual system.

3.2   The Visual System

We have gained a broad idea of how the human visual system works. Visual information passes from the eyes to the visual cortex at the back of the brain where processing begins. D H Hubel and T N Wiesel originally found that neurons in the visual cortex respond to particular primitive features: edges and corners either still or moving in certain directions.[15] Ungerleider and Mishkin found that processing then splits into two *streams*.[16] The *dorsal stream* runs into the parietal cortex and is generally associated with attention and location. The *ventral stream* passes into the temporal cortex and is associated with classification. Melvyn A Goodale and A David Milner note striking evidence for the distinction.[17] Neurological patients with damage to the temporal lobe, for example, may be incapable of naming objects (i.e., classifying them) but still able to interact with them in a generic way. More recent methods have allowed an even closer look at classification in humans. Kalanit Grill-Spector and Rafael Malach, in a review of functional magnetic resonance imaging (fMRI) studies of the visual cortex, reveal that certain parts of the ventral visual cortex respond to different categories of stimuli, such as faces, places and animals.[18]
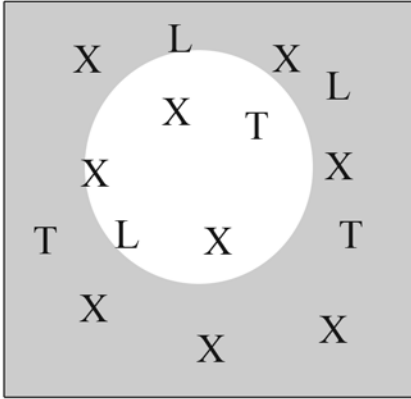
Figure 2(a): Space-based attention.

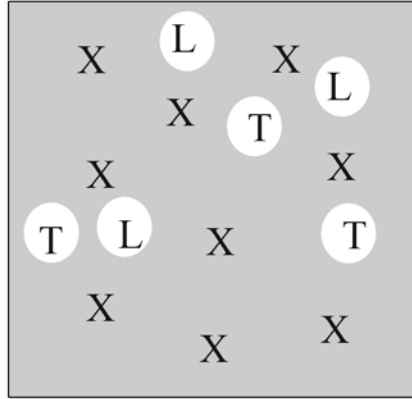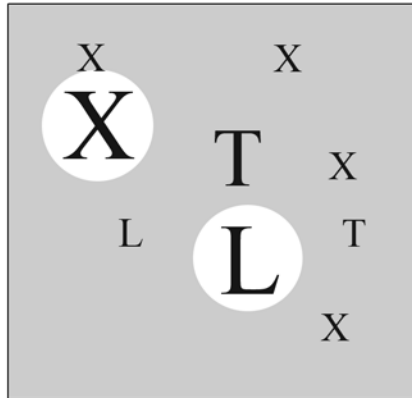Figure 2(b): Feature-based attention.

Figure 2(c): Scale-based attention.

Figure 2

3.2.1   Visual Attention

Attention, for psychologists, is quite a nebulous term which hides a great deal of complexity. It can apply equally to the meanest operations of peripheral processing and the highest levels of conscious thought. In everyday life we use it in the latter sense, but I want to talk about a more restricted sense – visual attention. By doing this I am limiting myself to methods that the brain uses to ensure one aspect of the visual field receives more emphasis, or processing, than the rest. Some examples should suffice to show what I mean.

Figure 2(a) shows an example of space-based attention. This is a conventional conception of visual attention.[19] The bright patch in the figure indicates the region attended to. Stimuli falling within are attended to – though what this entails is not certain. It is generally assumed that the entire field of view is too complicated to process at once, and attention decides which part of it should be processed. Attention then acts like a filter, picking out parts of a scene for close examination.

Space-based attention is not the only kind of visual attention, though. Figure 2(b) shows an example of feature-based attention. Feature here means primitive visual features, such as the direction, colour or length of line segments. In Figure 2(b) horizontal and vertical lines are the features attended to, picking out the Ls and Ts from the Xs. Note that although in the figure the letters are picked out by a bright circle, it is the *features* – the horizontal and vertical lines – that are picked out, irrespective of their location in the visual field. Feature-based attention is a more recent discovery than space-based attention, but its reality has emerged from psychological work on visual search, notably Anne M Treisman and Garry Gelade's Feature Integration Theory[20]. Such work shows that in certain circumstances a target object can be made to 'pop out' from a field of distracting objects based on feature differences without observers having to scan each object with space-based attention.

Figure 2(c) shows a further, more controversial, example of visual attention, which I have labelled scale-based attention. This refers to bias within the visual system to respond to stimuli of a particular size, the large letters in the figure. Evidence for this kind of attention comes from studies into *global precedence*, a term coined by David Navon[21] to describe the tendency for the global stimulus to swamp local components. In pictures like Figure 3 Navon found that the global figure (the A) interferes with tasks relating to the local figures (the Xs), slowing them down or increasing errors. However the Xs do not interfere with tasks relating to the A. This suggests that the visual system can choose to consider stimuli only at particular scales, though this kind of attention might not be as flexible as the others described above. Further evidence of scale-based attention comes from work by J Vincent Filoteo, Frances J Friedrich and John L Stricker[22], who found subjects were slower to respond to a stimulus at a different level (global or local, as in Figure 3) from the previous stimulus. This effect seemed to be independent of spatial attention.

There is a further useful distinction within attention. Endogenous, or top-down, attention is the term applied to attention directed from within. By contrast, exogenous, or bottom-up, attention is driven by the stimulus. Something that catches the eye stimulates exogenous attention; endogenous attention results from an internal choice.
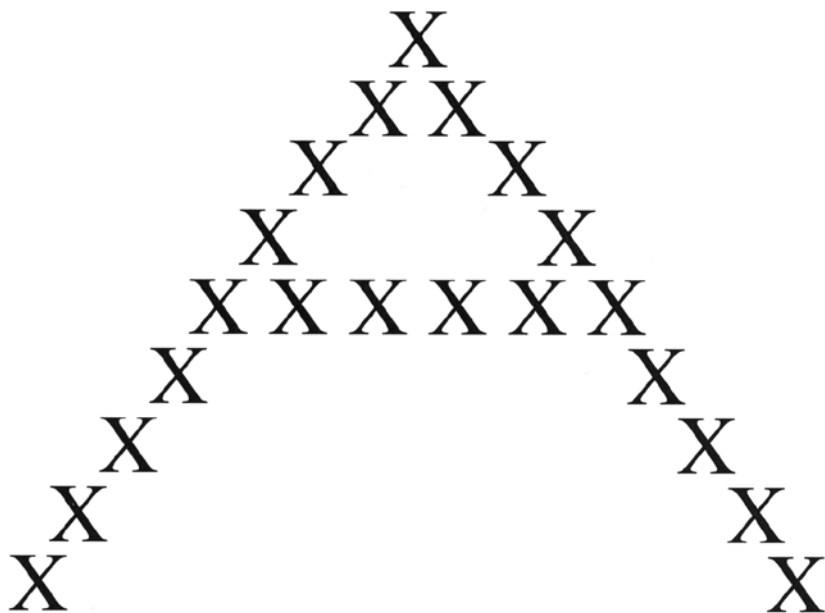
Figure 3: A typical stimulus used to demonstrate global precedence.

3.2.2   Classification

The classification of visual stimuli is an important problem in psychology and artificial intelligence, and one that has not yet been adequately solved. Progress has been made in some applications (such as optical character recognition used to read text), but the human ability to pick out and classify objects under a wide range of conditions eludes us as yet.

Object classification appears to go on in the ventral stream. Location information seems to be a dorsal stream specialty, so that classification in the ventral stream need not worry about where things are. All it need do is classify objects (presumably just the ones currently attended to). This is borne out by looking at the brain, where it is found that as one follows the ventral stream neurons respond to ever-larger regions of the retina. This implies that ventral stream processing is *spatially* or *location invariant* – it does not care where objects are. Obviously classification must also be *scale invariant* – it must be able to classify objects correctly no matter what their apparent size on the retina, and neurons in the ventral stream also seem to have this ability.[23]

I am interested in classification because it interacts with the attention system. Attention is supposed to filter raw visual input for later processing, and classification makes up a major part of that later processing.

4. A COMPUTATIONAL MODEL OF VISUAL ATTENTION AND CLASSIFICATION

Figure 4 illustrates the components of a model that I hope can link classification and attention to linguistic number. It is a model of part of the human visual system, based on widely used biologically plausible computational simulations. As the figure shows, after early processing, flow of information splits into the ventral and dorsal streams (analogous to vision in humans), and I will follow this broad division in the following discussion.

4.1   Early Processing

The model takes as input raw images. In reality this would be light impinging on the retina, here it can be thought of as a small bitmap image. This image is processed by a set of visual filters that model the effect of neurons in the primary visual cortex. The output of this processing is a set of *feature maps*.[24] Each feature map has the same spatial layout as the input (it is *retinotopic*) but instead of each point indicating light intensity, it indicates the strength of a particular feature. Each feature map corresponds to a different feature.

For example, one feature map might be sensitive to short vertical lines, so that 'bright' points on it would indicate places where the input image contains short vertical lines. Another feature map might be sensitive to short horizontal lines, another to short lines at a 45° angle. Feature maps sensitive to short lines are good for dealing with small objects, but other feature maps are sensitive to long lines, better for examining large objects. My model only extracts features based on line direction and length, but in reality many other kinds of features are available, like colour.

4.2 The Classifier: A Convolutional Neural Network

The classifier I have chosen is known as a convolutional neural network (CNN). Yan LeCun has made this a popular classifier for character classification.[25] My CNN, however, is based on the work of Michael C Mozer and Mark Sitton where it played a role in a computational model of attention.[26]

Mozer and Sitton were interested in the CNN because it resembles the kind of neural structures in early vision. The CNN takes as input a group of feature maps and produces as output a category. In between a series of layers reduce the feature maps to the category. Each layer consists of two parts: a convolving part, which extracts local feature combinations from its input, and an abstracting part, which reduces the size of the feature map. In my current implementation, for example, the input is eight feature maps, each measuring 31 x 31 pixels. The output from the first layer of the CNN is twelve feature maps (each feature being some combination of the eight input features) measuring 15 x 15 pixels. The output from the CNN is one feature map for each category containing only one 'pixel' which is active if the CNN detects an instance of that category.

The CNN learns the relationships between input shapes and categories using a conventional connectionist learning algorithm.[27]

The CNN has several notable characteristics. Mozer and Sitton were aware of its ability to abstract across the visual field, and so to classify objects at different locations. But this also, critically, allows it to classify homogeneous groups of objects. It cannot, however, reliably classify more than one type of object at a time, and so cannot deal with heterogeneous groups. It can also be trained to classify objects at different scales. A cluster of triangles produces the same output as a single triangle, and a small triangle produces the same output as a large triangle. Mixed objects – triangles and squares, for example, generally produce no output at all. So the CNN appears to be invariant to scale, location and number. As mentioned above, scale and space invariance are known properties of the human visual system. Number invariance is not, but is a prediction of my model. Figure 5 shows the CNN at work.

Figures 5(a) and 5(b) show classification of a small and large object, respectively. As expected, the classifier produces the correct answer when attending to the correct scale. The incorrect scale gives a nonsense answer, a 'best-guess', presumably. This is probably a result of training only with positive examples, such as always presenting a big shape when classifying at the large scale. By adding negative training examples -- in which there are no big shapes, in this case -- this problem could probably be eliminated.

In Figure 5(c) the large triangle is correctly classified, but the small shapes are not the same, and so the classifier produces no category. Mozer and Sitton noticed this, and used it to bolster their claim that spatial attention is necessary for adequate classification by concentrating on one object at a time. But in Figure 5(d), a very similar image made of identical small shapes, the classifier categorises them collectively as square. Mozer and Sitton did not notice this exception.

Finally, Figure 5(e) shows that the network responds to groups of shapes at the small scale no matter where they are in the input.
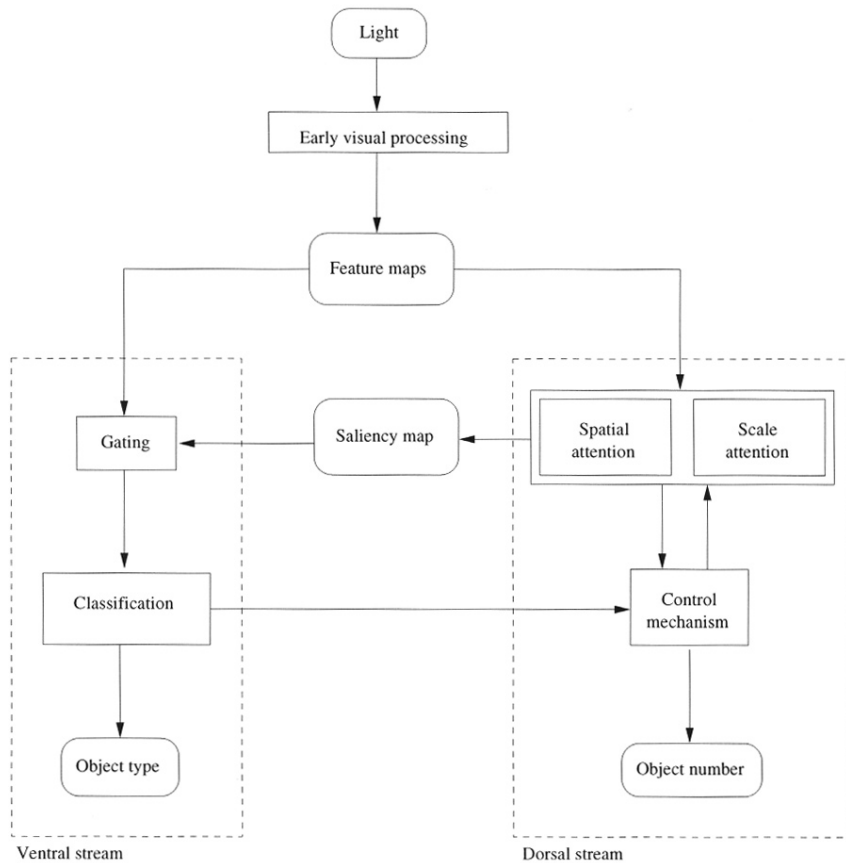
Figure 4: A diagram showing the major components of the model. Rounded boxes indicate representations of data, other boxes indicate processes. Arrows indicate flow of information.

The output from the classifier (which is either an object category – square, triangle, etc. – or nothing, indicating it could not classify the input) forms one of the outputs of the model, the *object type*.

4.3   Attention

Two attention systems also use the feature maps, as shown on the right of Figure 4. The space-based attention system uses feature maps to work out which regions contain interesting stimuli and the order in which these should be attended to. This is a common approach to spatial attention similar to that employed, for instance, by Jeremy M Wolfe[28] and Laurent Itti and Christof Koch[29] in their computational models. The scale-based attention system performs a similar job to select the best scale to operate at, and this is the result of my own work. These attention components also receive input from the control mechanism. This control mechanism can therefore direct attention to particular regions or scales, or at least influence those chosen from the feature maps. These two inputs to the attention modules correspond to endogenous and exogenous attention.

Attention limits what reaches the classifier from the feature maps, just as in Mozer and Sitton's model.[30] In Figure 4 this is performed in the gating component. After gating, only features that lie within the attended region *and* are of the attended scale (big or small) pass through to the classifier. The classifier then uses these features to decide what kind of object it is looking at.

4.4   The Control Mechanism

The classifier also signals its status to the control mechanism, indicating whether it can successfully classify its input or not. The control mechanism, by consulting the classification results and the attention system produces the other output of the model, the *object number*. This output only appears once the classifier produces a definite category. Until then, the control mechanism manipulates the attended region and scale until the gated features do produce a category from the classifier. Once a category appears the control mechanism can produce a number output, as laid out in Table 1 (at the end of part 4 below).

Simply stated, if the control mechanism detects a discrepancy between the size of the attended region and the attended scale it concludes that multiple objects are present.

The control mechanism is key to operation of the model, as an example of the model's operation will make clear. Consider the input of Figure 5(d). When presented to the model, the input image will first be processed into feature maps. At the beginning, no region is marked in the saliency map, meaning nothing reaches the classifier. The attention systems immediately go to work, though. The spatial attention module will identify the region of the large triangle as interesting. The scale attention module will identify the large scale as interesting. This is determined purely by the stimulus, without control mechanism intervention.

Now that the saliency map contains a region and scale, the gating module can filter the input into the classifier. Gating passes on feature maps corresponding to large features and everything corresponding to the large triangle. The classifier produces a classification of TRIANGLE, which appears at the object type output. The classifier also informs the control
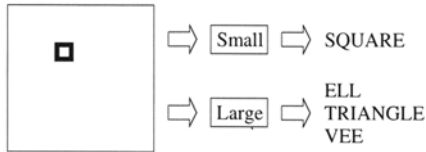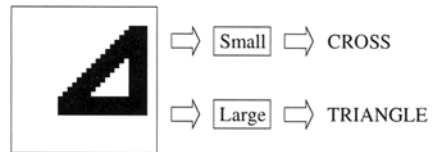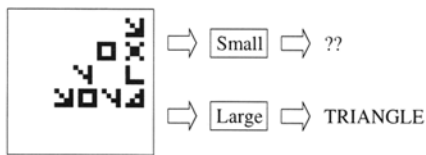
Figure 5(a)
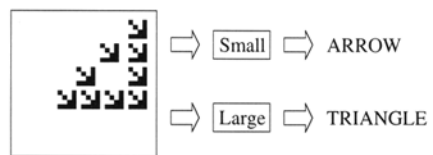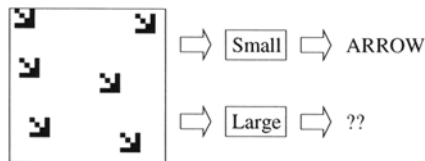

Figure 5(b)

Figure 5(c)

Figure 5(d)

Figure 5(e)

Figure 5: Examples of output from the classifier. No spatial filtering is applied (as would be done by space-based attention), but scale filtering is applied (as would be done by scale-based attention). The results of classifying a "small" and "large" object (or objects) are shown beside each input image.

mechanism of a successful match. The control mechanism consults the attention modules, finds that a large region has been classified as a large object, and from the rules in Table 1 produces ONE at the object number output.

This is not the end, though. Now the control mechanism commands the scale attention module to shift to smaller scale, without altering the region attended to by the spatial attention module. The gating module passes on feature maps corresponding to small features, and the classifier output changes to ARROW. The control mechanism, seeing a successful classification, consults its rules and produces MANY at the object number output.

Further operations are possible, too. Consider the input in Figure 5(c) instead. In this case the last operation would have failed because the classifier cannot classify all the small shapes at once. But now the control mechanism might order the spatial attention module to locate smaller regions of interest, and constrain attention sufficiently that it could classify a single small shape. This could be repeated as necessary.

The control mechanism produces its object number output quite separately from the object type produced by the classifier. Though there is limited communication between these two modules, they are quite distinct – Figure 4 shows them belonging to different processing streams (and so different parts of the brain). The outputs are independent of each other – *in the same way that noun and number are independent in language*. Indeed, one might speculate that the first example above corresponds to generation of the phrase 'triangle of arrows'.

The requirements of the control mechanism go beyond the simple relationship of Table 1, involving sequencing and decision making. While it might be possible to construct a system that followed these rules from scratch, it would not necessarily be consistent with the reality of attention in humans.

Instead, I am seeking a psychologically motivated system that not only reproduces the rules, but also exhibits other known attention phenomena. I hope the model will be able to reproduce global precedence and visual search results.[31] The attention and control mechanism form the focus of current and future work.

Attended region	Attended Scale	Object number
Big	Big	ONE
Big	Small	MANY
Small	Small	ONE

Table 1: The rules used by the control mechanism of my model to work out the object number output. The attended region size and attended scale come from the attention modules, as shown in Figure 4. Once the attended input has been classified the control mechanism can output an object number.

5. CONCLUSION

I have outlined general principles by which the language faculty can be grounded in the working of the brain. These suggest it may be profitable to explore conventionally non-linguistic aspects of cognition to shed light on how language works. In this spirit I presented my own model, which explores the psychology of visual attention and classification as a cause of linguistic structure. My model produces an object type output from the classifier, and an object number output from the attention and control mechanisms. These outputs (from a purely *visual* model) correspond neatly to the noun and number marker, respectively, in linguistic models. The one and many division of the model's object number output correspond to singular and plural markers in language. In addition, my model produces these outputs in different places, mirroring the separation of their linguistic counterparts.

As discussed earlier, there are two ways that vision could affect language. If something like the outputs of my model are the only information available to the language faculty, then it makes sense that the linguistic structures produced correspond closely to the model.

Alternatively, language could be partly 'piggy-backed' on older cognitive systems, like vision. In this scenario the parts of the brain responsible for visual processing would also be part of the language system, as in Figure 1(b). Humans clearly are not limited in their numerical ability to the one-many distinction, yet this distinction forms a mandatory part of language. This would make sense if the language module was serendipitously attached to a more primitive kind of numerical cognition, corresponding to my model, rather than more modern, sophisticated systems.

Either way, my account is firmly within the cognitive camp, since it explains the meaning of linguistic structure in terms of mental processes, rather than objects in the world (though those mental processes are supposed to be relevant to the world around us, of course). Non-cognitive accounts instead try to tie linguistic phenomena either to a logical representation of reality, or to nothing at all.

I have presented the model whole, though I have as yet only implemented the early processing, classifier and parts of the attention modules. I must complete and evaluate the model properly before drawing any firm conclusions. I am hopeful, however, from work completed so far, that the model will be capable of the behaviour described and provide a plausible explanation of the distinction between plurality and singularity in language.

1   Morten H Christiansen and Simon Kirby, "Language evolution: The hardest problem in science?", *Language Evolution,* eds, Morten H Christiansen and Simon Kirby (Oxford University Press, 2003). This volume provides excellent coverage of recent approaches to the origin of language.

2   Chomsky's classic work is Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, MA: MIT Press, 1965), though this theory has undergone several revisions the latest of which can be found in Noam Chomsky and Howard Lasnik, *The Minimalist Program* (Cambridge, MA: MIT Press, 1995). An introduction to more recent Chomskyan theory can be found in Liliane Haegeman, *Introduction to Government and Binding Theory* (Oxford: Blackwell, 1991).

3   Ronald W Langacker, "An introduction to cognitive grammar", *Cognitive Science* 10 (1986), 1-40.

4   George Lakoff, *Women, Fire and Dangerous Things: What Categories Reveal About the Mind* (Chicago: University of Chicago Press, 1987).

5   Jerry Fodor, *Modularity of Mind* (Cambridge, MA: MIT Press, 1983).

6   Chomsky, *Aspects of the Theory of Syntax* cit. and see also Noam Chomsky, "Language and nature", *Mind* 104 (1995), 1-61.

7   Ray Jackendoff, "The architecture of the linguistic-spatial interface", *Language and Space*, eds, Paul Bloom, Mary A Petersen, Lynn Nadel and Merrill F. Garrett (MIT Press, 1996) and Ray Jackendoff, *Foundations of Language: Brain, Meaning, Grammar, Evolution* (Oxford: Oxford University Press, 2002).

8   Brent Berlin and Paul Kay, *Basic Color Terms; their Universality and Evolution* (Berkeley: University of California Press, 1969).

9   Paul Kay and Chad K McDaniel, "The linguistic significance of the meanings of basic color terms", *Language* 54 (1978), 610-646.

10   See Mark Ridley, *Evolution* (Boston: Blackwell Scientific Publications, 1993), an introductory text, for details.

11   Ronald W. Langacker, "Nouns and verbs", *Language* 63 (1987), 53-94.

12   In linguistics, an example marked by a star indicates ungrammaticality.

13   Martin Haspelmath, *Understanding Morphology* (London: Arnold, 2002).

14   Joseph Greenberg, "Some universals of grammar with particular reference to the order of meaningful elements", *Universals of Language*, ed. Joseph H Greenberg (Cambridge, MA: MIT Press, 1963).

15   D H Hubel and T N Wiesel, "Receptive fields and functional architecture of monkey striate cortex", *Journal of Physiology* 195 (1968), 215-243.

16   Leslie G Ungerleider and Mortimer Mishkin, "Two cortical visual systems", *Analysis of Visual Behavior*, eds, David J Ingle, Melvyn A Goodale and Richard J W Mansfield (Cambridge, MA: MIT Press, 1982).

17   Melvyn A Goodale and A David Milner, "Separate visual pathways for perception and action", *Trends in Neuroscience* 15 (1992), 20-25.

18   Kalanit Grill-Spector and Rafael Malach, "The human visual cortex", *Annual Review of Neuroscience* 27 (2004), 649-677.

19   Michael I Posner, "Orienting of attention", *Quarterly Journal of Experimental Psychology* 32 (1980), 3-25.

20   Anne M Treisman and Garry Gelade, "A feature-integration theory of attention", *Cognitive Psychology* 12 (1980), 97-136 and also see Anne Treisman, "The perception of features and objects", *Visual Attention*, ed. Richard D. Wright (Oxford University Press, 1998) for a revised version.

21   David Navon, "What does a compound letter tell the psychologist's mind? ", *Acta Psychologica* 114 (2003), 273-309.

22   J Vincent Filoteo, Frances J Friedrich and John L Stricker, "Shifting attention to different levels within global-local stimuli: A study of normal participants and a patient with temporal-parietal lobe damage", *Cognitive Neuropsychology* 18 (2001), 227-261.

23   Grill-Spector et al., *The Human Visual Cortex* cit.

24   A well known work on modelling of early visual processing is David Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco: W H Freeman, 1982); I have been more influenced by work such as Laurent Itti and Christof Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", *Vision Research* 40 (2000),1489-1506 which employs feature maps.

25   Yan LeCun and Yoshua Bengio, "Convolutional networks for images, speech, and time-series", *The Handbook of Brain Theory and Neural Networks*, ed. Michael A. Arbib (MIT Press, 1995).

26   Michael C Mozer and Mark Sitton, "Computational modeling of spatial attention", *Attention*, ed. Harold E Pashler (Hove: Psychology Press, 1998).

27   The RPROP algorithm: Martin Riedmiller, "Rprop – description and implementation details", Technical report, Institut für Logik, Komplexität und Deduktionssyteme, University of Karlsruhe (1994).

28   Jeremy M Wolfe, "Guided search 2.0 – a revised model of visual-search", *Psychonomic Bulletin & Review* 1 (1994), 202-238.

29   Itti et al., *A saliency-based search mechanism for overt and covert shifts of visual attention* cit.

30   Mozer et al., *Computational Modeling of Spatial Attention* cit.

31   Navon, *What Does a Compound Letter Tell the Psychologist's Mind?* cit. Typical visual search results are described in Treisman, *The perception of features and objects* cit., but the alternative approach to visual search presented in John Duncan and Glyn W Humphreys, "Visual search and stimulus similarity", *Psychological Review* 91 (1989), 433-458 seems particularly relevant to my work because it relies on grouping by similarity rather than feature-based attention to explain the observations.

Hayden Walles has a broad background in mathematics, physics and computing. He is a member of the Artificial Intelligence Research Laboratory at the University of Otago, Dunedin, New Zealand and is currently completing his PhD in computer science with a thesis titled "Language and Sensory Motor Cognition". Hayden also writes popular science articles for magazines and newspapers (including the *Otago Daily Times* and *The Press*).